CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 119, No. 5 May 2006

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

ACOUSTICAL MEASUREMENTS AND INSTRUMENTATION [58]

ACOUSTIC SIGNAL PROCESSING [60]

PHYSIOLOGICAL ACOUSTICS [64]

PSYCHOLOGICAL ACOUSTICS [66]

SPEECH PRODUCTION [70]

(Continued)

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, 2005

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).

Aaron, Arnold
Aarts, Ronald
Abbas, Paul
Abbott, Robert
Abdelli-Beruh, Nassima
Abel, Sharon
Abom, Mats
Abraham, Bruce
Abramovich, Haim
Abu, Nur
Achenbach, Jan
Adachi, Seiji
Agullo, Joaquim
Ahnert, Wolfgang
Ajaev, Vladimir
Akagi, Masato
Akay, Adnan
Akeroyd, Michael
Alain, Claude
Alam, Sheikh
Albert, Donald
Aldridge, David
Alipour, Fariborz
Alkoy, Sedat
Alku, Paavo
Allard, Jean-François
Allen, Jont
Allen, Paul
Alshits, V.
Anand, G.
Anderson, David
Anderson, John
Ando, Yoichi
Andringa, Tjeerd
Andruski, Jean
Anemuller, Jorn
Angster, Judit
Aoyama, Katsura
Apoux, Frederic
Arehart, Kathryn
Arenas, Jorge
Arieh, Yoav
Arkhipkin, Alexander
Arnott, William
Aronov, Boris
Arvelo, Juan
Ashdown, Ian
Ashley, Ric
Askenfelt, Anders
Assmann, Peter
Astley, Jeremy
Astolfi, Arianna
Atalla, Noureddine
Attenborough, Keith
Au, Whitlow W. L.
Auer, Edward
Austeng, Andreas
Auther, Linda
Avan, Paul
Aydelott, Jennifer

Backhaus, Scott
Bacon, Sid
Baddour, Ralph
Bader, Rolf
Badiey, Mohsen
Badin, Pierre
Baer, Ralph
Baer, Thomas
Baggenstoss, Paul
Bai, Mingsian
Bailey, Michael
Baker, Richard
Baker, Steven
Balaban, Evan
Balachandran, Bala
Ballato, Arthur
Banai, Karen
Bank, Balazs
Barber, Frank
Barbone, Paul
Baritkar, Sunil
Barkmeier, Julie
Barlow, Jay
Barnard, Daniel J.
Barney, Anna
Barron, Michael
Barros, Allan Kardec
Barry, William
Bass, Andrew
Bastyr, Kevin
Bauer, Gordon
Bauer, Jay
Bazua Duran, Carmen
Bech, Soren
Becker, Adib
Beckman, Mary
Bedard, Alfred J.
Behler, Gottfried
Behzad, Mehdi
Bell, James
Bell, Thaddeus
Bellandese, Mary H.
Bell-Berti, Fredericka
Benaroya, Haym
Benech, Nicolas
Benké, Jose
Bensa, Julien
Bent, Tessa
Beran, Mark
Beranek, Leo
Berengier, Michel
Berg, Bruce
Berger, Elliott
Berglund, Birgitta
Berkhoff, Arthur
Berman, David Harry
Bernstein, Joshua
Bernstein, Leslie
Bernstein, Lynne
Berry, David
Berryman, James
Berthelot, Yves H.
Bertsch, Matthias
Beskow, Jonas
Besnainou, Charles
Best, Virginia
Betlehem, Terence
Beutelmann, Rainer
Bian, Lin
Bilsen, Frans
Bishop, Garner
Bissinger, George
Bistafa, Sylvio
Blacklock, Oliver
Blackstock, David
Blanc-Benon, Philippe
Blauert, Jens
Bliss, Donald
Blok, Hans
Bloothooft, Gerrit
Blotter, Jonathan
Blottman, John
Bochner, Joseph
Boersma, Paul
Bohnenstiehl, Delwagne
Boike, Kumiko
Boisvert, Jeffrey
Bolton, J. Stuart
Bond, Zinny
Bostrom, Anders
Boucher, Victor
Boulanger, Patrice
Bouzidi, Youcef
Boyce, Suzanne
Braasch, Jonas
Bradley, David
Braida, Louis
Brammer, Anthony
Brand, Thomas
Brandstadt, Peter
Branstetter, Brian
Braun, Martin
Breazeale, Mack
Breebaart, Jeroen
Brennan, Michael
Bressmann, Tim
Brick, Haike
Bronkhorst, Adelbert
Brooks, Bennett M.
Brooks, Thomas
Brown, Charles
Brown, David
Brown, Michael
Brujan, E. A.
Bruneau, Michel
Brunel, J. F.
Brungart, Douglas
Brungart, Timothy
Buck, John

 0001-4966/2006/119(5)/2541/8/$22.50

Buckingham, Michael
Buder, Eugene
Bunnell, H. Timothy
Bunton, Kate
Burgess, William
Burkard, Robert
Burkhardt, John
Burnett, David
Burnett, Theresa
Burnham, Denis
Busch-Vishniac, I. J.
Buss, Emily

Cable, Peter
Cagno, Enrico
Cai, Liang-Wu
Calamia, Paul
Campbell, Murray
Candy, James
Cannata, J. M.
Cantrell, John
Capone, Dean
Carcaterra, Antonio
Carevic, Dragana
Carey, William M.
Cargill, G. Slade
Cariani, Peter
Carlile, Simon
Carlyon, Robert
Carney, Laurel
Caronti, Alessandro
Carre, Rene
Carroll, Gerard
Carstensen, Edwin
Cartling, Bo
Carvalho, António
Castaings, Michel
Cathignol, Dominique
Cato, Douglas
Causse, Rene
Cawley, Peter
Cazals, Yves
Cazzolato, Ben
Chaigne, Antoine
Chambers, David
Chambers, James
Chan, Roger
Chapman, David
Chapman, John
Chapman, N. Ross
Chatterjee, Monita
Cheatham, Mary Ann
Chen, Hongbin
Chen, Jeng
Chen, Sin-Horng
Cheng, Anping
Cheng, Arthur
Cheng, Hsien
Chevret, Patrick
Chi, Taishih
Chimenti, Dale
Chin-Bing, Stanley
Chiroiu, Veturia
Chiu, Ching-Sang
Cholewiak, Roger
Chotiros, Nicholas
Choy, Boris
Choy, Yat-sze
Chu, Dezhang
Chunchuzov, Igor
Chung, Jing-Yau
Church, Charles
Church, Michael
Cipolla, Jeffrey
Ciric, Dejan
Citrin, David
Clark, Cathy Ann
Clark, Christopher
Clark, Robert
Clarke, Constance
Clarke, Eric
Cleary, Miranda
Clermont, Frantz
Clopper, Cynthia
Coakley, William
Cobbold, Richard
Cobo, Pedro
Colburn, H. Steven
Cole, Ronald
Collier, Robert
Collier, Sandra
Collins, Leslie
Collins, Michael
Colonius, Tim
Colosi, John
Coltman, John W.
Cone-Wesson, Barbara
Cooke, Martin
Coombs, Sheryl
Cooper, Nigel
Cornuelle, Bruce
Corsaro, Robert
Costa, Marcio H.
Cottingham, James
Coulouvrat, François
Coussios, Constantin
Cowan, M. L.
Cowan, Robert S.
Cox, Trevor
Cranen, Bert
Cranford, Ted
Cray, Benjamin
Creamer, Dennis
Cross, Eric
Cugnet, Boris
Culling, John
Culver, Richard
Cummings, William
Cuschieri, Joseph
Cutler, Anne

Dacol, Dalcio
Dahl, Peter
Dai, Huanping
Daigle, Gilles
D'Alessandro, C.
Daley, Tom
Dance, Stephen
Dang, Changjiu
Dang, Jianwu
D'Antonio, Peter
Darwin, Christopher
Dasgupta, Nilanjan
Dau, Torsten
Davidson, Lisa
Davis, Anthony
Davis, Barbara
Davis, Chris
Davy, John
Daya, El Mostapha
Dayton, Paul
Deane, Grant
DeBlok, Kees
De Boer, Egbert
De Cheveigne, Alain
Declerq, Nico
De Donato, Stefano
Defrance, Jerome
De Groot-Hedlin, C.
De Jong, Kenneth
DelBalzo, Donald
Delgutte, Bertrand
DeLillo, Thomas
DeLong, Caroline
Demany, Laurent
Demiroglu, Cenk
Denardo, Bruce
Deng, Cheri
Deng, Li
Denzinger, Annette
Derode, Arnaud
DeRosny, Julien
Deshmukh, Om
Desloge, Jay G.
Desmet, Wim
Deutsch, Diana
Deverge, Mickale
De Vries, Diemer
Dhar, Sumit
D'hooge, Jan
Diachok, Orest
DiBiase, Joseph
Diederich, Chris
Ding, Desheng
Ding, Heping
Divenyi, Pierre
Djelouah, Hakim
Dobie, Robert
Dodd, Tony
Doellinger, Michael
Doinikov, Alexander
Dolson, Mark
Donaldson, Gail
Donskoy, Dimitri
Dooling, Robert
Dorman, Michael
Dos Santos, Serge
Dosso, Stan
Dowling, David
Dreisbach, Laura
Drennan, Ward
Dromey, Christopher
Druyvesteyn, Willem
D'Spain, Gerald
Dual, Jurg
Dubno, Judy
Ducasse, Eric
Duda, Richard
Duda, Timothy
Duhamel, Denis
Duifhuis, Hendrikus
Durgin, William
Durlach, Nathaniel
Dusan, Sorin
Dye, Raymond
Dzieciuch, Matthew

Ebbini, Emad
Ebenezer, D.
Eck, Douglas
Eddington, Donald
Eddins, David
Edds-Walton, Peggy
Edelmann, Geoffrey
Edwards, Brent
Eldredge, Jeff
Elko, Gary W.
Elliott, Stephen
Ellis, Dale
Ellis, Daniel
Elsberry, Wesley
Enflo, Bengt
Engdahl, Bo
Engwall, Olov
Epstein, Michael
Erbe, Christine
Ermann, Michael
Escudero, Paola
Espy-Wilson, Carol
Evans, Bronwen
Evans, Jack
Evans, Richard
Everbach, E. Carr
Everstine, Gordon
Ewert, Stephan D.

Füllgrabe, Christian
Fagerlund, Al
Fahey, Paul
Fairweather, Graeme
Fallat, Mark
Faller, Christof
Fant, Gunnar
Fantini, Deborah
Farabee, Theodore
Farassat, Feri
Farouk, Bakhtier
Fastl, Hugo
Fatemi, Mostafa
Faulkner, Andrew
Fawcett, John
Feeney, M. Patrick
Feit, David
Feleppa, Ernest Joseph
Fellah, Zine Abiddine
Ferguson, Sarah
Ferri, Aldo A.

Feth, Lawrence
Fidell, Sanford
Fields, James
Finan, Don
Fine, Michael
Finette, Steven
Finneran, James
Fisher, Karl
Fitzgibbons, Peter
Fitzpatrick, John
Flandrin, Patrick
Flatté, Stanley
Fletcher, Neville H.
Flint, John
Florentine, Mary
Folegot, Thomas
Folkow, Peter
Forbes, Barbara
Ford, John
Forrest, Karen
Fosler-Lussier, Eric
Fourakis, Margaritis
Fowler, Carol
Fox, Robert
Fox, Warren
Frampton, Kenneth
Franck, Andreas
Frank, Scott
Franzoni, Linda
Frazer, L. Neil
Fredericksen, Erling
Freeman, Dennis
Frendi, Kader
Freyman, Richard
Frijns, J. H. M
Frisina, Robert D.
Frisk, George V.
Fromm, David
Fromme, Paul
Fu, Qian-Jie
Fuks, Iosif
Fung, K. Y.
Fung, Pascale
Furusawa, Masahiko
Fuzessery, Zoltan

Gaja, Esteban
Gallun, Frederick
Gan, Rong
Gan, Woon
Garces, Milton
Gardonio, Paolo
Garrelick, Joel
Garrett, Steven
Gaumond, Charles F.
Gaunaurd, G. C.
Gauss, Roger
Gavin, Joseph
Gedamke, Jason
Gedeon, David
Gendron, Paul
Gentry, Roger
Genuit, Klaus
Georgios, Natsiopoulos
Gershfeld, Jonathan
Gerstein, Edmund
Gerstoft, Peter
Gibbs, Gary
Gifford, Rene
Giguere, Christian
Gilbert, Kenneth
Gilbert, Richard
Gilkey, Robert
Ginsberg, Jerry
Giordano, Nicholas J.
Glauber, Claus-C.
Glegg, Stewart
Glorieux, Christ
Gockel, Hedwig
Godin, Oleg
Goedecke, George
Goffaux, Cecile
Goldinger, Stephen
Goldrick, Matthew
Goldstein, Albert
Gordon, Grant
Gordon, Jonathan
Gordon-Salant, Sandra
Gorga, Michael
Gottfried, Terry
Gottlieb, Hans
Goussev, Vitalyi
Gover, Bradford
Gow, David
Gracewski, Sheryl
Granqvist, Svante
Grantham, D. Wesley
Graziani, Giorgio
Green, Tim
Greenberg, Julie
Greenberg, Steven
Greene, Charles
Greenleaf, James
Gregory, Michelle
Griesinger, David
Grill, Wolfgang
Grimault, Nicolas
Grinchenko, Victor
Grose, John
Grosh, Karl
Gruber, Fred
Gspan, Stefan
Gucunski, Nenad
Guillot, Francois
Guinan, John
Guion, Susan
Gumerov, Nail A.
Gummer, Anthony
Guo, Ningqun
Gurevich, Boris
Gusev, Vitali
Guski, Rainer
Gussenhoven, Carlos
Guyader, Jean-Luis

Haberman, Michael
Haberman, Robert
Hachiya, Hiroyuki
Hadimioglu, Babur
Hagen, Mechthild
Hahn, Steven
Hailman, Jack
Hall, Joseph W., III
Hall, Donald
Hall, Marshall
Hall, Neal
Hammershoi, Dorte
Hansen, Colin
Hansen, Ole
Hansen, Uwe J.
Hanson, Helen M.
Harari, Isaac
Hardin, Jay
Harding, Sue
Hare, James
Harland, Andy
Harma, Aki
Harnsberger, James
Harrison, Christopher
Hartley, Douglas
Hartmann, William M.
Hasegawa-Johnson, M.
Hassan, Scott
Hay, Jen
Hayama, Atsushi
Hayek, Carleton
Hayes, Erin
Hayes-Harb, Rachel
Hazan, Valerie
Healy, Eric
Heaney, Kevin
Hedberg, Claes
Hedlin, Michael
Hedrick, Mark
Heffner, Henry
Heffner, Rickeye
Heil, Peter
Heimann, Dietrich
Heinz, Michael
Helfer, Karen
Henderson, Paul
Henry, James
Henyey, Frank
Herdic, Peter
Hermand, Jean-Pierre
Hermansky, Hynek
Herrmann, Barbara
Herrmann, Harald
Hertrich, Ingo
Herzel, Hans-Peter
Herzing, Denise
Hildebrand, John
Hillenbrand, James
Hillyard, Steven
Hinders, Mark
Hirata, Yukari
Hirose, Sohichi
Hirschberg, Avraham
Hirsekorn, Martin
Hirsekorn, Sigrun
Hobaek, Halvor
Hodgkiss, William
Hodgson, Murray
Hoekje, Peter
Hoffmeister, Brent
Hofman, Paul
Hojen, Anders
Hole, Lars
Holland, Charles
Holliday, D. Vance
Holm, Sverre
Holmes, Jason
Holste, Fredi
Holt, Lori
Homm, Anton
Honda, Kiyoshi
Honorof, Douglas
Hoole, Philip
Horacek, Jaromir
Horonjeff, Richard
Horoshenkov, Kirill
Horst, Johannes
Horton, William
Hosokawa, Atsushi
Hosten, Bernard
Houde, John
Houser, Dorian
Houtgast, Tammo
Houtsma, Adrianus
Hovem, Jens
Howard, Carl
Howe, Bruce
Howe, Michael
Hu, Yi
Huang, Lixi
Huber, Jessica
Huffman, Marie
Hull, Andrew
Humes, Larry
Humphrey, Joseph
Humphreys, William
Hunt, William
Hunter, Eric
Hurley, David
Hursky, Paul
Hutchins, David
Hwang, Yun-Fan

Ih, Jeong-Guon
Ihlenburg, Frank
Ilgamov, MaratA
Ilinskii, Yurii
Imai, Satomi
Imaizumi, Satoshi
Insley, Stephen
Isabelle, Scott
Isakov, Victor
Isakson, Marcia
Ishimaru, Akira
Iskarous, Khalil
Ison, James
Ito, Masashi
Iverson, Paul

Jackson, Darrell R.

Jacobs, Laurence
Jacobsen, Finn
Jaeger, Stephen
Janssen, Thomas
Jansson, Erik
Jarzynski, Jacek
Jastreboff, Pawel
Javel, Eric
Jen, Philip
Jensen, Finn
Jensen, Jorgen
Jerger, James
Jerzak, Wayne
Jesse, Alexandra
Jesteadt, Walt
Jesus, Sergio
Jia, Xiaoping
Jiang, Jack
Johnson, Joel
Johnson, Mark
Johnson, Marty
Johnson, Michael
Johnston, Patrick
Jones, Joie P.
Jongman, Allard
Jot, Jean-Marc
Juve, Daniel

Kaernbach, Christian
Kalb, Joel
Kalinowski, Joseph
Kallivokas, Loukas
Kallman, Jeffrey
Kalveram, Karl
Kamakura, Tomoo
Kang, Jian
Kaplunov, Julius
Kar, T.
Kargl, Steven
Karnell, Michael
Kastak, David
Kato, Yuichi
Katz, Brian
Katz, William
Kaufman, Jonathan
Kawahara, Hideki
Keefe, Douglas
Keiffer, Richard
Keilman, George
Keith, William
Kellermann, Walter
Kelly, Jack
Kelly, Mark
Keltie, Richard
Kemp, David
Kenbu, Teramoto
Keolian, Robert
Kergomard, Jean
Kewley-Port, Diane
Khismatullin, Damir
Khokhlova, Vera
Kidd, Gary
Kiefte, Michael
Kim, Benjamin
Kim, Duck
Kim, Jaehwan
Kim, Jin-Yeon
Kim, Kang
Kim, Katherine
Kim, Sang-Myeong
Kim, Yang-Hann
King, Simon
Kingston, John
Kinra, Vikram
Kirby, Raymond
Kirkeby, Ole
Kishon-Rabin, Liat
Kistler, Doris
Kitahara, Mafuyu
Kitamura, Christine
Klapuri, Anssi
Klis, S. F.
Klump, Georg
Knight, Richard
Knobles, David
Kob, Malte
Kochanski, Greg
Kochetov, Alexai
Koehnke, Janet
Koenig, Laura
Koessl, Manfred
Kohler, Klaus
Komatitsch, Dimitri
Kong, Ying-Yee
Konofagou, Elisa
Kook, Hyungseok
Korman, Murray
Kozick, Richard
Krakow, Rena
Krane, Michael
Krasovitski, Boris
Krause, Jean
Kreider, Kevin
Kreiman, Jody
Kripfgans, Oliver
Krishnan, A.
Krishnaswamy, Sridhar
Kristiansen, Ulf
Krog, Norun
Kromine, Alex
Krumbholz, Katrin
Krutyansky, Leonid
Krylov, Victor
Kuhl, Patricia K.
Kundu, Tribikram
Kuperman, William A.
Kuttruff, K. Heinrich
Kwon, Hyu-Sang

Labov, William
Lacerda, Francisco
Lachs, Lorin
Ladd, Robert
Ladich, Friedrich
Laeufer, Christiane
Lagree, Pierre-Yves
Lam, Yiu
Lammers, Marc
Lane, Courtney
Langton, Christian
Lanza di Scalea, F.
Larose, Eric
Larson, Gregg
Lasaygues, Philippe
Lasky, Robert
Lauchle, Gerald
Laugier, Pascal
Laura, Patricio
Lauriks, Walter
Lauter, Judith
Lauterborn, Werner H.
Lavery, Andone
Lawrie, Jane
Leamy, Michael
Leble, Sergey
Lee, Jang Moo
Lee, Moohyung
Lee, Tan
Leek, Marjorie
Legrand, Olivier
Lehman, Sean F.
Leibold, Lori
Leishman, Timothy
Lemonds, David
Lentz, Jennifer
Leon, F.
LePage, Kevin
Lerch, Reinhard
Lerro, Donald
Letowski, Tomasz
Levitt, Harry
Lewin, Peter
Li, Kai Ming
Lim, Raymond
Lin, Feng
Lindberg, Jan
Lineton, Ben
Ling, Ju
Lingevitch, Joseph
Lipkens, Bart
Lister, Jennifer
Litovsky, Ruth
Litt, Mitchell
Liu, Daben
Liu, Qing
Liu, Tianshu
Lo, Kam
Lobkis, Oleg
Lofqvist, Anders
Loizou, Philipos
Lopez-Poveda, Enrique
Lorenzi, Christian
Lotto, Andrew
Louie, John
Lowe, Michael
Lowenstein, Joanna
Lubman, David
Lucero, Jorge
Lucks Mendel, Lisa
Ludlow, Christy
Lueptow, Richard
Lunde, Per
Lynch, James
Lyon, Richard H.

Macaulay, Gavin
Mackersie, Carol
Macpherson, Ewan
Madsen, Peter
Maeda, Shinji
Maev, Roman
Mah, Marko
Mahshie, James
Makarov, Sergey
Makris, Nicholas
Malocha, Donald
Maltsev, Nick
Mamou, Jonathan
Manasseh, Richard
Mandelis, Andreas
Manley, Martin
Mann, David
Mann, Virginia
Mapp, Peter
Marburg, Steffen
Marengo, Edwin
Marshall, Lynne
Marshall, William
Marston, Philip
Martens, William
Martin, Glen
Martin, Paul
Martinez, Rudolph
Martinez-Sala, Rosa
Marvit, Peter
Mason, Christine
Mason, Matthew
Massaro, Dominic
Masson, Patrice
Mast, T. Douglas
Masuda, Mitsuhiro
Mathevon, Nicolas
Mattei, Pierre-Olivier
Matthies, Melanie
Matveev, Konstantin
Mauermann, Manfred
Maugin, Gerard
Maury, Dominique
Max, Ludo
Maxfield, Bruce
Maye, Jessica
Maynard, Julian
Mayo, Catherine
Maysenhoelder, W.
Maze, Gérard
McCoy, John
McDaniel, James
McDermott, Hugh
McDermott, Joshua
McFadden, Dennis
McGough, Robert
McGowan, Richard
McKay, Colette
McLaughlin, Stephen
McPherson, David
McRoberts, Gerald

Meadows, Guy
Meddis, Ray
Mellinger, David
Meloni, Tommaso
Mendez, Rafael
Menn, Lise
Meo, Michele
Meyer, Jens
Mi, Bao
Michaels, Jennifer
Michaels, Thomas
Michalopoulou, Z.
Micheyl, Christophe
Michielssen, Eric
Middlebrooks, John
Miedema, Henk
Miles, Ronald
Miller, Douglas
Miller, James
Miller, James H.
Miller, Lee
Millman, Rebecca
Milner, Ben
Mironov, Mikhail
Misun, Vojtech
Mitri, Farid
Mitson, Ron
Mitterer, Holger
Miyara, Federico
Miyazaki, Kenichi
Mobley, Joel
Mohl, Bertel
Mokhtari, Parham
Moldover, Michael
Molis, Michelle
Monahan, Edward
Monchalin, Jean-Pierre
Mongeau, Luc
Montaldo, Gabriel
Moon, Won-kyu
Moore, Brian
Moore, James
Mooshammer, Christine
Moreau, Andre
Moreton, Elliott
Moretti, David
Morfey, Christopher
Morris, Philip
Mould, John
Mountain, David
Mourad, Pierre
Mozurkewich, George
Muehleisen, Ralph
Muesch, Hannes
Muir, Thomas
Mukai, Tohru
Mullennix, John
Mullins, Joe
Munhall, Kevin
Munjal, M.
Munson, Benjamin
Murphy, Peter
Murphy, William
Murray, Narelle
Murray, Todd

Nabelek, Anna K.
Nachtigall, Paul
Nagem, Raymond
Naghshineh, Koorosh
Nagy, Peter
Nakai, Takayoshi
Nakajima, Yoshitaka
Narayanan, Shrikanth
Naugolnykh, K.
Naylor, Graham
Neal, Steven
Nearey, Terrance
Nederveen, Cornelis
Neel, Amy
Neely, Stephen
Neff, Donna
Neild, Adrian
Neitzel, Richard
Nelson, David
Nelson, Douglas A.
Nelson, Peggy B.
Newhall, Bruce
Newman, Rochelle
Ng, Manwa
Nicholson, Patrick
Niedzielski, Nancy
Nielsen, Peter
Niezrecki, Christopher
Nightingale, K. R.
Nishi, Kanae
Nissen, Shawn
Nittrouer, Susan
Niyogi, Partha
Noble, John
Nolan, Francis
Nolle, Alfred
Nordborg, Anders
Norris, Andrew
Norris, David
Norton, Guy
Norton, Stephen
Nosal, Eva-Marie
Nowacek, Douglas
Nusbaum, Howard
Nyborg, Wesley
Nye, Patrick
Nygaard, Lynne

Oba, Roger
Obara, Clifford
O'Brien, W. D., Jr.
O'Donnell, Matthew
Oelze, Michael
Ogren, Mikael
Ohde, Ralph
Ohl, Claus-Dieter
Ohno, Masahiro
Oliphant, Travis
Olive, Sean
Olson, Elizabeth
Orduna-Bustamante, F.
Ostashev, Vladimir
Oswald, Julie
Ouni, Slim
Ozyoruk, Yusuf
Pabon, Peter
Page, John
Pagneux, Vincent
Pallier, Christophe
Palmer, Alan
Pan, Jie
Papanicolaou, G. C.
Parastates, Elias
Pardo, Jennifer
Parker, Scott
Parkins, John
Parrott, Tony
Parthasarathi, Anand
Pasewark, Bruce
Patel, Aniruddh
Patterson, Roy
Patuzzi, Robert
Pavan, Gianni
Payton, Karen
Peake, Nigel
Peat, Keith
Pelorson, Xavier
Pereira, Jose Carlos
Peskin, Charles
Peters, Robert
Petersson, Björn
Pfingst, Bryan
Phillips, Dennis
Photiadis, Douglas
Picaut, Judicaël
Pichora-Fuller, M.
Picone, Joseph
Pierce, Allan D.
Pierucci, Mauro
Piquette, Jean
Pisoni, David
Pittman, Andrea
Plack, Christopher
Plant, Randall
Plesniak, Michael
Plotkin, Kenneth
Pluymers, Bert
Poese, Matthew
Polka, Linda
Pollack, Gerald
Pols, Louis
Polyzos, Demosthenes
Poncelet, Olivier
Port, Robert
Porter, Mike
Potel, Catherine
Poterek, Thomas
Potty, Gopu
Prada, Claire
Predoi, Mihai
Preis, Anna
Preisig, James
Pressnitzer, Daniel
Preston, Dennis
Preston, John
Prosek, Robert
Purcell, David
Putnam, Gabe

Qi, Yingyong
Qian, Menglu
Qin, Michael
Qu, Jianmin
Quazi, Azizul
Quené, Hugo
Quinlan, Angela

Rabbitt, Richard
Radcliffe, Clark
Radebaugh, Ray
Rafaely, Boaz
Raichel, Daniel
Rajan, Subrahaniam
Rajapan, Dhilsha
Raju, P.
Rakerd, Brad
Ram, Yitshak
Ramakrishnan, Ramani
Ramis, Jaime
Ramsay, Gordon
Rand, Richard
Rapids, Brian
Rasmussen, Gunnar
Raspet, Richard
Ratilal, Purnima
Recio, Alberto
Reed, Charlotte
Reeder, D.
Reep, Roger
Rehrig, Paul
Reichmuth-Kastak, C.
Reid, John
Reissland, Nadja
Remez, Robert
Remington, Paul
Ren, Tianying
Repp, Bruno
Ressler, Patrick
Rhebergen, Koenraad
Rhode, William
Richards, Roger
Richards, Virginia
Richardson, Bernard
Richardson, W.
Richie, Carolyn
Richter, Utz
Rienstra, Sjoerd
Ritcey, James
Rivens, Ian
Roberts, Brian
Rogers, Catherine
Rogers, John
Rokhlin, Stanislav
Romeo, Maurizio
Roosnek, Nico
Rose, Joseph
Rose, Marina
Rose, Philip

Rosenblum, Lawrence
Rosowski, John
Rossing, Thomas
Rothpletz, Ann
Rouseff, Daniel
Roux, Philippe
Roy, Ronald
Royston, Thomas
Rubin, Philip
Rubinstein, Jay
Rudenko, Oleg
Rumerman, Melvyn
Rumsey, Francis
Ryaboy, Vyacheslav
Ryan, James
Ryden, Nils

Sabra, Karim
Sagayama, Shigeki
Saijo, Yoshifumi
Saijyou, Kenji
Saito, Shigemi
Sakagami, Kimihiro
Salt, Alec
Salzberg, David
Samuel, Arthur
Sanchez-Morcillo, V.
Sarkissian, Angie
Sarradj, Ennes
Sas, Paul
Sato, Hiroshi
Saylor, John
Scales, John
Scandrett, Clyde
Scarborough, Rebecca
Schairer, Kim
Schanz, Martin
Scharf, Bertram
Scheich, Henning
Schlauch, Robert
Schlaug, Gottfried
Schmitt, Denis
Schmitt, Douglas
Schneider, Bruce
Schneider, John
Schobben, Daniel
Schomer, Paul D.
Schubert, Frank
Schuhmacher, Andreas
Schuller, Gerd
Schulte-Fortkamp, B.
Schultz, Tanja
Schwartz, Jean-Luc
Schwartz, Joshua
Scordilis, Michael
Seeber, Bernhard
Sek, Aleksander
Selamet, Ahmet
Seo, Sang-Hyun
Seong, Woojae
Serbyn, M. Roman
Sereno, Joan
Serrano, Elba
Sethares, William
Sgard, Franck
Shackleton, Trevor
Shadle, Christine
Shah, Amee
Shamma, Shihab
Shams, Qamar
Shang, Er-Chang
Shannon, Robert
Shattuck-Hufnagel, Stephanie
Shaw, Neil
Sheft, Stanley
Sheng, Ping
Shepard, W.
Shepherd, Robert
Shera, Christopher
Sherlock, Laguinn
Shih, Chilin
Shinn-Cunningham, B.
Shorter, Philip
Shung, K.
Sibul, Leon
Siderius, Martin
Siegel, Jonathan
Siegmann, William L.
Simmons, Andrea M.
Simons, Dick
Simpson, Brian
Simpson, Roger
Sinex, Donal
Sinha, Bikash
Sinha, Dipen
Sinor, Milan
Sisto, Renata
Skowronski, Mark
Slaney, Malcolm
Slaton, William
Slepyan, Leonid I.
Small, Steve
Smeulders, David
Smiljanic, Rajka
Smith, Bruce
Smith, David
Smith, Jerry
Smith, John
Smith, Julius
Smith, Kevin B.
Smith, Suzanne
Smurzynski, Jacek
Smyth, Tamara
Snell, Karen
Snieder, Roel
Snyder, Russell
Solodov, Igor
Solokhin, Nick
Sommerfeldt, Scott D.
Sommers, Mitchell
Sondhi, Mohan
Song, Hee-Chun
Sorokin, Victor
Southall, Brandon
Souza, Luiz L.
Spahr, Anthony
Sparrow, Victor W.
Spicer, James
Spies, Martin
Spiesberger, John
Spoor, Philip
Sproat, Richard
Stafford, Kathleen
Stanic, Steve
Stanton, Timothy
Stecker, G. Christopher
Steele, Charles
Stellmack, Mark
Stelmachowicz, Pat
Stenfelt, Stefan
Stephan, Yann
Stephen, Ralph
Stephenson, Uwe
Stevens, Kate
Stevens, Kenneth
Stickney, Ginger
Stinson, Michael
Stockwell, Robert
Stojanovic, Milica
Stokes, Ann
Stone, Maureen
Stoughton, John W.
Strange, Winifred
Strasberg, Murray
Strickland, Elizabeth
Strifors, Hans
Strik, Helmer
Strube, Hans Werner
Studebaker, Gerald
Sturdy, Christopher
Su, Zuhre
Succi, George
Sugimoto, Nobumasa
Sujbert, Laszlo
Sujith, R.
Sullivan, Anders
Sullivan, Edmund
Sum, K. S.
Summerfield, Quentin
Summers, Jason
Summers, Van
Sundara, Megha
Sundberg, Johan
Sung, Chia-Chi
Supin, Alexander
Surprenant, Aimee
Susan, Voss
Sussman, Elyse
Sussman, Joan
Sutherland, Louis
Sutin, Alexander
Svec, Jan
Svensson, Peter
Szabo, Thomas

Tabain, Marija
Taherzadeh, Shahram
Tajima, Keiichi
Takagi, Kenshiro
Takahashi, Terry
Talavage, Thomas
Tam, C. K. W.
Tang, S. K.
Tanter, Mickael
Tarnow, Viggo
Taroudakis, Michael
Tasko, Stephen
Tatarinov, Alexey
Tattersall, J.
Taylor, S. Larry
Telschow, Ken
TenCate, James
TerHaar, Gail
Terhune, Jack
Ternström, Sten
Teutsch, Heinz
Thawani, P.
Thibodeau, Linda
Thode, Aaron
Thompson, David J.
Thompson, Lonny L.
Thompson, Stephen C.
Thorsos, Eric I.
Tichy, Jiri
Tiede, Mark
Tim, Bressmann
Timmerman, Nancy
Tindle, Christopher
Tinianov, Brandon
Tisseyre, Alain
Tito, Frank
Titze, Ingo
Tjaden, Kris
Toivanen, Jari
Tollin, Daniel
Tomilina, Tatiana
Tomsovic, Steven L.
Tourin, Arnaud
Trahey, Gregg
Tranter, Sue
Tremblay, Kelly
Tronchin, Lamberto
Tsahalis, Dimosthenis
Tseng, Chiu-yu
Tsogka, Chrysoula
Tuller, Betty
Tun, Pat
Turgut, Altan
Turner, Christopher
Turner, Joseph
Tyack, Peter
Tzanetakis, George

Umnova, Olga
Ungar, Eric E.
Unger, Evan
Unoki, Masashi
Uscinski, Barry

Vafin, Renat
Vainio, Martti
Valdivia, Nicolas
Valeau, Vincent
Valiere, J.-C.
Van de Par, Steven
Van der Heijden, M.

VanDijk, Pim
VanLieshout, Pascal
VanOpstal, John
VanParijs, Sofie
VanWieringent, Astrid
VanWijngaarden, S.
VanderMeulen, F.
Varghese, Tomy
Vatikiotis-Bateson, Eric
Vavrycuk, Vaclav
Verhey, Jesko
Verkerke, Gijsbertus
Vermeir, Gerrit
Verrillo, Ronald
Versfeld, Niek
Verweij, Martin
Vidmar, Paul
Vieira, Maurilio
Viemeister, Neal
Vignola, Joseph
Vilain, Coriandre
Vipperman, Jeffrey
Vitevich, Michael
Vladivia, Nicolas
Vliegen, Joyce
Vogel, Irene
Voon, L. C. Lew
Vorländer, Michael
Voronovich, Alexander
Vos, Henk
Vos, Piet
Voss, Susan
Vu-Quoc, Loc

Waag, Robert
Wada, Hiroshi
Wade, Travis
Wagstaff, Ronald
Wahlberg, Magnus
Wakefield, Gregory
Wakeland, Ray
Walhberg, Magnus
Walker, Bruce
Walsh, Timothy
Wang, Can-yun
Wang, DeLiang
Wang, Jiqing
Wang, Lugen
Wang, M.
Wang, Quan
Wang, Yadong
Wang, Yue
Ward, Lawrence
Warnock, Alf
Warren, Richard
Washburn, Karl
Watkin, Kenneth
Watkins, Anthony
Watson, Charles
Watson, Willie
Waubke, Holger
Waxler, Roger
Wayland, Ratree
Wear, Keith
Weaver, Richard
Weber, Reinhard
Weisenberger, Janet M.
Weismer, Gary
Weiss, Lora
Welkowitz, Walter
Wereley, Norman
Werner, Lynne
Werner, Stefan
Westwood, Evan
Wiegrebe, Lutz
Wiggins, Sean
Wightman, Fred
Wilcox, Paul
Wilen, Larry
Wilhelms-Tricarico, R.
Wilkens, Roy
Wilkens, Volker
Willatzen, Morten
Williams, Earl
Williams, Kevin
Wilmut, Mike
Wilson, D. Keith
Wilson, Preston
Wilson, Richard
Wilson, Robin
Wismer, Margaret
Withnell, Robert
Wittenberg, Thomas
Wojtczak, Magdalena
Wolfe, Joe
Wolfe, Patrick
Wolfson, Michael
Woloszyn, Philippe
Wong, George
Wong, Kainam
Wong, Patrick
Wood, Warren
Woodhouse, James
Worcester, Peter
Wouters, Jan
Wright, Andrew
Wright, Matthew
Wright, Richard
Wu, Chung-Hsien
Wu, Junru
Wu, Kuangcheng
Wu, Lixue
Wu, Ting-Wen
Wu, Tsung-Tsong

Xiang, Ning
Xu, Mubing
Xu, Yi

Yan, Hong
Yan, Zhongyu
Yang, Tsih
Yang, Xinmai
Yevick, David
Yiu, Edwin
Yost, William A.
Yunusova, Yana

Zabolotskaya, Evgenia
Zacharov, Nick
Zagrai, Andrei
Zahorian, Stephen
Zahorik, Pavel
Zahui, Marcellin
Zaitsev, V.
Zajac, David
Zalesak, Joseph
Zampolli, Mario
Zeddies, David
Zeegers, Jos
Zeng, Fan-Gang
Zhang, Xiaoming
Zhang, Zhaoyan
Zhong, Pei
Zhou, Ji-Xun
Zimmer, Karin
Zimmer, Walter
Ziomek, Lawrence
Zsiga, Elizabeth
Zurek, Patrick
Zurk, Lisa
Zwolan, Teresa

USA Meetings Calendar

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future.

	2006
6–9 June	151st Meeting of the Acoustical Society of America, Providence, RI [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
17–21 Sept.	INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA [www.interspeech2006.org; http://www.interspeech2006.org/].
28 Nov.–2 Dec.	152nd Meeting of the Acoustical Society of America joint with the Acoustical Society of Japan, Honolulu, HI [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747- 4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org]. Deadline for receipt of abstracts: 30 June 2006.
	2007
4–8 June	153rd Meeting of the Acoustical Society of America, Salt Lake City, UT [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, LA (note Tuesday through Saturday) [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747- 4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
	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, P. O. Box 1609, Groton, CT 06340-1609; Tel.: 860-572-0680; Web: www.icben.org; E-mail: icben2008@att.net].

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 STANDARDS NEWS

Susan B. Blaeser, Standards Manager
ASA Standards Secretariat, Acoustical Society of America, 35 Pinelawn Rd., Suite 114E, Melville, NY 11747 [Tel.: (631) 390-0215; Fax: (631) 390-0217; e-mail: asastds@aip.org]

George S. K. Wong
Acoustical Standards, Institute for National Measurement Standards, National Research Council, Ottawa, Ontario K1A 0R6, Canada [Tel.: (613) 993-6159; Fax: (613) 990-8765; e-mail: george.wong@nrc.ca]

American National Standards (ANSI Standards) developed by Accredited Standards Committees S1, S2, S3, and S12 in the areas of acoustics, mechanical vibration and shock, bioacoustics, and noise, respectively, are published by the Acoustical Society of America (ASA). In addition to these standards, ASA publishes Catalogs of Acoustical Standards, both National and International. To receive copies of the latest Standards Catalogs, please contact Susan B. Blaeser.

Comments are welcomed on all material in Acoustical Standards News.

This Acoustical Standards News section in JASA, as well as the National and International Catalogs of Acoustical Standards, and other information on the Standards Program of the Acoustical Society of America, are available via the ASA home page: http://asa.aip.org.

Standards Meetings Calendar National

The 151st ASA Meeting, Providence, RI, the Westin Providence Hotel, 5-9 June 2006, the ASA Committee on Standards (ASACOS), ASACOS STEERING Committees, and Accredited Standards Committees S1 Acoustics, S2 Mechanical Vibration and Shock, S3 Bioacoustics, and S12 Noise, and the Standards Plenary Group will meet as follows:

•Monday, 5 June 2006 7:00 P.M.

ASACOS Steering Committee

•Tuesday, 6 June 2006 8.30 A.M.

ASA Committee on Standards (ASACOS). Meeting of the Committee that directs the Standards Program of the Acoustical Society.

•Tuesday, 6 June 2006 10:30 A.M.

Standards Plenary Group. To discuss national and international items relevant to Accredited Standards Committees S1, S2, S3, and S12. This meeting also includes the annual meetings of the U.S. Technical Advisory Groups (TAGs) for ISO/TC 43 Acoustics, ISO/TC 43/SC 1 Noise, IEC/TC 29 Electroacoustics, ISO/TC 108 Mechanical Vibration and Shock, ISO/TC 108/SC 2 Measurement and Evaluation of Mechanical Vibration and Shock as Applied to Machines, Vehicles and Structures; ISO/TC 108/SC3 Use and Calibration of Vibration and Shock Measuring Instruments, ISO/TC 108/SC 4 Human Exposure to Mechanical Vibration and Shock; ISO/TC 108/SC 5 Condition Monitoring and Diagnostics of Machines, and ISO/TC 108/SC 6 Vibration and Shock Generating Systems.

•Tuesday, 6 June 2006 1:45 P.M.

Accredited Standards Committee S1 on Acoustics.

•Tuesday, 6 June 2006 3:30 P.M.

Accredited Standards Committee S12 on Noise.

•Wednesday, 7 June 2006 8:30 A.M.

Accredited Standards Committee S2 on Mechanical Vibration and Shock.

•Wednesday, 7 June 2006 10:30 A.M.

Accredited Standards Committee S3 on Bioacoustics.

American National Standards Call for Comment on Proposals Listed

This section solicits comments on proposed new American National Standards and on proposals to revise, reaffirm, or withdraw existing standards. The dates listed in parentheses are for information only.

ASA (ASC S1) (Acoustical Society of America)

REAFFIRMATIONS

BSR S1.4-1983 (R200x), Specification for Sound Level Meters [reaffirmation of ANSI S1.4-1983 (R 2001)]

This standard specifies minimum requirements for three basic types of sound level meters, types 0, 1, and 2, with performance requirements that become progressively less stringent, proceeding from type 0 to type 2. For each type, the standard requires three frequency weightings, *A*, *B*, and *C*; and two exponential- time-averaging characteristics, slow and fast.

BSR S1.4 A -1985 (R 200x), Amendment to ANSI S1.4-1983 [reaffirmation of ANSI S1.4 A-1985 (R 2001)]

The intent of this amendment is (a) to specify the electrical design goal for the relative response characteristics of the *A*-weighted frequency response up to 100 kHz; and (b) to specify type 0 and type 1 tolerance limits on relative electrical response for frequencies between 16 kHz and 100 kHz.

BSR S1.6-1984 (R 200x), Preferred Frequencies, Frequency Levels, and Band Numbers for Acoustical Measurements [reaffirmation of ANSI S1.6-1984 (R 2001)]

This standard defines the preferred frequencies, or nominal band-center frequencies to be used for acoustical measurements. Frequency levels or band numbers are associated with these sets of frequencies, and the preferred frequencies are rounded values obtained from those for which the corresponding frequency levels or band numbers are integers.

BSR S1.8-1989 (R 200x), Reference Quantities for Acoustical Levels [reaffirmation of ANSI S1.8-1989 (R 2001)]

This standard includes reference quantities for commonly used levels in acoustics, electroacoustics, and mechanical vibration. Acoustical levels of various kinds are commonly used to describe acoustical measurements in gases, liquids, and solids. A reference quantity, preferably independent of the medium, is needed for each kind of level. The preferred unit for an acoustical level is the decibel.

BSR S1.9-1996 (R200x), Instruments for the Measurement of Sound Intensity [reaffirmation of ANSI S1.9-1996 (R 2001)]

Specifies the requirements for instruments to measure sound intensity employing the two-microphone technique and methods for performance verification to meet the requirements. Its primary application is to instruments used for the determination of sound power of sources, in accordance with the requirements of ANSI S12.12-1992. The requirements and methods of performance verification are specified for the complete instrument system and separately for the probes and processors forming the complete system.

BSR S1.15, Part 1-1997 (R200x), Measurement Microphones—Part 1: Specifications for Laboratory Standard Microphones [reaffirmation of ANSI S1.15, Part 1-1997 (R 2001)]

This Standard specifies mechanical dimensions and certain electroacoustical characteristics for capacitor (condenser) microphones used as laboratory standards for sound pressure measurements of the highest attainable accuracy. The specifications are intended to ensure that primary calibration by

 0001-4966/2006/119(5)/2549/4/$22.50

the reciprocity method can be readily carried out. This Standard establishes a system to classify laboratory standard microphones into a number of types according to their dimensions and properties.

BSR S1.42-2001 (R200x), Design Response of Weighting Networks for Acoustical Measurements (reaffirmation of ANSI S1.42-2001)

This Standard provides the design criteria for both the frequency-domain response (amplitude and phase) and time-domain of the *A*- and *C*-weighting networks used in acoustically related measurements. The poles and zeros for each weighting network are given, along with equations for computing the amplitude and phase responses as functions of frequency and impulse and step responses as functions of time. Other known weighting networks that had been standardized, such as the *B*-, *D*-, and *E*-weightings, or weightings that were published in the past, are listed in the Annexes for reference.

ASA (ASC S2) (Acoustical Society of America)

REAFFIRMATIONS

BSR S2.2-1959 (R200x), Methods for the Calibration of Shock and Vibration Pickups [reaffirmation of ANSI S2.2-1959 (R 2001)]

This standard is designed to acquaint the user with the general principles of calibration of shock and vibration pickups and to describe concisely several standard methods that have proven to give reliable and reproducible results. Further details concerning these methods are given in the Appendix. Also, other methods that have not as yet reached the stage of development of the standard methods are described briefly in the Appendix. (20 March 2006)

BSR S2.8-1972 (R200x), Guide for Describing the Characteristics of Resilient Mountings [reaffirmation of ANSI S2.8-1972 (R 2001)]

This standard sets forth suggestions as to subject matter and format for describing resilient mountings, so that there will be a clear understanding by both the user and the manufacturer. It is beyond the scope of this standard to present characteristics of resilient mountings.

BSR S2.9-1976 (R200x), Nomenclature for Specifying Damping Properties of Materials [reaffirmation of ANSI S2.9-1976 (R 2001)]

This standard presents the preferred nomenclature (parameters, symbols, and definitions) for specifying the damping properties of uniform materials and uniform specimens, where "uniform" implies homogeneity on a macroscopic scale.

BSR S2.20-1983 (R200x), Estimating Air Blast Characteristics for Single Point Explosions in Air, with a Guide to Evaluation of Atmospheric Propagation and Effects [reaffirmation of ANSI S2.20-1983 (R 2001)]

Provides qualitative definitions of explosion characteristics for a single-point explosion in air, along with methodologies for scaling these characteristics for a wide range of yield and ambient air conditions. Factors for use with common solid explosives are also included. Methods are provided for predictions of long-range propagation under atmospheric refractive influences. Target damage estimation procedures are provided for use in explosion operation planning and evaluation.

BSR S2.24-2001 (R200x), Graphical Presentation of the Complex Modulus of Viscoelastic Materials [reaffirmation of ANSI S2.24-2001]

This Standard specifies the procedure for generating a graphical presentation of the frequency and temperature dependence of the complex modulus of viscoelastic materials (13 March 2006).

BSR S2.71-1983 (R200x), Guide to the Evaluation of Human Exposure to Vibration in Buildings [reaffirmation and redesignation of ANSI S3.29-1983 (R 2001)]

Assesses reactions of humans to vibrations of 1 to 80 Hz inside buildings by use of degrees of perception and associated vibration levels and durations. Accelerations or velocities inside buildings may be measured to assess perceptibility and possible adverse reactions from those inside. (20 March 2006)

WITHDRAWALS

ANSI S2.13-Part 1-1996 (R2001), Mechanical Vibration of Non-Reciprocating Machines—Measurements on Rotating Shafts and Evaluation—Part 1: General Guidelines [withdrawal of ANSI S2.13-Part 1-1996 (R 2001)]

This Standard provides the test procedure for the measurement and evaluation of the mechanical vibration of nonreciprocating machines, as measured on rotating shafts. The standard also provides guidelines for adapting evaluation criteria for different types of machines.

ANSI S2.41-1985 (R2001), Mechanical Vibration of Large Rotating Machines with Speed Range from 10 to 200 rev/s—Measurement and Evaluation of Vibration Severity in situ [withdrawal of ANSI S2.41-1985 (R 2001)]

Focuses on the measurement and evaluation of vibration severity of large rotating machinery in situ. It is not applicable to reciprocating machinery. The values recommended as limits in this standard are intended to serve as standard values for machines of similar type, when measured in accordance with the procedures describes herein. The recommended vibration limits may be used for acceptance standards or for monitoring the performance of the machine during service operations.

ASA (ASC S12) (Acoustical Society of America)

REVISIONS

BSR S12.65-200x, Rating Noise with Respect to Speech Interference [revision and redesignation of ANSI S3.14-1977 (R 1997)]

This standard defines a simple numerical method for rating the expected speech-interfering aspects of noise using acoustical measurements of the noise.

ASABE (American Society of Agricultural and Biological Engineers)

NEW NATIONAL ADOPTIONS

BSR/ASABE/ISO 5007-200x, Agricultural wheeled tractors—Operator's seat—Laboratory measurement of transmitted vibration (identical national adoption)

Specifies a laboratory method for measuring and evaluating the effectiveness of the suspension of operator seats on agricultural wheeled tractors. (13 March 2006)

BSR/ASABE/ISO 5008-2002 W/Cor. 1-200x Agricultural wheeled tractors and field machinery—Measurement of whole-body vibration of the operator (identical national adoption).

Specifies methods for measuring and reporting the whole body vibration to which the operator of an agricultural wheeled tractor or other field machine is exposed when operating on a standard test track.

ATIS (Alliance for Telecommunications Industry Solutions)

NEW STANDARDS

BSR ATIS 0600005-200x, Acoustic Measurement

This standard identifies sound power as the preferred method of measuring the emission of acoustic noise from telecommunications equipment. The main focus is to use sound power to gain repeatability and accuracy over sound pressure methods. This standard will also provide the emission limits for the temperature-controlled environment (i.e., central office, data centers) for the North American environment. (Comment Deadline: April 10, 2006)

ARI (Air-Conditioning and Refrigeration Institute)

NEW STANDARDS

BSR/ARI 530-200x, Rating of Sound and Vibration for Refrigerant Compressors

This standard applies to external-drive, hermetic and semihermetic refrigerant compressors. In the case of external-drive refrigerant compressors, the driving mechanism shall be excluded from the sound and vibration mea-

surements. However, for semihermetic refrigerant compressors, where the driving mechanism is an integral part of the compressor assembly, it shall be included in the measurements. (25 April 2006)

IEEE (ASC C63) (Institute of Electrical and Electronics Engineers)

REVISIONS

BSR C63.19-200x, Methods of Measurement of Compatibility between Wireless Communications Devices and Hearing Aids (revision of ANSI C63.19-2001)

This standard sets forth uniform methods of measurement and parametric requirements for the electromagnetic and operational compatibility and accessibility of hearing aids used with wireless communications devices operating in the range of 800 MHz to 3 GHz. However, this version is focused on existing services, which are in common use.

Project Initiation Notification System (PINS)

ANSI Procedures require notification of ANSI by ANSI-accredited standards developers (ASD) of the initiation and scope of activities expected to result in new or revised American National Standards (ANS). Early notification of activity intended to reaffirm or withdraw an ANS and in some instances a PINS related to a national adoption is optional. The mechanism by which such notification is given is referred to as the PINS process. For additional information, see clause 2.4 of the ANSI Essential Requirements: Due Process Requirements for American National Standards. Following is a list of proposed actions and new ANS that have been received recently from ASDs.

ASA (ASC S3) (Acoustical Society of America)

BSR S3.20-200x, Bioacoustical Terminology (revision of ANSI S3.20-1995 (R 2003))

This standard provides definitions for a wide variety of terms used in human bioacoustics, including hearing, speech, psychoacoustics, and physiological acoustics. It is intended to supplement ANSI S1.1, American National Standard for Acoustical Terminology, in which more-generally-used terms in acoustics are defined, including a number of terms from physiological and psychological acoustics and music.

ARI (Air-Conditioning and Refrigeration Institute)

BSR/ARI 370-200x, Sound Rating of Large Outdoor Refrigerating and Air-Conditioning Equipment (new standard)

This standard applies to the outdoor portions of factory-made commercial and industrial Large Outdoor Refrigerating and Air-Conditioning Equipment, including heat pumps, used for refrigerating or air-conditioning of spaces.

BSR/ARI 530-200x, Rating of Sound and Vibration for Refrigerant Compressors (new standard)

This standard applies to External-drive, Hermetic and Semi-Hermetic Refrigerant Compressors. In the case of External-drive Refrigerant Compressors, the driving mechanism shall be excluded from the sound and vibration measurements. However, for Semi-Hermetic Refrigerant Compressors, where the driving mechanism is an integral part of the compressor assembly, it shall be included in the measurements.

SMACNA (Sheet Metal and Air-Conditioning Contractors' National Association)

BSR/SMACNA 007-200x, Residential Comfort System Installation Standards Manual (new standard)

The document will provide installation standards for residential heating, ventilating and, air conditioning (HVAC) systems. It will include the most current mechanical and control technology so contractors and designers can design, construct, and install all HVAC systems, from the simplest to the state-of-the-art. Forced-air heating, heat pumps, automatic controls and thermostats, flues, vents, sound and vibration, air cleaning, and other subjects and technologies appropriate for this new century will be included.

IEEE (Institute of Electrical and Electronics Engineers)

BSR/IEEE C57.151-200x, Sound Level Measurement Guide for Liquid Filled and Dry Type

Transformers and Reactors (new standard)

This user guide provides supporting information to help both manufacturers and purchasers apply the measurement techniques described in IEEE C57.12.90 and IEEE C57.12.91. The sources and characteristics of transformer and reactor sound are described. Practical guidance on making measurements is given, and factors that may influence the accuracy of the methods are discussed.

Final actions on American National Standards

The standards actions listed below have been approved by the ANSI Board of Standards Review (BSR) or by an ANSI-Audited Designator, as applicable.

ASA (ASC S12) (Acoustical Society of America)

REAFFIRMATIONS

ANSI S12.7-1986 (R 2006) Measurements of Impulse Noise [reaffirmation of ANSI S12.7-1986 (R 1998)]. (30 January 2006)

AHAM (Association of Home Appliance Manufacturers)

NEW STANDARDS

ANSI/AHAM AC-2-2006, Method for Sound Testing of Portable Household Electric Room Air Cleaners (new standard). (5 January 2006)

ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.)

REAFFIRMATIONS

ANSI/ASHRAE 87.3P-2001 (R 2006) Method of Testing Propeller Fan Vibration (reaffirmation of ANSI/ASHRAE 87.3P-2001). (23 January 2006)

IEEE (Institute of Electrical and Electronics Engineers)

NEW STANDARDS

ANSI/IEEE C57.136-2000 (R 2005), Guide for Sound Level Abatement and Determination for Liquid-Immersed Power Transformers and Shunt Reactors Rated Over 500 kVA (reaffirmation of ANSI/IEEE C57.136-2000). (29 December 2005)

Standards News from Abroad

(Partially derived from *ANSI Reporter* and *ANSI Standards Action*, with appreciation)

Newly Published ISO and IEC Standards

Listed here are new and revised standards recently approved and promulgated by ISO—the International Organization for Standardization.

ISO Standards

ROLLING BEARINGS (TC 4)

ISO 15242-3:2006, Rolling bearings—Measuring methods for vibration—Part 3: Radial spherical and tapered roller bearings with cylindrical bore and outside surface

SHIPS AND MARINE TECHNOLOGY (TC 8)

ISO 9875/Cor1:2006, Shipbuilding—Marine echo-sounding equipment—Corrigendum

ISO Draft Standards

ACOUSTICS (TC 43)

ISO/DIS 3741, Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Precision methods for reverberation test rooms. (24 May 2006)

MECHANICAL VIBRATION AND SHOCK (TC 108)

ISO/DIS 18436-3, Condition monitoring and diagnostics of machines—Requirements for training and certification of personnel—Part 3: Requirements for training bodies and the training process. (28 April 2006)

ISO/DIS 18436-6, Condition monitoring and diagnostics of machines—Requirements for training and certification of personnel—Part 6: Acoustic emission. (25 May 2006)

IEC Standards

PERFORMANCE OF HOUSEHOLD ELECTRICAL APPLIANCES (TC 59)

IEC 60704-3 Ed. 2.0 b: 2006, Household and similar electrical appliances—Test code for the determination of airborne acoustical noise—Part 3: Procedure for determining and verifying declared noise emission values

IEC 60704-2-13 Amd.1 Ed. 1.0 en: 2005, Amendment 1—Household and similar electrical appliances—Test code for the determination of airborne acoustical noise—Part 2-13: Particular requirements for range hoods

AUDIO, VIDEO AND MULTIMEDIA SYSTEMS AND EQUIPMENT (TC 100)

IEC/PAS 62458 Ed. 1.0 en: 2006, Sound system equipment—Electroacoustical transducers—Measurement of large signal parameters

IEC/PAS 62459 Ed. 1.0 en: 2006, Sound system equipment—Electroacoustical transducers—Dynamic measurement of suspension parts

U.S. TAG	ISO and IEC documents
S1	**IEC/CD 60645-3 (29/595/CDV)** "Electroacoustics—Audiometric equipment—Part 3: Test signals of short duration" (16 June 2006)
S12	**ISO/DIS 3741**, Acoustics—Determination of sound power levels and sound energy levels of noise sources using sound pressure—Precision methods for reverberation test rooms. (30 June 2006)
	Second ISO/CD 11201 (N1638), Acoustics—Noise emitted by machinery and equipment—Measurement of emission sound pressure levels at a work station and at other specified positions—Precision and engineering methods in an essentially free field over a reflecting plane without applying environmental corrections (Revision of ISO 11201:1995). (24 April 2006)
	Second ISO/CD 11202 (N1639), Acoustics—Noise emitted by machinery and equipment—Measurement of emission sound pressure levels at a work station and at other specified positions—Engineering and survey methods applying approximate environmental corrections (Revision of ISO 11202:1995). (24 April 2006)
	Second ISO/CD 11204 (N1640), Acoustics—Noise emitted by machinery and equipment—Measurement of emission sound pressure levels at a work station and at other specified positions—Engineering and survey methods applying accurate environmental corrections (Revision of ISO 11204:1995). (24 April 2006)
	Third ISO/DTR 25417 (N1637), Acoustics—Definitions of basic quantities and terms. (28 April 2006)

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).]*

Model-Based Signal Processing

James V. Candy

Wiley-IEEE Press, New York, 2006, 677 pp. $125.00 hardcover. ISBN 0471236322

The use of models in signal processing already plays a major role in the area of inverse problems, in which the goal is to extract information concerning the medium or certain of its contents. In such problems, the model plays the role of the structure in which the unknown parameters lie. Seismics is a good example of this. However, using models to improve the detection or estimation performance of a processor is a fairly new idea. In the acoustics world, its origins go back to people like Parvelescu, Clay, Hinich, and Bucker, whose ideas led to what is now called matched-field processing. Their ideas made clear that the inclusion of physical models in a processing scheme can be an effective means of introducing *a priori* information and thereby improving performance.

As this concept was developed, it soon became clear that there was a down side. Unless the models (or their parameters) were sufficiently well known, the method would perform poorly or, indeed, fail completely. In the underwater acoustics world, this became known as the "mismatch problem." Methods of dealing with this usually entailed either desensitizing the processor or trying to include some of the troublesome parameters themselves as jointly estimated "nuisance" parameters. In the theory developed in this book, the mismatch problem is seen to be simply a consequence of poor processor design, and, in the case where the model itself is insufficient for the task, it is obvious to the designer.

At some point in the development of any useful discipline, someone usually comes along and organizes things into a coherent and self-consistent structure. This is precisely what James Candy has done for the field of model-based processing. The book is organized in a hierarchical structure. After some necessary introductory material, a thorough review of estimation theory is given. This is followed by the development of parametric signal processing methods, such as time-series analysis and its generalizations. Next is the development of state-space processors, which naturally focuses on the Kalman filter, and it is shown how the parametric processors fit into this scheme. After developing the linear Kalman filter, the nonlinear and adaptive cases are studied. Although, strictly speaking, parametric processors are model based, the models do not necessarily represent the physics in a transparent sense. Thus, the book treats physics-based models as a distinct class of problems.

The Kalman filter is a state-space formulation of a multidimensional Gauss-Markov process and thereby forms a natural basis in which physics-based models can be included. This follows from the fact that the state-space-based equations of the Kalman filter form a natural basis for including the differential equations of the relevant physics. The result is a formulation that can be recursive in either space or time, is adaptive, and can be "augmented" by including the nuisance parameters that were the bane of matched-field processors. This means that such processors provide the basis for self-correcting models. This capability is due, in no small measure, to the fact that the Kalman filter provides a direct measure of the "goodness of fit" of a model via the innovation sequence, one of the natural outputs of the processor.

After this development, the author takes us through a broad spectrum of applications. These include reentry vehicle tracking, laser ultrasonics, towed-array processing, internal waves, acoustic source localization, oceanic inversion, and nondestructive evaluation. The author's experience in these fields lends a great deal to the value of this text. Such a comprehensive emphasis on the reduction to practice of the theory is a critical factor found lacking in many signal processing works.

Besides providing a formal basis for model-based processing, the book serves another purpose. The Kalman filter, in my opinion, is not being used to its full potential in the signal processing world. This is partly due to the fact that it is often viewed simply as a control theory tool and, consequently, many signal processing professionals are not terribly conversant with it, and therefore are not aware of its power beyond the control theory world. To put it succinctly, it is a generalized, multi-channel, recursive, minimum-variance estimator into which the user can insert essentially any model he desires. The modern signal-processing professional ignores it at his own loss.

It should be mentioned that the Kalman filter is strictly optimal only in the linear Gaussian case. The nonlinear case, which usually entails non-Gaussian statistics, is dealt with by the so-called extended Kalman filter (EKF) in which the nonlinearities are treated by a Taylor series approximation. Although the EKF works well and is widely applied with great success, new research in this area has been focusing on the *unscented* Kalman filter, which avoids the linearization approach used by the EKF, and the author gives an introduction to this field.

The book is aimed at the graduate student, and therefore it will help if the reader is already somewhat conversant in modern digital signal processing, estimation theory, time-series analysis, and spectral estimation. For those who do not quite fulfill these requirements, the book is still accessible, since it contains a comprehensive set of problems for each chapter, and appendices on probability and statistics and the algorithmic mechanics of sequential model-based processing. In my judgment, it belongs in the library of every practicing signal processor.

EDMUND J. SULLIVAN
EJS Consultants
46 Lawton Brook Lane
Portsmouth, Rhode Island 02871

 0001-4966/2006/119(5)/2553/1/$22.50

OBITUARY

Peter Nielsen Ladefoged ● 1925–2006

Peter Nielsen Ladefoged, a Fellow of this Society and Society Silver Medal recipient, passed away unexpectedly in London on January 24, 2006 due to a stroke. Professor Ladefoged was returning from a fieldwork trip to India where he was studying Toda, an endangered language with few remaining speakers. This was the last of many fieldwork trips Ladefoged conducted with an array of colleagues, including stays in Nigeria, Botswana, Ghana, Uganda, Tanzania, Sierra Leone, Senegal, India, Yemen, Papua New Guinea, Nepal, Kenya, Korea, South Africa, Brazil, Mexico, Australia, and China. At the time of his death, Ladefoged was Distinguished Professor of Phonetics Emeritus and Research Linguist at University of California, Los Angeles, as well as Leverhulme Professor at University of Edinburgh (2005) and Adjunct Professor at the University of Southern California.

Professor Ladefoged is considered the father of the field of linguistic phonetics, which integrates phonetic sciences with linguistic theory. His seminal textbook used throughout the field, *A Course in Phonetics*, is now in its 5th edition (1st edition, Harcourt Brace 1982; 5th edition, Thomson Wadsworth 2006). Professor Ladefoged emphasized the importance of considering the full diversity of human speech sounds—their articulatory, acoustic, and perceptual properties—in seeking a linguistic characterization of the sounds of language. In particular, he melded field and experimental data collection with pioneering analytical techniques. The volume *The Sounds of the World's Languages*, coauthored with Ian Maddieson (Blackwell 1996), stands as a comprehensive account of their research in this area. At the time of his death he was engaged in developing a web archive of diverse speech sounds gathered from scores of languages in over 50 years of research (archive.phonetics.ucla.edu/ and phonetics.ucla.edu/).

In addition to Ladefoged's contributions investigating and documenting the sounds of the world's languages, his important research on vowel acoustics, perception, and classification began with his Ph.D. thesis and continued throughout his career in collaboration with Donald Broadbent, Walter Lawrence, Richard Harshman, Louis Goldstein, Keith Johnson, and Mona Lindau [Lingua **5.2**, 113–128 (1956); Nature **178**, 815–816 (1956); J. Acoust. Soc. Am. **29**, 98–104 (1957); J. Acoust. Soc. Am. **62**, 693–707 (1977); J. Acoust. Soc. Am. **94**, 701–714 (1993)]. He demonstrated that vowels can be well classified by their auditory properties and advocated that auditory properties, in addition to articulatory ones, are relevant in the categorization of speech sounds.

Ladefoged valued instrumental techniques in the field and in the laboratory, and developed a long-standing interest in seeing that they were appropriately applied outside these arenas as well. He was particularly concerned that spectrographic analysis was being abused in the field of forensic "voiceprinting," and was concerned about factors affecting the reliability of earwitness testimony [Language and Speech **21**, 373–374 (1978); Echoes **14**, 2, 14–15 (2004)]. His scholarly efforts and public advocacy in these areas introduced a measure of healthy skepticism regarding expert and naïve trial testimony.

Ladefoged's seminal contributions led to numerous prestigious awards. During his career, Professor Ladefoged was elected Fellow of the Acoustical Society of America, of the American Speech and Hearing Association, and of the American Academy of Arts and Sciences. He served as President of the Linguistic Society of America (1978) and of the International Phonetic Association (1985), receiving the Gold Medal from the International Congress of Phonetic Sciences in 1991 and the Silver Medal from the Acoustical Society of America in 1994 "for advancing the theory of acoustic phonetics and phonology using acoustic field data from many of the world's languages" [J. Acoust. Soc. Am. **96**(5), 2 (1994)]. He received honorary doctoral degrees from the University of Edinburgh (D.Litt., 1993) and Queen Margaret University (D.Sc., 2002).

Professor Ladefoged was born in Sutton, England in 1925. After his war service from 1944-1947, he received his Ph.D. from University of Edinburgh in 1959, where he studied with David Abercrombie. He continued research on vowel perception with Donald Broadbent and on respiration with David Whitteridge. In 1953, Ladefoged married Jenny Ladefoged (née MacDonald), who became his partner and collaborator throughout his life. Ladefoged joined the faculty of the University of California, Los Angeles in 1962 and founded the UCLA Phonetics Laboratory shortly thereafter. Among his more than 20 Ph.D. students were Victoria Fromkin, John Ohala, Ian Maddieson, Louis Goldstein, and Catherine Browman. He retired in 1991 but remained active in both teaching and research until his death at the age of 80. More details regarding Professor Ladefoged's impact as a teacher and a phonetician can be found at www.linguistics.ucla.edu/people/ladefoge/remember/index.htm. A recognition by the field of his recent 80th birthday, celebrated with a special session at the 150th Meeting of the Acoustical Society [J. Acoust. Soc. Am. **118**(3), 2 (2005)], can be found at sail.usc.edu/peterladefoged80.html. In addition to his wife Jenny, Peter Ladefoged is survived by his three children and his grandchildren.

DANI BYRD
JODY KREIMAN

 0001-4966/2006/119(5)/2555/1/$22.50

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

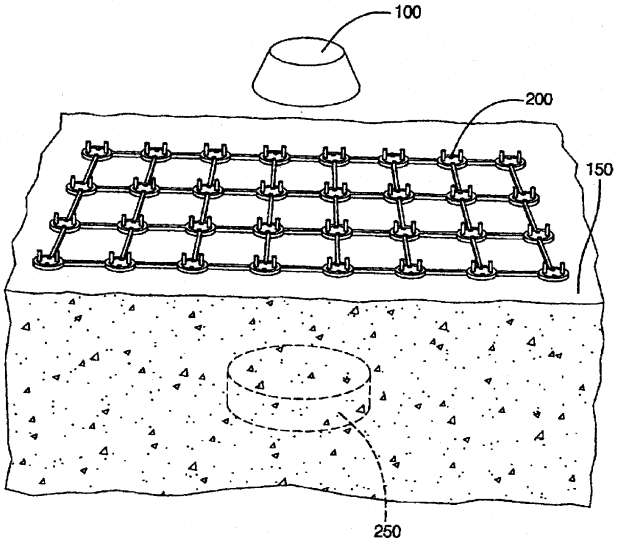
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
DIMITRI DONSKOY, *Stevens Institute of Technology, Castle Point on the Hudson, Hoboken, New Jersey 07030*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
JOHN ERDREICH, *Ostergaard Acoustical Associates, 200 Executive Drive, West Orange, New Jersey 07052*
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*
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
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, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,862,252

43.20.Gp METHOD AND APPARATUS FOR ACOUSTIC DETECTION OF BURIED OBJECTS

Robert Hickling, Huntington Woods, Michigan
1 March 2005 (Class 367/88); filed 9 September 2003

Detection of buried land mines is accomplished using an array of ''acoustic vector probes'' **200** placed near the surface of the ground **150**. Each probe consists of a tetrahedral arrangement of four microphones allowing for the measurement of sound velocity vector and sound pressure.



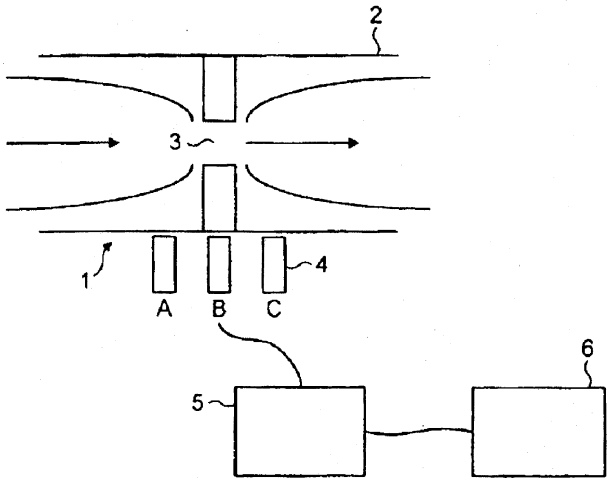
An impulse acoustic source **100** above the array broadcasts 0.05–0.2-ms sound pulses. The detection is based on multichannel processing of the signals echoed from the mine **250**, computing a sound-intensity vector at each probe in the array.—DMD

6,945,094

43.20.Ye VISCOSITY MEASUREMENT

Svein Eggen *et al.*, assignors to Borealis Technology Oy
20 September 2005 (Class 73/54.41); filed in the United Kingdom 22 December 2000

The rheological characterization of a polymer when it leaves the polymerization reactor is accomplished by analysis of the acoustic emissions of the melt when it flows through apparatus **1** consisting of conduit **2** and



constriction **3**. The accelerometers **4** detect the acoustic emissions. Signals from **4** are sent to amplifier/preprocessor **5** and analysis is carried out by computer **6**.—NAS

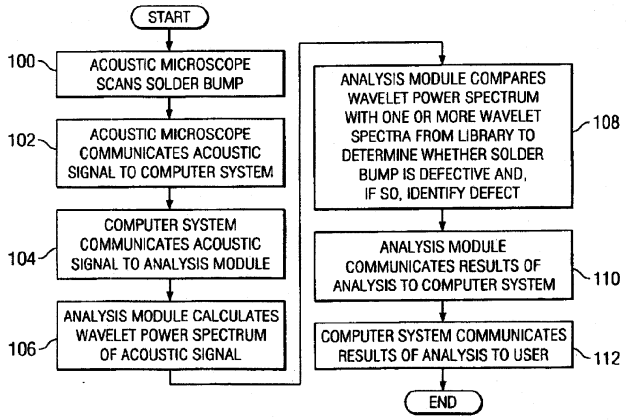
 0001-4966/2006/119(5)/2557/15/$22.50

6,950,761

43.20.Ye WAVELET ANALYSIS OF ONE OR MORE ACOUSTIC SIGNALS TO IDENTIFY ONE OR MORE ANOMALIES IN AN OBJECT

Kartik Ramanujachar, assignor to Texas Instruments Incorporated
27 September 2005 (Class 702/48); filed 31 December 2003

The wavelet analysis of a signal or signals from an acoustic scan of an object is compared to a reference library to determine if the object is known or unknown. The reference library contains both good and bad scans. In the



instant case, this technique is applied specifically to solder bumps in integrated circuits.—NAS

6,954,175

43.30.Xm ACCURATE POSITIONING OF DEVICES SUSPENDED UNDERWATER FROM A FLOATING BODY

Wayne H. Cox, assignor to Input/Output, Incorporated
11 October 2005 (Class 342/357.07); filed 2 November 2004

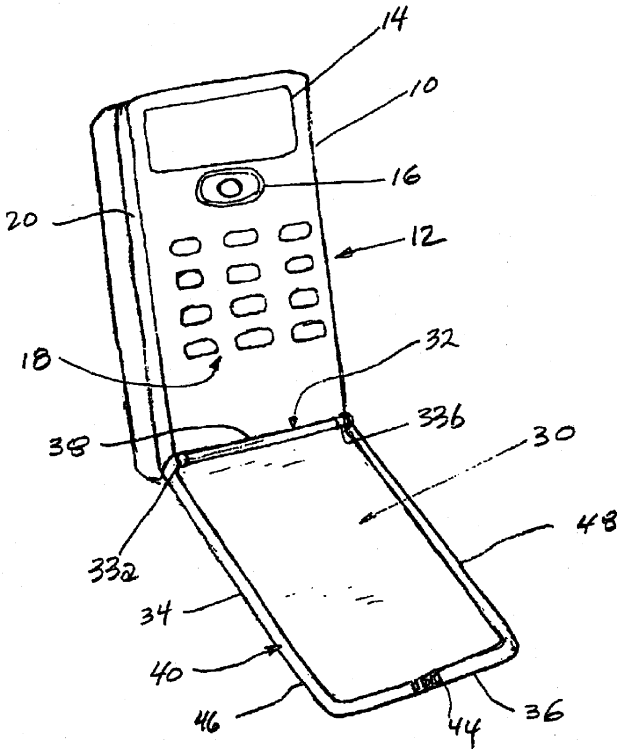
A system is described for accurately determining the position of a sonar device attached to the bottom of a floating buoy. The sonar device is mounted to the buoy collinearly with a pair of GPS antennas, or other antennas, on the top side of the buoy. The position of the sonar device can be computed from knowledge of the range of the two antennas relative to the source of the EM wave that excites them, the known distance between these two antennas, and the known distance between one of the antennas and the sonar device.—WT

6,965,790

43.38.Ar ACOUSTIC PIPE FOR FLIP STYLE CELLULAR TELEPHONE

Gregory S. Mendolia, assignor to Ericsson Incorporated
15 November 2005 (Class 455/575.3); filed 23 June 1997

In conventional cell phone design there is a small electret microphone located in the fold-down ''flap'' portion of the hand-held device. This requires electrical wiring from the electret element back to the main chassis of the system. The patent proposes an ''acoustical pipe'' for this purpose. The



microphone element is located in the main assembly and the acoustical path works its way through the hinge down to the bottom of the flap. The path length is probably no more than 4 or 5 in. and, with reasonable care, should cause no degradation of the voice signal. Manufacturing advantages are obvious.—JME

6,967,430

43.38.Dv FLAT RESONATING ELECTROMECHANICAL DRIVE UNIT

Stefan Johansson, assignor to Piezomotor Uppsala AB
22 November 2005 (Class 310/328); filed 1 October 2003

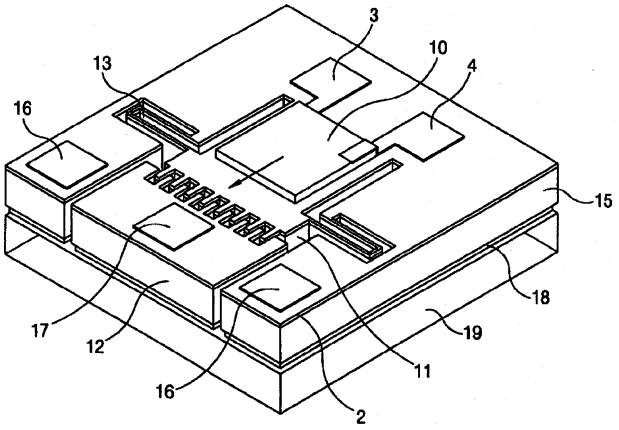
This patent describes an actuator suited to the control of short-throw, high-precision positioning devices for which incremental, not absolute, positioning is adequate. The mode of operation is that of intermittent scrubbing contact between the driver and the driven stage. The drawings are descriptive, but the means by which the requisite hinges could be fabricated without incurring excessive play and losses is not described. There are no dimensions given, although it is suggested that they are all something like a few cm. It is not very well explained in terms of design formulas and considerations, and no performance measurements are given. The patent is not recommended for insights into actuator design.—JAH

6,965,274

43.38.Ew THIN FILM BULK ACOUSTIC RESONATOR FOR CONTROLLING RESONANCE FREQUENCY AND VOLTAGE CONTROLLED OSCILLATOR USING THE SAME

Heon Min Lee and Young Joo Yee, assignors to LG Electronics Incorporated
15 November 2005 (Class 331/107 A); filed 23 June 2003

The present invention concerns a bulk mechanical resonator with spring **13** and mass **12** having an electrically controlled and modifiable spring constant via a comb-shaped gap connected to electrodes **16** and **17**. Utilizing the distance-dependent attraction between the combs when a voltage is applied to them, the mechanical spring constant of the suspension



spring **13** is altered by the existence of a parallel electrical spring. The electronic control is used to create a voltage-tunable oscillator having the capability of feedback control of phase noise. The whole device is conveniently made using surface micromachining techniques.—JAH

6,956,792

43.38.Fx OPENWORK SHELL PROJECTOR

Jason W. Osborn *et al.*, assignors to BAE Systems Information and Electronic Systems Integration Incorporated
18 October 2005 (Class 367/174); filed 28 February 2002

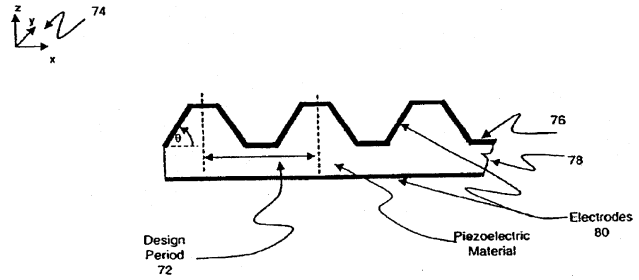
The solid shell of either a flextensional transducer or a slotted-shell cylindrical transducer is replaced by a honeycomb shell. The honeycomb shell provides high specific stiffness with much less weight than a solid shell, thus increasing the transducer's bandwidth, because the honeycomb shell can be thicker, resulting in a greater radiating surface while also reducing total weight and cost.—WT

6,965,189

43.38.Fx BENDING ACTUATORS AND SENSORS CONSTRUCTED FROM SHAPED ACTIVE MATERIALS AND METHODS FOR MAKING THE SAME

Christoph P. Menzel, assignor to Monodrive Incorporated
15 November 2005 (Class 310/330); filed 20 September 2002

This patent describes several different unimorph-style actuators made by nonuniform electroding of piezoelectric materials. The general principle is embodied in the figure, which depicts a slab of piezoelectric material with a corrugated surface. Upon application of a voltage to the top and bottom surfaces, the material will bend due to the nonuniform electric field applied.



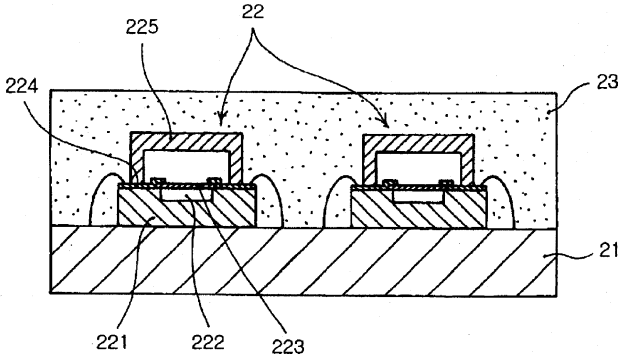
There are several other geometries presented for consideration, but no discussion of the coupling constants or poling details of these configurations is given. Simple electromechanical considerations would suggest that these designs are not as efficient as those created from traditional unimorphs.—JAH

6,965,281

43.38.Fx FBAR, FBAR BASED DUPLEXER DEVICE AND MANUFACTURING METHOD THEREOF

Kook Hyun Sunwoo and Jong Oh Kwon, assignors to Samsung Electro-Mechanics Company, Ltd.
15 November 2005 (Class 333/133); filed in the Republic of Korea 24 September 2003

This patent describes a process for making film bulk acoustic resonators (FBARs) on ordinary PC board or low-temperature cofired ceramic substrates as illustrated by the pair **22**. The objective is to reduce the cost of



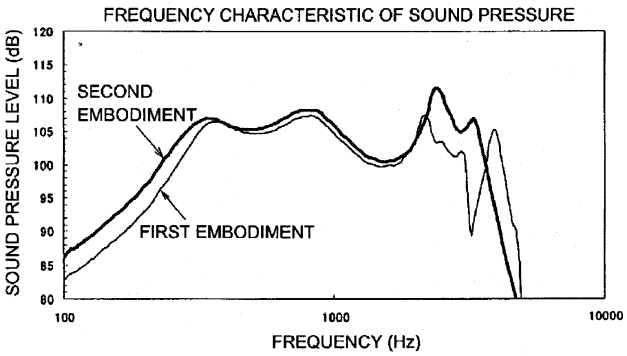
the devices by easing fabrication. However, the performance of such devices is not stated or even hinted at. There is little to be learned from the details.—JAH

6,965,680

43.38.Fx PIEZOELECTRIC ELECTRO-ACOUSTIC TRANSDUCER

Masakazu Yamauchi *et al.*, assignors to Murata Manufacturing Company, Limited
15 November 2005 (Class 381/190); filed in Japan 26 April 2002

The patent first describes the earlier development of a small, multilayer piezoelectric transducer capable of producing high sound levels over a relatively wide frequency range. An improved version is then disclosed which is self-contained and requires no special mounting provisions. Piezoelectric ceramic layers are bonded to the front and back of a resin film.



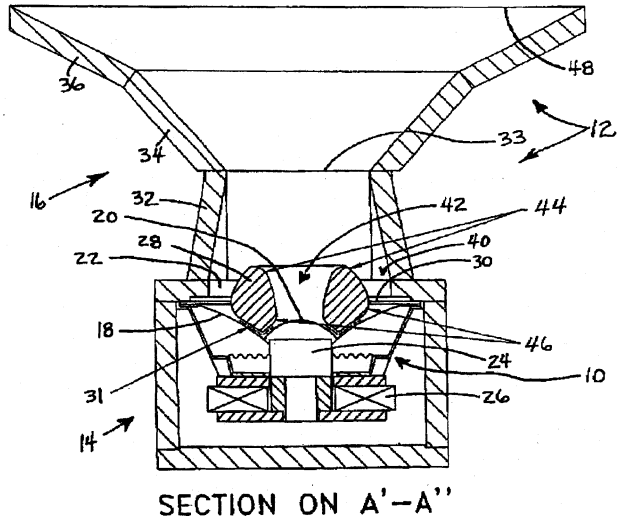
The area of the ceramic vibration elements is 40% to 70% of the area of the film, which is sealed to, and supported by, a rectangular case. The fundamental resonance of the assembly is relatively low, facilitating natural voice reproduction.—GLA

6,950,530

43.38.Ja DIRECTIONAL LOUDSPEAKER UNIT

Jason Baird and William T. Webb, assignors to Martin Audio Limited
27 September 2005 (Class 381/343); filed in the United Kingdom 31 January 2002

This is a mid-frequency, horn-loaded loudspeaker assembly with what is described as a phase-coherent radiating wavefront. Among the items described are a phasing plug **28** of a toroidal (or part-spherical or part-elliptical) shape, a central aperture **42** in the plug **28** which serves princi-



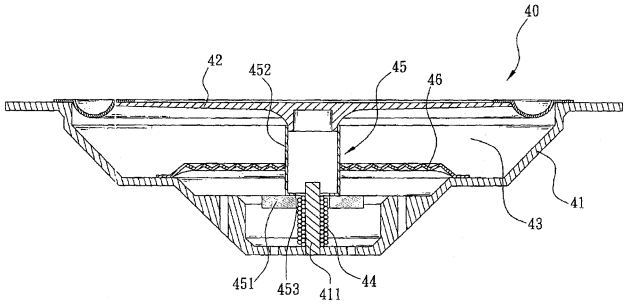
pally as a high-frequency sound path, a second annular path **40** around the plug which is the predominate path for low-frequency sound, and a single driving means for all sounds in the passband.—NAS

6,968,071

43.38.Ja SPEAKER HAVING MAGNETIC MEMBER INSTALLED ON DIAPHRAGM

Chao-Lang Wang, Taipei, Taiwan, Province of China
22 November 2005 (Class 381/423); filed 15 January 2003

The inventor has already patented a moving-magnet loudspeaker in which vaguely defined lines of flux interact with a fixed voice coil (United States Patent 6,415,037 [reviewed in J. Acoust. Soc. Am. **113**(2), 683 (2003)]). This subsequent patent describes the operation of an improved variant in terms that are impossible to decipher: "The primary objective of the present invention is to provide a speaker having a diaphragm disposed around the periphery of a frame on one lateral side of the speaker, such that the diaphragm and the frame enclose a resonant space, and a magnetic



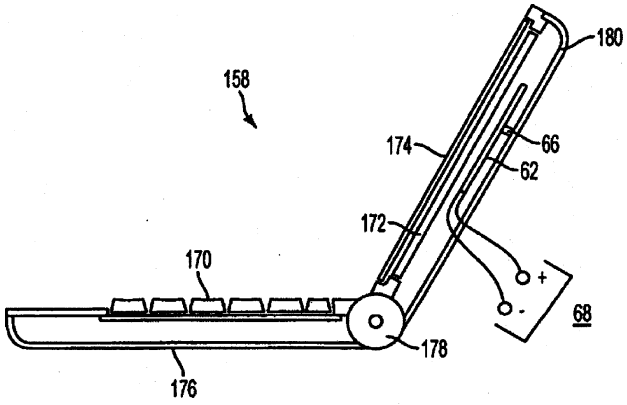
member disposed at the center of the diaphragm and along the direction of the resonant space, thereby when a winding in the middle of the inner side of the frame is excited, the resonant space generates varying magnetic lines, so that an attraction-repulsion effect is generated in the middle of the winding or the reserved space at the position of the periphery to vertically slide the magnetic member and cause the diaphragm to vibrate and generate sound."—GLA

6,965,678

43.38.Ja ELECTRONIC ARTICLE COMPRISING LOUDSPEAKER AND TOUCH PAD

Graham Bank *et al.*, assignors to New Transducers Limited
15 November 2005 (Class 381/152); filed in the United Kingdom 20 July 2001

This patent is concerned with small electronic devices that incorporate touch pads. Although the presence of a touch pad is a requirement of the patent claims, it has no necessary connection with the invention itself. Instead, the patent gives us a detailed analysis of several ways in which mul-



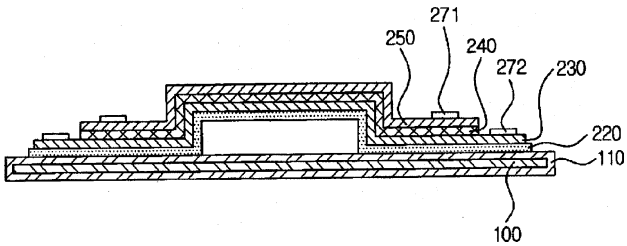
tiresonant, distributed-mode, and/or balanced-mode transduction can be achieved using the outer case of a device as a sound-emitting element. There is a great deal of interesting and valuable information here.—GLA

6,967,362

43.38.Kb FLEXIBLE MEMS TRANSDUCER AND MANUFACTURING METHOD THEREOF, AND FLEXIBLE MEMS WIRELESS MICROPHONE

Yun-woo Nam and Suk-han Lee, assignors to Samsung Electronics Company, Limited
22 November 2005 (Class 257/254); filed in the Republic of Korea 26 September 2002

This patent describes the fabrication of a micromachined microphone said to be flexible and wireless, the motivation for it being to make a wearable information input device. The sensing layer **220** is any of PVDF, Tr-Ef, polyimide, nylon, etc., and is configured as a diaphragm here or, in other implementations, as a cantilever. It all seems miraculous that any such arrangements should create functional microphones without regards to cavity



volume, leaks, materials used for construction, poling direction, and the like. It is probable that none of these devices have been actually made, for there are but one or two dimensions given in the text and no performance measurements. The reader ought to look elsewhere for information on construction of polymer-film microphones.—JAH

6,960,137

43.38.Md SOUND CONTROL METHOD AND DEVICE FOR EXPRESSING GAME PRESENCE

Manabu Nishizawa *et al.*, assignors to Sony Computer Entertainment Incorporated
1 November 2005 (Class 463/35); filed in Japan 19 January 2001

Now Sony is patenting the trivial: In a computer game, divide the sounds into headset and loudspeaker catagories. Now, during playback, you can listen to the headset sounds in one ear and the loudspeaker sounds in the other.—MK

6,965,546

43.38.Md SOUND CRITICAL POINTS RETRIEVING APPARATUS AND METHOD, SOUND REPRODUCING APPARATUS AND SOUND SIGNAL EDITING APPARATUS USING SOUND CRITICAL POINTS RETRIEVING METHOD

Junichi Tagawa and Hiroaki Yamane, assignors to Matsushita Electric Industrial Company, Limited
15 November 2005 (Class 369/30.19); filed in Japan 13 December 2001

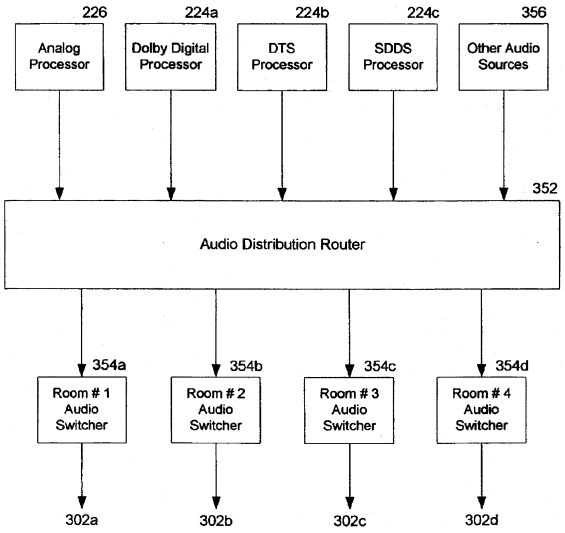
Essentially, the inventor proposes automatically finding "sound critical points" in audio material without listening to the data. These critical points can be used to segment the sound for preview as well as other editing tasks. The features proposed are not given good psychoacoustic justification unless you count, "these properties are determined by the human auditory system and cerebrum." These properties include rms, beat noisiness, and attack ratio.—MK

6,972,829

43.38.Md FILM SOUNDTRACK REVIEWING SYSTEM

Tom McCormick *et al.*, assignors to NT Audio Visual Supply, Incorporated
6 December 2005 (Class 352/40); filed 17 November 2003

This Hollywood-specific invention illustrates how a 35-mm film with multiple audio tracks (e.g., with different spatial encoding methods) can be



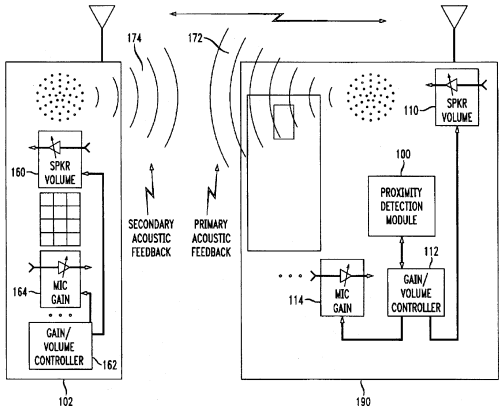
previewed in multiple viewing rooms. As presented, it's reasonable engineering but borders on the obvious.—MK

6,952,471

43.38.Si HANDSET PROXIMITY MUTING

Joseph M. Cannon *et al.*, assignors to Agere Systems Incorporated
4 October 2005 (Class 379/390.01); filed 9 June 2000

The figure "shows a cordless telephone including a proximity detection module in a base unit, and gain/volume controllers in both the base and the handset, for muting or attenuating a microphone and/or speaker path when a simultaneously (sic) hand set and base unit come within close prox-



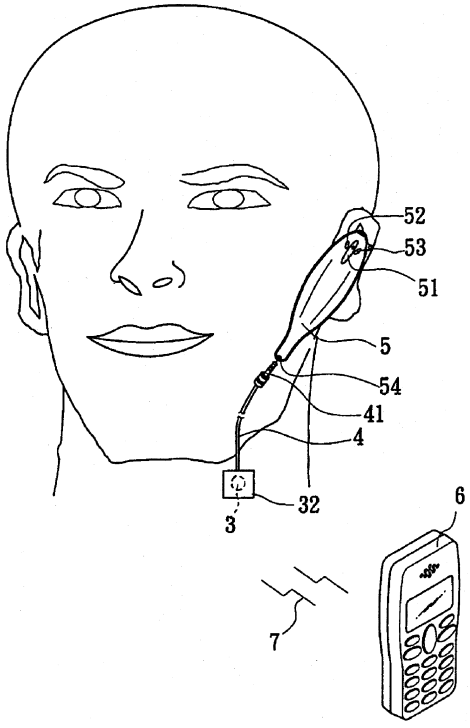
imity to one another." This is done using a receive-strength indicator, a proximity-detection module, and even GPS satellites, among other methods, both explicitly described and implied.—NAS

6,968,066

43.38.Si WIRELESS HANDS-FREE APPARATUS WITH SOUND-ABSORBING FUNCTION

Wen Hsiang Yueh, assignor to Partner Tech Corporation
22 November 2005 (Class 381/151); filed 1 December 2003

This invention is defined throughout the patent as "A wireless and hands-free earphone with a sound-absorbing function." However, it is really a throat-contact microphone held in place by an adhesive patch that also provides an airtight seal. The inventor is confident that this arrangement cuts



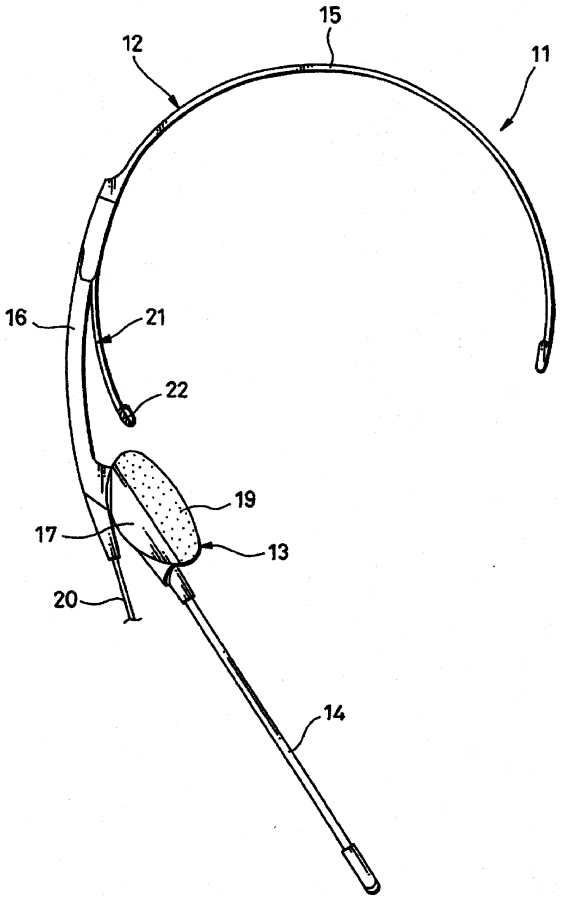
out ambient sound and also resolves the drawbacks of too much bass and lack of high frequencies. Not surprisingly, there is no indication that a working model has actually been tried.—GLA

6,970,572

43.38.Si HEADPHONE DEVICE

Katsunori Murozaki and Koji Nageno, assignors to Sony Corporation
29 November 2005 (Class 381/381); filed in Japan 31 August 2000

This is the latest Sony nonslip headset. It includes a headphone **17** that



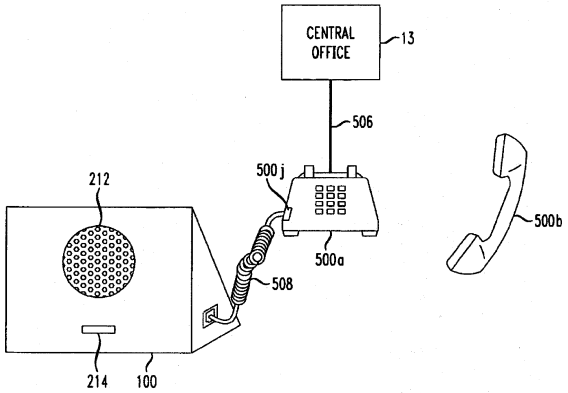
is tilted to match the inclination of the user's ear, and a little clip **22** to keep the assembly from creeping down.—GLA

6,973,179

43.38.Si POCKET SPEAKERPHONE

Steven B. Kaufman and John P. Veschi, assignors to Agere Systems Incorporated
6 December 2005 (Class 379/388.02); filed 11 September 1998

According to this patent, ''There is no conventional technique or apparatus to allow use of the common telephone as a speakerphone. ... There is thus a need for a convenient technique and apparatus for temporarily adapting a conventional common telephone as a speakerphone on an as needed



basis.'' What is proposed is a little box containing a microphone, loudspeaker, and associated electronics that plugs into the telephone's handset jack. This rather modest concept has been expanded into 23 patent claims, none of which require it to be pocket-size.—GLA

6,947,569

43.38.Vk AUDIO SIGNAL PROCESSING DEVICE, INTERFACE CIRCUIT DEVICE FOR ANGULAR VELOCITY SENSOR AND SIGNAL PROCESSING DEVICE

Yuji Yamada and Hirofumi Kurisu, assignors to Sony Corporation
20 September 2005 (Class 381/310); filed 24 July 2001

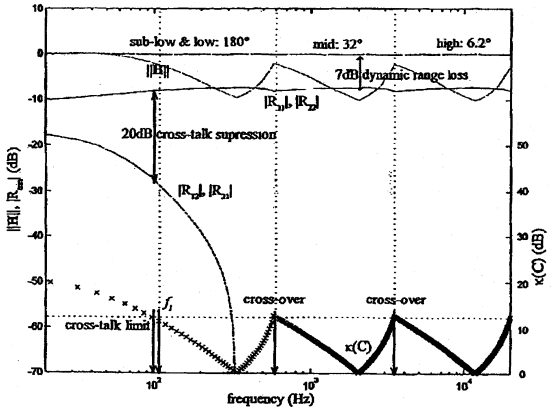
Although giving few references to related devices, this patent acknowledges various prior versions of essentially the same invention. The argument here seems to be for more accurate measurement of the user's head motions as a basis for improved HRTF computations. In various embodiments, alternate sensors, including a vibration gyroscope, a geomagnetic rate sensor, or multiple instances of such sensors, are specified. Details of the actual HRTF computations are not discussed.—DLR

6,950,524

43.38.Vk OPTIMAL SOURCE DISTRIBUTION

Philip Arthur Nelson and Takashi Takeuchi, assignors to Adaptive Audio Limited
27 September 2005 (Class 381/17); filed in the United Kingdom 24 June 2000

This intense (in the best patentese) patent, with 30 figures, describes a solution for multichannel reproduction, which is said to correct a litany of problems with prior art, including loss of dynamic range. There is a detailed



discussion and illustration of some of the patent's various implementations.—NAS

6,973,192

43.38.Vk DYNAMIC ACOUSTIC RENDERING

Alan Gerrard and Nam Do, assignors to Creative Technology, Limited
6 December 2005 (Class 381/17); filed 4 May 1999

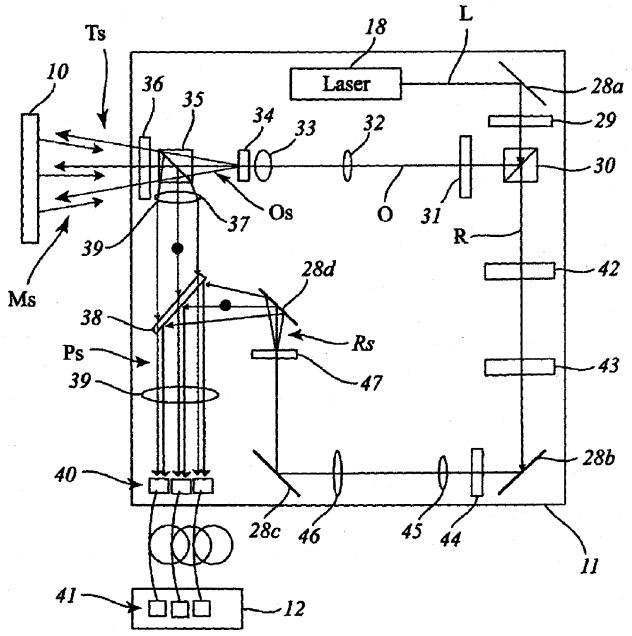
This patent relates to enhanced stereo/binaural audio to accompany on-screen action in video games. Using some of the techniques previously developed for auralization in spaces yet unbuilt, rapid on-screen action can be backed up by a parallel audio scenario in which acoustical effects, bouncing off walls, coming around corners, or simply being reverberated, can be realistically rendered over stereo loudspeakers or via binaural headphones—all in synchronism with what is taking place on-screen.—JME

6,972,846

43.40.Le MULTI-BEAM HETERODYNE LASER DOPPLER VIBROMETER

Amit K. Lal *et al.*, assignors to MetroLaser, Incorporated
6 December 2005 (Class 356/486); filed 31 March 2003

As this patent explains, laser interferometry is a well-established technique for nondestructive measurement and analysis of the vibration of an object. In many applications it is used in conjunction with acoustical elements. For example, to search for buried land mines, an acoustical signal from a loudspeaker can induce vibrations in the ground. The transient vibrations at the surface of the soil can then be detected and mapped to detect the object. Although sequential scanning can be used to gather data from a



relatively large area, an ideal system would measure multiple locations simultaneously. The patent discloses such a simultaneous multibeam system based on heterodyne detection and describes numerous variants. In view of the substantial body of prior art in this field, it is a little surprising that only two previous patents (one cited by the examiner) are referenced, plus one other publication.—GLA

6,971,491

43.40.Tm MAGNETO-RHEOLOGICAL FLUID ENCASED IN FLEXIBLE MATERIALS FOR VIBRATION CONTROL

Faramarz Gordaninejad *et al.*, assignors to The Board of Regents of the University and Community College System of Nevada, on behalf of the University of Nevada, Reno
6 December 2005 (Class 188/267.2); filed 26 December 2002

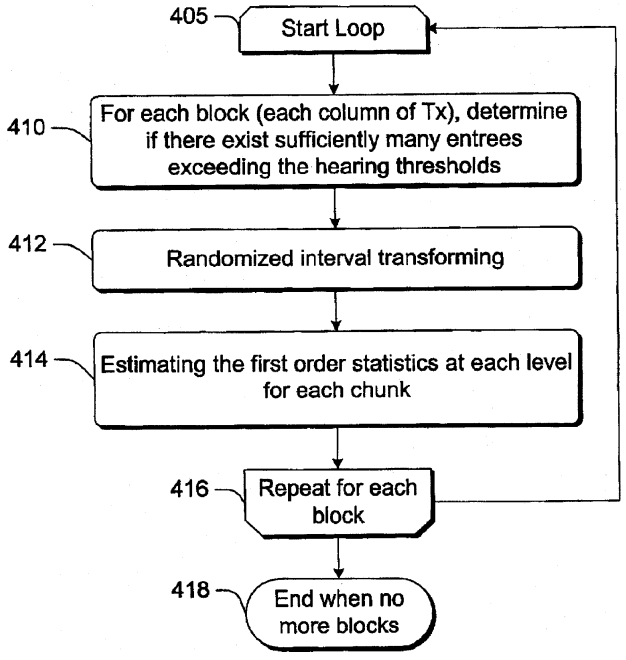
A magneto-rheological fluid is one whose effective viscosity can be varied by applying a magnetic field. In most devices that have used such fluids, the fluid has been contained in pipes or other rigid-walled structures. In devices according to the present patent the magneto-rheological fluid fills the space in a containment element that is flexible and resilient, and ideally also tough and durable. The containment elements may be in the shape of pillows or hoses, for example, and may contain several fluid-filled interior spaces. Illustrative data presented in the patent indicate that application of a magnetic field to such a device results in significant increases in the effective storage and loss moduli in certain frequency ranges, but in reductions of the loss factor (the ratio of the loss modulus to the storage modulus).—EEU

6,973,574

43.60.Cg RECOGNIZER OF AUDIO-CONTENT IN DIGITAL SIGNALS

M. Kivanc Mihcak and Ramarathnam Venkatesan, assignors to Microsoft Corporation
6 December 2005 (Class 713/180); filed 24 April 2001

"Hashing" is a well-known search technique—the hash function is a map from a large domain to a much smaller and fixed range. This audio hash function is computed by (1) transforming a windowed signal into the frequency domain via a lapped transform, (2) estimating the signal statistics



from the frequency-domain blocks, (3) applying an adaptive quantizer, and, lastly, (4) using an error correcting code (in this case Reed-Muller). The text is well written and is a useful introduction to audio content hashing.—MK

6,965,676

43.60.Mn VOLUME-RESPONSIVE LOUDNESS COMPENSATION CIRCUITS, SYSTEMS, AND METHODS

Rustin W. Allred, assignor to Texas Instruments Incorporated
15 November 2005 (Class 381/104); filed 27 June 2000

This patent describes a more complex incarnation of the notorious ''loudness circuit'' found on many audio amplifiers, which is intended to provide a bass boost at low overall listening volumes to compensate for the hearing curve's elimination of apparent bass at low average sound-pressure level. The method here is difficult to follow and seems to refer to previous patents for exposition; the loudness compensation is a gain-dependent variable function that uses the base 2 log of the overall volume, multiplied by a slope function whose details are left to the implementer.—SAF

6,968,063

43.60.Mn DYNAMIC VOLUME ADJUSTMENT IN A SLOT MACHINE

Scott Boyd, assignor to Acres Gaming Incorporated
22 November 2005 (Class 381/57); filed 11 March 2003

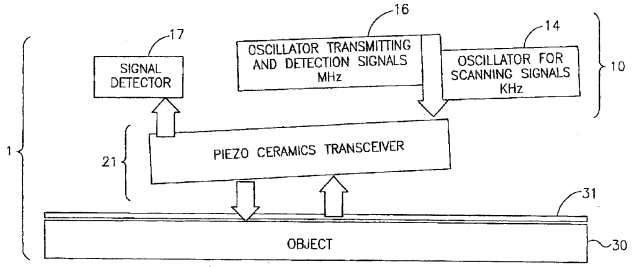
A simple component of a gaming machine is described which adjusts the volume of the machine's sounds in accord with a measurement of ambient noise level.—SAF

6,964,640

43.60.Vx SYSTEM AND METHOD FOR DETECTION OF MOTION

Jona Zumeris *et al.*, assignors to P M G Medica L I D
15 November 2005 (Class 600/459); filed 22 January 2003

The Doppler effect forms the basis for this motion-detection device that consists of two electronic oscillators, one to produce a scanning signal and the other to produce a Doppler signal. A vibrating element vibrates in response to the scanning and Doppler signals and a vibration transducer receives signals reflected from a structure inside the motion detector, converting the reflected signals into electrical signals. The velocity of the structure from which the vibrations are reflected can then be estimated by comparing the frequencies of the transmitted and reflected vibrations. The transducers may be made of piezo-ceramic materials.—DRR

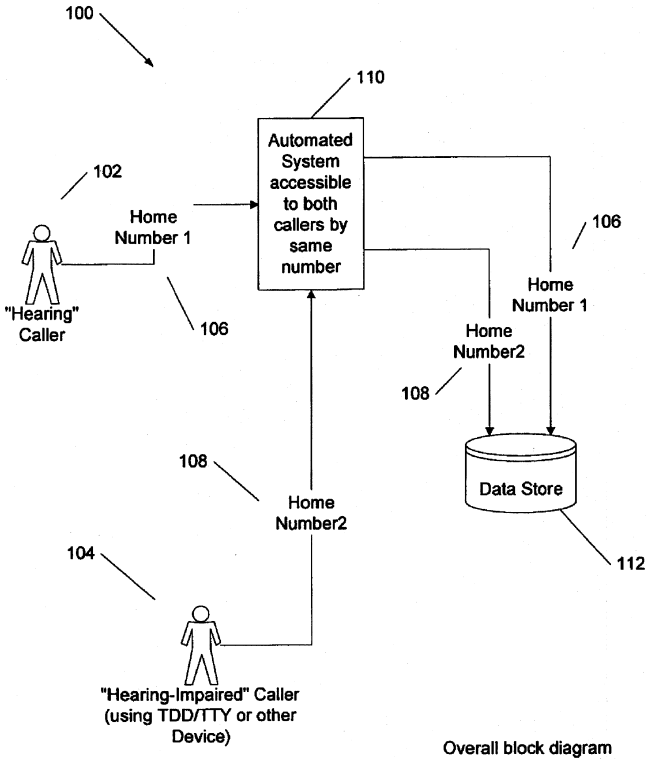


6,965,664

43.66.Ts SYSTEM, METHOD, AND COMPUTER-READABLE MEDIUM FOR ENABLING BOTH HEARING AND HEARING-IMPAIRED CALLERS TO ACCESS A RESOURCE

Jonathan P. McIntosh *et al.*, assignors to West Corporation
15 November 2005 (Class 379/52); filed 25 June 2003

This is a system that enables either a hearing or a nonhearing caller to make a call to either a hearing or nonhearing party through an automated system. The call will be routed to the appropriate platform for a hearing call recipient or for a hearing-impaired recipient who depends on a TDD/TTY or another device to receive messages.—DRR



Overall block diagram

6,968,238

43.66.Ts METHOD FOR INSERTING COCHLEAR ELECTRODE AND INSERTION TOOL FOR USE THEREWITH

Janusz A. Kuzma, Parker, Colorado
22 November 2005 (Class 607/137); filed 5 August 2003

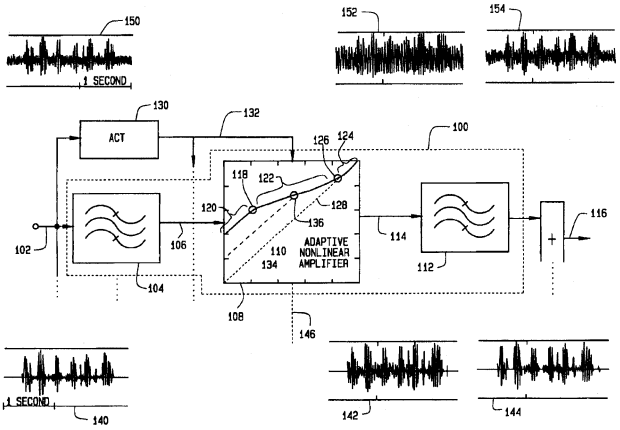
To ensure that an electrode array better conforms to the modiolus during implantation, a stylet wire is inserted into a lumen projection while the array is held within a sheath. Subsequently, the sheath and stylet wire are removed while the array is pushed into the cochlea.—DAP

6,970,570

43.66.Ts HEARING AIDS BASED ON MODELS OF COCHLEAR COMPRESSION USING ADAPTIVE COMPRESSION THRESHOLDS

Julius L. Goldstein, assignor to Hearing Emulations, LLC
29 November 2005 (Class 381/321); filed 23 August 2001

Traditional hearing aid multichannel-compression circuitry is augmented by making the compression threshold (CT) in each channel variable in response to input signal changes. The CT increases or decreases, respectively, if the peak value of the input signal increases or decreases repeatedly



or stays the same if the peak value of the input signal has only minor fluctuations. Increases in CT are instantaneous, whereas decreases in CT are performed slowly.—DAP

6,965,681

43.66.Vt ARRANGEMENT IN ACOUSTIC HEADSETS

Christer Almqvist, assignor to Peltor AB
15 November 2005 (Class 381/371); filed in Sweden 15 August 1997

A radio receiver has been added to a muff-type hearing protector with buttons on the earcup to control the electronics.—JE

6,970,571

43.66.Vt LOW COST HEARING PROTECTION DEVICE

Jon P. Knorr and Benjamin P. Knapp, assignors to Jackson Products, Incorporated
29 November 2005 (Class 381/370); filed 3 February 2003

This ''low-cost'' hearing protector incorporates circuitry to permit gain control of the sound at the speakers inside the earmuff-type headset and to allow configuration changes. Of course, the degree of protection is limited by the earmuffs themselves. One wonders if the cost of the two batteries required to power the unit, let alone the circuits shown in six accompanying figures, are as ''low cost'' as a pair of foam earplugs.—JE

6,959,276

43.72.Dv INCLUDING THE CATEGORY OF ENVIRONMENTAL NOISE WHEN PROCESSING SPEECH SIGNALS

James G. Droppo *et al.*, assignors to Microsoft Corporation
25 October 2005 (Class 704/226); filed 27 September 2001

A problem is identified with conventional adaptive-noise-canceling methods in speech processing systems, namely that their effectiveness is often hampered by unwelcome switching between noise-correction vectors resulting from imperfect matches between training noise and operation noise. The solution proposed here is to incorporate a separate noise-category classifier (e.g., vehicle, airplane) whose decision then governs the action of a conventional noise-canceling system, thereby forcing a possibly locally suboptimal match between training and operational noise, but resulting in a globally improved performance thanks to the elimination of rapid switching between correction vectors.—SAF

6,959,277

43.72.Dv VOICE FEATURE EXTRACTION DEVICE

Shingo Kiuchi *et al.*, assignors to Alpine Electronics, Incorporated
25 October 2005 (Class 704/233); filed in Japan 12 July 2000

This patent proposes to augment a basic Fourier spectral-feature extraction scheme [and suggests numerous times to employ a Humming (sic) window] with a ''noise reduction system.'' The main innovation over conventional adaptive-filtering NR methods is the apparently novel insight that one need only apply the noise-reduction filter (obtained by established adaptive-noise-canceling routines) in the frequency domain to half the Fourier-transform points, the other half being first discarded as the redundant frequency spectrum.—SAF

6,965,860

43.72.Dv SPEECH PROCESSING APPARATUS AND METHOD MEASURING SIGNAL TO NOISE RATIO AND SCALING SPEECH AND NOISE

David Llewellyn Rees and Robert Alexander Keiller, assignors to Canon Kabushiki Kaisha
15 November 2005 (Class 704/227); filed in the United Kingdom 23 April 1999

A method is described for compensating speech processing frame by frame in the presence of noise. The signal-to-noise ratio is estimated from the signal prior to the onset of speech, so it presumably works well only with fairly stationary noise. Working with cepstral feature vectors for the average noise and the current frame, the noise vector is subtracted from the frame vector when the SNR is low and performs no correction when the SNR is high.—SAF

6,959,082

43.72.Ew METHOD AND SYSTEM FOR AUTOMATIC GAIN CONTROL WITH ADAPTIVE TABLE LOOKUP

Lee F. Holeva, assignor to 3Com Corporation
25 October 2005 (Class 379/388.03); filed 10 July 2001

A simple method is outlined for automatic gain control in a speech communication device, which involves the generation and use of a ''lookup table'' storing a fixed set of gain values which can be selected by a standard sort of adaptive signal-power estimation scheme. The goal is presumably to reduce the computational complexity of deciding on the gain at each step, although an additional element of the proposal involves the capability to recalculate the gain values in the table by means of an adaptive error signal.—SAF

6,959,275

43.72.Ew SYSTEM AND METHOD FOR ENHANCING THE INTELLIGIBILITY OF RECEIVED SPEECH IN A NOISE ENVIRONMENT

Adoram Erell, assignor to D.S.P.C. Technologies Limited
25 October 2005 (Class 704/225); filed 30 May 2001

Systems which enhance speech received under noisy conditions often involve automatic gain control and dynamic range compression components which operate adaptively. This patent proposes to improve the adaptive performance of this type of method by invoking a long-term peak-amplitude value calculated for each frame containing a speech signal. The intuition is that speech is an inherently dynamic and fluctuating signal which can easily trick a simplistic adaptive system that relies entirely on the short-term signal characteristics.—SAF

6,961,702

43.72.Fx METHOD AND DEVICE FOR GENERATING AN ADAPTED REFERENCE FOR AUTOMATIC SPEECH RECOGNITION

Stefan Dobler *et al.*, assignors to Telefonaktiebolaget LM Ericsson (publ)
1 November 2005 (Class 704/244); filed in the European Patent Office 7 November 2000

Speech recognition of an utterance normally involves comparison against a reference model. The proposal here is to take account of each different speaker or environment by adapting the reference model and trying again. If this iteration gets somewhere, the adapted reference may be kept and used for future recognition.—SAF

6,961,703

43.72.Fx METHOD FOR SPEECH PROCESSING INVOLVING WHOLE-UTTERANCE MODELING

Alan Lawrence Higgins and Lawrence George Bahler, assignors to ITT Manufacturing Enterprises, Incorporated
1 November 2005 (Class 704/249); filed 13 September 2000

The basic task at hand is speaker verification from a single utterance, and it is suggested that instead of using traditional frame-by-frame speech recognition sorts of algorithms, for this purpose it will be better to model and recognize one huge vector of features for the entire utterance at once. The normal computational obstacle with this idea arises from the extreme dimensionality of such a vector, preventing computation of linear discriminant analysis. This problem is said to be solved here by a new eigenvector approach in which a subspace is first selected upon which the observed between-speaker variations lie—the intuition being that it will not be worth keeping account of a large fraction of the feature differences.—SAF

6,961,696

43.72.Gy CLASS QUANTIZATION FOR DISTRIBUTED SPEECH RECOGNITION

Tenkasi V. Ramabadran and Alexander Sorin, assignors to Motorola, Incorporated
1 November 2005 (Class 704/207); filed 7 February 2003

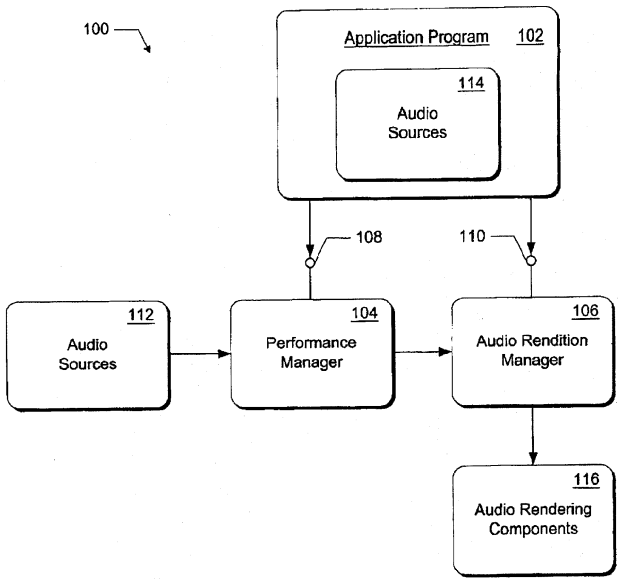
This patent presents a new bit-coding scheme which incorporates pitch class and other sound class information into code words that ultimately involve a one-bit savings over more well-known coding schemes. This small compression amounts to large network bandwidth gains for wireless communications sending coded audio back to servers for distributed speech recognition from cell phones and such.—SAF

6,970,822

43.72.Gy ACCESSING AUDIO PROCESSING COMPONENTS IN AN AUDIO GENERATION SYSTEM

Todor J. Fay and Brian L. Schmidt, assignors to Microsoft Corporation
29 November 2005 (Class 704/270); filed 7 March 2001

To provide a multimedia application program direct access to the audio data-processing components, a performance manager (audio source manager) and an audio rendition manager are provided by the application. The performance manager instantiates audio data processing components for



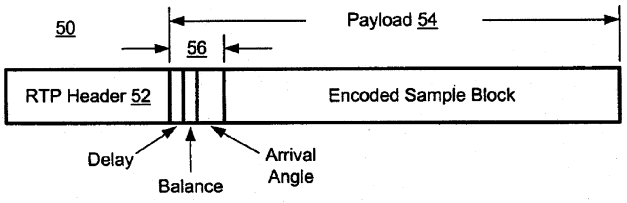
each audio source. The audio rendition manager includes a synthesizer component that generates audio-waveform data from the received audio instructions and buffers the processed data.—DAP

6,973,184

43.72.Gy SYSTEM AND METHOD FOR STEREO CONFERENCING OVER LOW-BANDWIDTH LINKS

Shmuel Shaffer and Michael E. Knappe, assignors to Cisco Technology, Incorporated
6 December 2005 (Class 379/420.01); filed 11 July 2000

A pseudo-stereo conferencing encode/decode scheme is implemented in about the same bandwidth as a monaural signal by sampling the sound field at two locations. The relative delay between the two sound-field signals is recorded and transmitted over a packet network with a single audio pay-



load formed by combining the separate sound signals. At the decoder, the relative delay is used to form two spatially separated playback signals, so the listener receives source location information for multiple participants.—DAP

6,947,731

43.72.Ja METHOD FOR CONVERTING STATUS MESSAGES OUTPUT IN SPOKEN FORM

Erich Kamperschroer, assignor to Siemens Aktiengesellschaft
20 September 2005 (Class 455/419); filed in Germany 30 September 1999

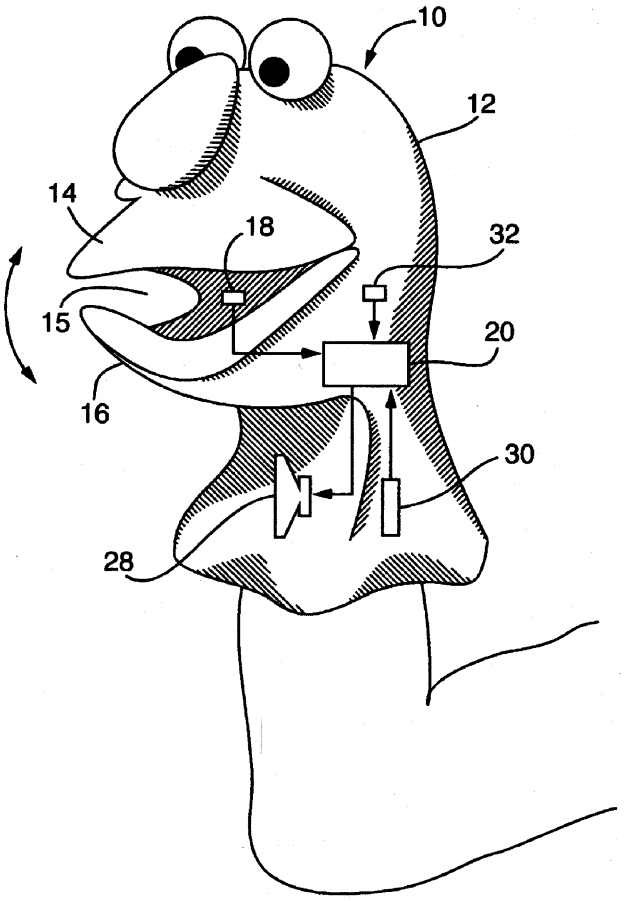
Well into the realm of the completely obvious, this patent deals with the use of speech synthesis to notify a user of the status of a device which was not equipped for speech output. However, it is also assumed that the device in question does have some means of data signalling and that a speech-synthesis device is available. The only remaining task is then the translation of various device signals into the appropriate spoken phrases. And for this, we maintain a Patent Office?—DLR

6,971,943

43.72.Ja INTERACTIVE SOUND PRODUCING TOY

Arne Schulze, Santa Monica, California
6 December 2005 (Class 446/327); filed 9 August 2004

A hand-operated puppet **10** has a switch in the mouth **18** that controls the voice synthesizer **20** (element **32** is a program select switch). The idea is to synchronize an on-chip recording with the hand movements. The switch



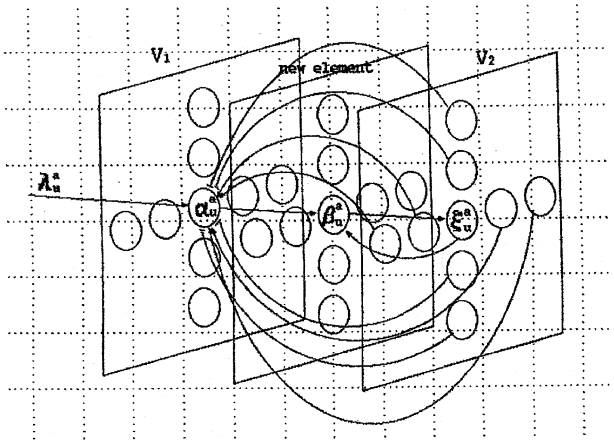
18 provides a very coarse quantization of mouth position, but child amusement is assured.—MK

6,947,890

43.72.Ne ACOUSTIC SPEECH RECOGNITION METHOD AND SYSTEM USING STEREO VISION NEURAL NETWORKS WITH COMPETITION AND COOPERATION

Tetsuro Kitazoe, Miyazaki, Japan *et al.*
20 September 2005 (Class 704/232); filed in Japan 28 May 1999

This speech-recognition architecture begins with a typical (Mel cepstrum) analysis of each acoustic frame. This is followed by a neural network analyzer which incorporates both time and vocabulary contexts to recognize the input speech. Written in the language of neural net math, the patent



includes an extensive discussion of the frame-comparison processes used. Some evaluations are presented in support of the assertion that the method is superior to the more traditional HMM recognizer architecture.—DLR

6,957,183

43.72.Ne METHOD FOR ROBUST VOICE RECOGNITION BY ANALYZING REDUNDANT FEATURES OF SOURCE SIGNAL

Narendranath Malayath and Harinath Garudadri, assignors to Qualcomm Incorporated
18 October 2005 (Class 704/246); filed 20 March 2002

The typical speech recognition front end involves the generation of a feature vector for each analysis frame, for example, by spectral or cepstral analysis. Such feature vectors are preferred when their components are decorrelated. This patent lays claim to a general method whereby, in addition to the primary feature vector, some sort of secondary feature vector is also obtained for each analysis frame, and this secondary vector is assumed to be somehow correlated with the primary. The intention is to improve recognition performance by somehow combining the two correlated feature vectors—nearly all conceivable ways of generating and combining the vectors are covered in the patent claims, although no specific embodiment is proven to accomplish the desired improvement.—SAF

6,959,278

43.72.Lc SYSTEMS AND METHODS FOR IMPLEMENTING SEGMENTATION IN SPEECH RECOGNITION SYSTEMS

Chang-Qing Shu and Han Shu, assignors to Verizon Corporate Services Group Incorporated
25 October 2005 (Class 704/243); filed 5 April 2001

The proposed method attacks the problem of segmenting speech into phones for further processing by examining tracks of cepstral peaks. The intuition is that where cepstral peaks change dramatically from one frame to the next, one is likely to find a transition between speech segments.—SAF

6,961,701

43.72.Ne VOICE RECOGNITION APPARATUS AND METHOD, AND RECORDING MEDIUM

Hiroaki Ogawa *et al.*, assignors to Sony Corporation
1 November 2005 (Class 704/236); filed in Japan 2 March 2000

The rather unbelievable suggestion here is that one can get around all of the usual out-of-vocabulary problems with speech recognition by somehow involving a huge dictionary of over a million forms. The general approach is to search this massive dictionary (don't the authors realize that no

dictionary could ever be sufficient?) using restricted or partial matching criteria which might help the computational complexity.—SAF

6,963,837

43.72.Ne ATTRIBUTE-BASED WORD MODELING

Michael Finke *et al.*, assignors to Multimodal Technologies, Incorporated
8 November 2005 (Class 704/256); filed 6 October 2000

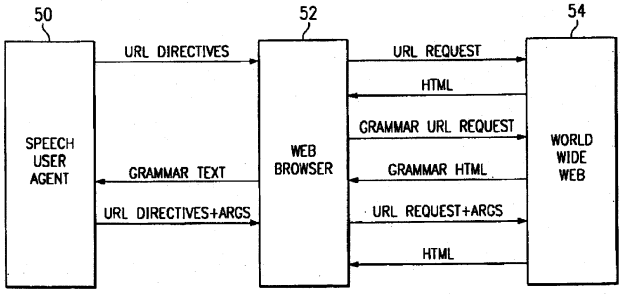
This patent describes an apparently novel approach to speech recognition that moves beyond the usual phone-based segmental methodology. The general idea involves creating a new kind of acoustic model for recognition that uses phonetic ''attributes'' together with suprasegmental attributes. ''A word is transcribed as a sequence of instances ... which are bundles of instantiated attributes (i.e., attribute-value pairs).'' The attributes may involve traditional phonetic features, but also dialect, gender, word or syllable position, word class, etc. This new approach addresses the fundamental problems with the prior art involving multiple possible pronunciations for each word in the dictionary, which can paradoxically lead to decreased recognition performance.—SAF

6,965,864

43.72.Ne VOICE ACTIVATED HYPERMEDIA SYSTEMS USING GRAMMATICAL METADATA

Philip R. Thrift and Charles T. Hemphill, assignors to Texas Instruments Incorporated
15 November 2005 (Class 704/275); filed 10 April 1995

A flexible-vocabulary speech interface to the Web is proposed for accessing information via voice-activated control information embedded in HTML pages. The system utilizes phonetic models for speaker-independent



voice recognition without any training required. Grammars are dynamically modified for particular applications.—DAP

6,968,310

43.72.Ne METHOD, SYSTEM, AND APPARATUS FOR SPEECH RECOGNITION

Kazuo Nemoto, assignor to International Business Machines Corporation
22 November 2005 (Class 704/231); filed in Japan 2 May 2000

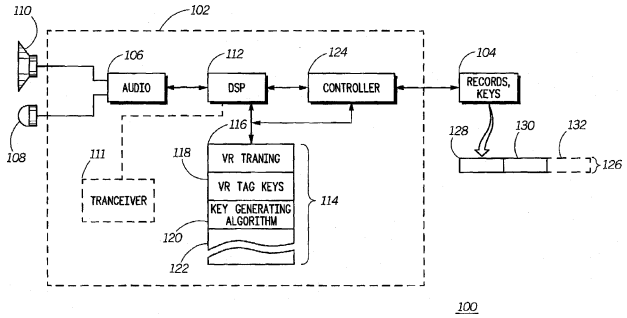
Here is another speech-recognition tweak which suggests matching whole words against stored pronunciations found to match the current-session speaker in favor of other stored models, when recognition accuracy would thereby improve. The objective is to reduce the number of possible online templates in the grammar in response to the actual pronunciations employed by the speaker. The context is digit recognition in Japanese.—SAF

6,970,817

43.72.Ne METHOD OF ASSOCIATING VOICE RECOGNITION TAGS IN AN ELECTRONIC DEVICE WITH RECORDS IN A REMOVABLE MEDIA FOR USE WITH THE ELECTRONIC DEVICE

Douglas Eugene Ross *et al.*, assignors to Motorola, Incorporated
29 November 2005 (Class 704/201); filed 31 October 2001

Voice-recognition training data stored in a remote device, such as a cellphone or PDA, is associated with corresponding records in removable



media by generating a key when a new voice-recognition tag is trained. The keys are stored in the remote device and/or the removable media.—DAP

6,970,818

43.72.Ne METHODOLOGY FOR IMPLEMENTING A VOCABULARY SET FOR USE IN A SPEECH RECOGNITION SYSTEM

Xavier Menedez-Pidal and Lex S. Olorenshaw, assignors to Sony Corporation
29 November 2005 (Class 704/236); filed 14 March 2002

This is yet another patent describing a method for employing an n-best list which ranks speech-recognition hypotheses in decreasing order of likelihood. How many variations on this idea can be patented?—SAF

6,973,427

43.72.Ne METHOD FOR ADDING PHONETIC DESCRIPTIONS TO A SPEECH RECOGNITION LEXICON

Mei-Yuh Hwang *et al.*, assignors to Microsoft Corporation
6 December 2005 (Class 704/249); filed 26 December 2000

A speech-recognition methodology is outlined which augments the standard art by attempting to recognize a spoken word phonetically when the usual vocabulary-based recognition fails, though a phone-based approach is not employed since these are known to be poor performers. Instead, a scheme based on ''syllable-like units'' is put forth, which has several variants including an n-gram language-model implementation for these empirical approximations to syllables. In this way, unknown utterances such as personal names can at least be represented ''syllabically'' rather than treated as garbage, as is commonly done.—SAF

6,973,426

43.72.Pf METHOD AND APPARATUS FOR PERFORMING SPEAKER VERIFICATION BASED ON SPEAKER INDEPENDENT RECOGNITION OF COMMANDS

John E. Schier and Patrick E. Jackson, assignors to Cisco Technology, Incorporated
6 December 2005 (Class 704/246); filed 29 December 2000

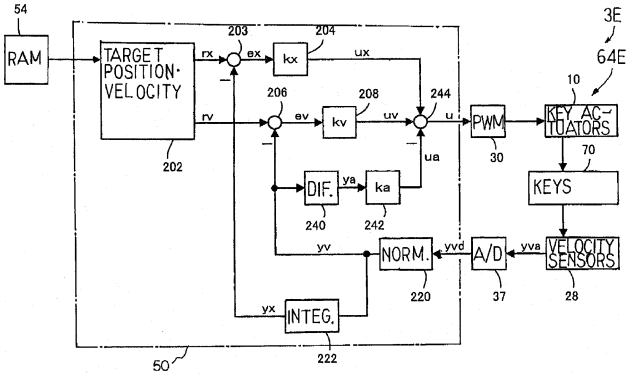
The patent describes a system that combines speaker-independent recognition of simple commands with speaker-verification computations on the same spoken utterance. Each speaker's verification template is updated with their subsequent utterances of the command. The recognition and verification procedures themselves are black boxes here.—SAF

6,969,791

43.75.Mn AUTOMATIC PLAYER KEYBOARD MUSICAL INSTRUMENT EQUIPPED WITH KEY SENSORS SHARED BETWEEN AUTOMATIC PLAYING SYSTEM AND RECORDING SYSTEM

Yuji Fujiwara, assignor to Yamaha Corporation
29 November 2005 (Class 84/13); filed 8 March 2004

This patent follows in the long line of Synclavier patents previously reviewed. The inventor proposes changing the actuator/sensor open-loop control with closed-loop control to eliminate errors and improve performance. Multiple implementation schemes are illustrated and discussed. —MK



6,972,362

43.75.Wx METHOD AND DEVICE FOR GENERATING ELECTRONIC SOUNDS AND PORTABLE APPARATUS UTILIZING SUCH DEVICE AND METHOD

Yutaka Nakamura, assignor to Rohm Company, Limited
6 December 2005 (Class 84/603); filed in Japan 9 January 2002

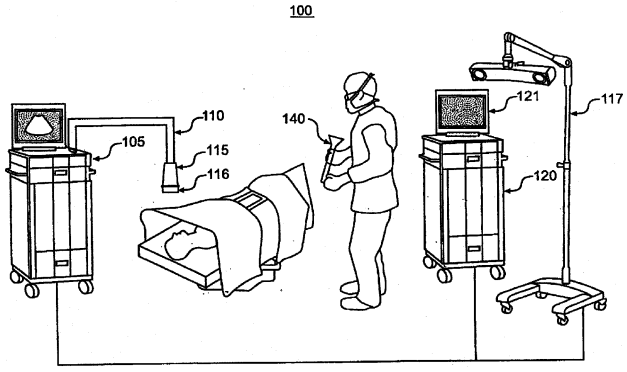
This patent attacks the so-called "parameter update problem" where a realtime DSP must update synthesis parameters and process samples in a fixed sample period. The inventor makes several assumptions: (a) complex timbres demand more processor cycles, (b) missed samples are acceptable if the stale sample data is held, and (c) everything will be acceptable if the mean deviation of the processing time is zero. None of these are absolutely true. Complex timbres depend on the algorithm, not just cycles. Missed samples can be audible depending on the program material and, finally, updates come in large batches, ruining the short-term mean.—MK

6,968,224

43.80.Qf METHOD OF DETECTING ORGAN MATTER SHIFT IN A PATIENT

Paul Kessman *et al.*, assignors to Surgical Navigation Technologies, Incorporated
22 November 2005 (Class 600/407); filed 19 September 2003

This surgical instrument tracking method uses a navigation system incorporating an ultrasound machine with an attached computer. The computer generates an icon representing the tip of a surgical instrument and the instrument's trajectory. The icon is overlayed on a real-time ultrasound image in an image plane, so that when the instrument traverses the image plane, the trajectory display is altered to represent the instrument's travel. The system also contains a localizer coupled to the ultrasound machine and a display which shows the generated icon superimposed on the real-time image.—DRR



6,967,275

43.75.St SONG-MATCHING SYSTEM AND METHOD

Daniel Ozick, assignor to iRobot Corporation
22 November 2005 (Class 84/616); filed 24 June 2003

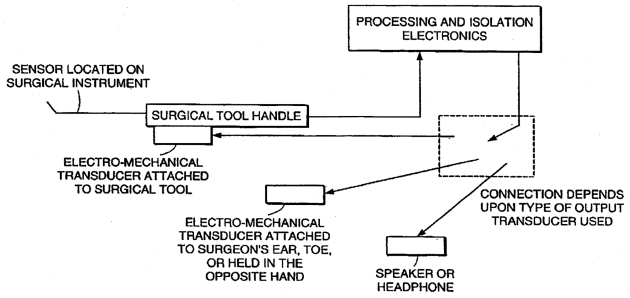
This inventor argues that previous art has not included automatic accompaniment in synchrony with a song-matching system. It's a pity he didn't read (or cite) Dannenberg's 1988 patent (United States Patent 4,745,836, filed in 1985). This well-known work used dynamic time warping to do the score match and provide an accompaniment to a solo player.—MK

6,969,384

43.80.Qf SURGICAL DEVICES AND METHODS OF USE THEREOF FOR ENHANCED TACTILE PERCEPTION

Eugene de Juan, Jr. and Patrick S. Jensen, assignors to The Johns Hopkins University
29 November 2005 (Class 606/1); filed 3 January 2001

In order to provide a "feel" feedback in the course of microsurgery, a surgical instrument is provided with a sensor element that is capable of detecting interaction of the device with the environment. Information relat-



ing to the device interaction, particularly of the nonvisual sort, can be communicated in real time in the form of tactile or auditory information.—DRR

6,970,733

43.80.Qf SYSTEM AND METHOD FOR ELECTRODE LOCALIZATION USING ULTRASOUND

N. Parker Willis *et al.*, assignors to SciMed Life Systems, Incorporated
29 November 2005 (Class 600/424); filed 21 October 2002

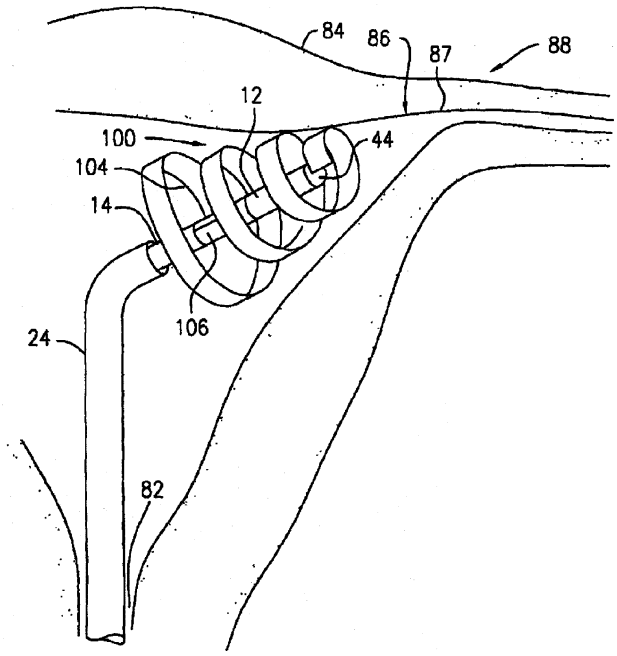
This localization system uses one or more ultrasound reference catheters to establish a fixed three-dimensional coordinate system within a patient's heart using the principles of triangulation. The coordinate system is represented graphically in three dimensions on a video monitor and helps the clinician guide other medical devices equipped with ultrasound transducers through the body to locations at which they are needed to perform clinical procedures. For example, the system can be used in the heart to help a physician guide mapping catheters for measuring electrical activity and ablation catheters for ablating selected regions of cardiac tissue at specific locations.—DRR

6,972,018

43.80.Sh APPARATUS AND METHOD FOR TRANSCERVICAL STERILIZATION BY APPLICATION OF ULTRASOUND

Thomas P. Ryan and John S. Crombie, assignors to Gynecare a division of Ethicon, Incorporated
6 December 2005 (Class 606/50); filed 28 June 2002

This is an ultrasound device for sterilizing a female patient in a transcervical procedure. The device consists of an elongated catheter featuring a piezoelectric ultrasound transducer at its distal end. The catheter is inserted transcervically into the uterus and directed to the intramural region of a fallopian tube, to be in direct contact with the surrounding tissue. The transducer produces a radially dispersing acoustical wave front, heating the adjacent tissue and forming a thermal lesion therein. A fibrous tissue forms during the healing process, thereby sealing off the fallopian tube. One embodiment of the device includes a movable thermal sensor for monitoring the tissue temperature at multiple locations along the lesion.—DRR



6,969,354

43.80.Vj ADAPTABLE INTRAOPERATIVE OR ENDOCAVITY ULTRASOUND PROBE

Vaughn R. Marian, assignor to Acuson Corporation
29 November 2005 (Class 600/459); filed 25 September 2001

This probe has an adjustable section between the handle and the transducer. This section allows the transducer to be rotated relative to the handle prior to insertion within a patient. The shape of the section is maintained while the probe is in use and need not be further adjusted.—RCW

6,971,991

43.80.Vj APPARATUS FOR MULTIMODAL PLANE WAVE ULTRASOUND IMAGING

Robert S. Lasser *et al.*, assignors to Imperium, Incorporated
6 December 2005 (Class 600/437); filed 7 March 2003

This apparatus contains an ultrasonic imaging unit and an x-ray imaging unit as well as a capability to hold in compression the object to be imaged. The x-ray source and the ultrasonic source can each be moved from an active imaging position to an inactive or nonimaging position. Both x-ray images and ultrasonic images are made in the transmission mode and the images are registered to assist user correlation of images. Speckle contained in the ultrasonic image is reduced to improve contrast resolution.—RCW

6,971,992

43.80.Vj ULTRASONIC IMAGING METHOD AND APPARATUS

Marino Cerofolini, assignor to Esaote, S.p.A.
6 December 2005 (Class 600/447); filed in Italy 13 August 2002

Transducers in this system are excited with delays to produce a rela-

tively homogeneous acoustic field over a large region of interest. Two subgroups are used, one transmitting a waveform and the other transmitting an inverted version of the waveform. Because the transmitted waveform produces a relatively homogeneous distribution of energy over the whole region of interest, received signals can be processed with different delays to cover the entire region with receive beams that are focused along scan lines to permit a higher frame rate than would be possible with a narrow transmit beam.—RCW

FORUM

Forum is intended for communications that raise acoustical concerns, express acoustical viewpoints, or stimulate acoustical research and applications without necessarily including new findings. Publication will occur on a selective basis when such communications have particular relevance, importance, or interest to the acoustical community or the Society. Submit such items to an appropriate associate editor or to the Editor-in-Chief, labeled FORUM. Condensation or other editorial changes may be requested of the author.

Opinions expressed are those of the individual authors and are not necessarily endorsed by the Acoustical Society of America.

Poincaré's criterion in room acoustics

James Bernard Lee[a)]
6016 S.E. Mitchell Street, Portland, Oregon, 97206

(Received 13 January 2006; revised 28 February 2006; accepted 6 March 2006)
[DOI: 10.1121/1.2190188]

Three papers by W. B. Joyce, published in this Journal, are crucial in applying statistical methods to the physics of sound in rooms.[1–3] They are well known and occasionally cited, but rarely applied to justify statistical approximations to the acoustics of large rooms, either in presentations at meetings of this Society or in articles published in this Journal.

They are difficult papers, presuming readers understand abstract terms like "mixing" and "ergodicity," as well as knowing recondite theorems and methods of mathematical analysis; perhaps this is why their conclusions most often are honored in the breach. Following is a précis of an analogous problem to help authors, editors, and reviewers readily to apply Joyce's findings to submissions in room acoustics:

"…let us imagine for a moment a gas whose molecules cannot collide mutually, but can be deviated by the sides of the vessel in which the gas is enclosed. If the form of the vessel is sufficiently complicated, it will not be long before the distribution of the molecules and that of their velocities become uniform. This will not happen if the vessel is spherical, or if it has the form of a rectangular parallelepiped. And why not? Because in the former case the distance of any particular trajectory from the center remains constant, and in the latter case we have the absolute value of the angle of each trajectory with the sides of the parallelepiped."[4]

Thus Henri Poincaré in his essay, "Chance." This system is formally identical to Joyce's concept of sound as a "phonon gas," the "particles" of which do not affect each other and rebound only from the surfaces of a room. Joyce does not cite Poincaré, but in the final paragraphs of his last paper does remark that "spheres" and "rectangular parallelepipeds" are not mixing if their surfaces are not complicated. Clearly the two authors are in accord.

Following is a contrary assertion, and the original basis for statistical methods in room acoustics:

"The above results, then, all point to the evident, but perhaps not appreciated, fact that the dispersion of sound between all parts of a hall is very rapid in comparison with the total time required for its complete absorption, and that in a very short time after the source has ceased the intensity of the residual sound, except for the phenomenon of interference to be considered later, is very nearly the same everywhere in the room."[5]

Thus Wallace Clement Sabine expressly acknowledges that admitting the wave nature of sound concomitantly admits the existence of nonuniform fields of sound in rooms, but asserts that absent physical acoustics sound in rooms is effectively uniform for purposes of calculation.

So could Poincaré be wrong and Joyce's effort to reconcile these assertions pointless? This must be resolved.

The kinetic theory of gases, new in Poincaré's time, is a well tested and widely applied theory of physics. But the system Poincaré analyzes in the passage quoted is unusual, applying only to tenuous gases in small containers, with the mean free path much larger than the linear dimensions of the container. In such systems the particles do not interact with each other, do not exchange momentum and energy, and so do not converge to equilibrium of themselves. Such particles become distributed uniformly only if the shape of their container "…is sufficiently complicated." This is termed "mixing," and requires a finite time, called the "mixing time," to eventuate.

Clearly Sabine does not understand this need of suitable geometry for mixing, although he does acknowledge the requirement of a definite time, assumed to be negligible, for its completion. Sabine, contemporary with Poincaré and using the kinetic theory of gases as an analogy to sound, does not know that concepts such as uniform density and mean free path come freighted with pertinent restrictions.

Joyce successfully quantifies both these criteria vital to Sabine's model of sound in rooms, and following Joyce we may assume that practical auditoriums are "sufficiently complicated" to fulfill Poincaré's requirement. But mixing time is another matter. Sabine asserts, without proof, that uniform energy density should obtain in a time much shorter than what we call "reverberation time." He says nothing about growth of sound.

This cannot be correct, for the decay of sound, characterized by reverberation time, is mediated by the logarithmic scale with which it is measured; on the same scale the time for the growth of sound is much shorter, reaching only 2 dB below its asymptotic level in one time constant, about 1/14th of the reverberation time. The time constant therefore is the realistic datum for temporal events in room acoustics, as Allan Pierce and others understand.[6]

Close study of Joyce's papers reveals a change of emphasis, as he gradually comes to grasp the central fact of room acoustics: it is absorption that primarily determines uniformity of sound in most practical rooms; "Poincaré geometry" and mixing time are not important if much absorption is present. For example, in concert halls and opera houses about 1/3 of the surface area is occupied by the audience, which is strongly absorbent; uniform fields of sound, even as the severe abstraction of a "phonon gas," cannot exist in such rooms. Some rooms, notably reverberation chambers, are designed for very small absorption, and in these complexity and mixing time are relevant.[7]

I propose that articles and presentations in room acoustics presuming a statistical basis, as submitted to the Acoustical Society of America, be vetted *a priori* by the following criteria:

(1) Is the room "sufficiently complicated" to ensure mixing in Poincaré's sense?

(2) Is the mixing time much shorter than one time constant?

(3) Is the absorption small enough for uniform fields of sound to be generated?

These are basic, reasonable, and answerable questions. Well conducted relevant research reported in well written papers should supply convincing answers to them.

a)Electronic mail: cadwal@quik.com

[1]W. B. Joyce, "Sabine's reverberation time and ergodic auditoriums," J. Acoust. Soc. Am. **58**, 643–655 (1975).
[2]W. B. Joyce, "Exact effect of surface roughness on the reverberation time of a uniformly absorbing spherical enclosure," J. Acoust. Soc. Am. **64**, 1429–1436 (1978).
[3]W. B. Joyce, "Power series for the reverberation time," J. Acoust. Soc. Am. **67**, 564–571 (1980).

[4]J. H. Poincaré, *Science and Method* (Dover, Mineola, New York, 2003) (republication of 1908 original), p. 80.
[5]W. C. Sabine, *Collected Papers on Acoustics* (Dover, Mineola, New York, 1964) (republication of 1922 original), p. 19.
[6]A. D. Pierce, *Acoustics, an Introduction to Its Physical Principles and Applications* (The Acoustical Society of America, Melville, New York, 1989), p. 263.
[7]M. d. Godoy, P. J. Barry, and S. R. Bistafa, "A study of the influence of mounting conditions on the measured sound absorption of suspended ceilings," J. Acoust. Soc. Am. **119**, 33–36 (2006).

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.

Pseudo-3D modeling of a surface-bonded Lamb wave source (L)

Emmanuel Moulin,[a] Sébastien Grondel, Mustapha Baouahi, and Jamal Assaad
IEMN, UMR CNRS 8520, OAE Department, Université de Valenciennes et du Hainaut Cambrésis, Le Mont Houy, 59313 Valenciennes Cedex 9, France

(Received 21 July 2005; revised 1 March 2006; accepted 1 March 2006)

In this paper, a simple technique allowing the prediction of the Lamb wave field excited in an isotropic plate by a transducer of finite dimensions is presented. The basic idea is to separate the problem into two uncoupled parts. First, the Lamb wave excitation problem, in the plane defined by the thickness and width directions of the transducer, is treated. Then the source diffraction effects in the plane of the plate can be quantitatively estimated. The subsequent formulation thus offers a simple way of developing three-dimensional (3D) solutions from a two-dimensional (2D) modeling. Compared to a heavy full-3D numerical modeling, this technique appears to be a very satisfying alternative. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2189447]

PACS number(s): 43.20.Ks, 43.40.Le, 43.60.Qv, 43.38.Ar [RMW] Pages: 2575–2578

I. INTRODUCTION

For several years, Lamb wave testing has been seen as a promising technique for health monitoring of civil or aerospace structures. In order to make such monitoring faster, less expensive, and effortless, the integration of piezoelectric patches to the material has been considered.[1,2] The behavior of such an integrated transducer and the type of waves emitted depend strongly on the nature of the host material and the insertion conditions. Then a proper modeling tool, allowing to take into account the whole complexity of the problem, is required.

In this context, previous works have focused on the development of a 2D hybrid technique for the modeling of plates containing integrated sources.[3–5] The imposed plane-strain condition implied the infiniteness along one dimension, which is only a rough approximation of a bar-shaped transducer. In order to account for the finite length of an actual transducer, a 3D modeling is required. However, instead of developing unpractical full 3D numerical models, the idea suggested here is to uncouple the divergence effects in the plane of the plate from the mode characteristics with respect to the plate thickness.

This concept has already been presented in a previous paper[6] and applied to multi-element sources for Lamb wave beam-steering.[7] However, the simplistic approach used then only provided qualitative results. This paper aims at extending these works so that quantitative results are accessible, allowing direct comparison to experimental or 3D numerical results.

Up to now, only a very small number of works have dealt with 3D modeling of Lamb wave excitation by finite sources. Among them, a recent paper by Rose *et al.*[8] is particularly noticeable. Based on Mindlin plate theory, the authors derived analytical solutions for point forces and bending moments and then deduced solutions for circular-or rectangular-shaped surface-mounted sources by spatial integration. However, the formulation is relatively delicate to handle and since Mindlin theory is an approximate plate model, a number of physical aspects are not correctly described (symmetric modes, higher order antisymmetric modes,…).

The modeling method proposed here is based on line- and point-source solutions derived from the works by Viktorov and Ditri, respectively.[9,10] A remarkable synthesis and useful applications of these works have been presented by Wilcox.[11] Though not fundamentally different in nature, the formulation introduced in the present paper is more suited to the objective. As will be seen indeed, the link established with the normal mode expansion method[12] results in a very versatile and lightweight pseudo-3D modeling technique.

II. MODELING PRINCIPLE

A. Finite length excitation

Let us consider an isotropic plate with thickness d, excited by a given harmonic stress applied to a localized area of its upper surface (Fig. 1). The excitation area S is supposed to be rectangular shaped, with length L and width W. The applied stress profile is described by a function $\sigma_0(x,y)$ defined over S and representing the action of a transducer. In order to simplify the notations, all the derivations presented

[a]Electronic mail: emmanuel.moulin@univ-valenciennes.fr

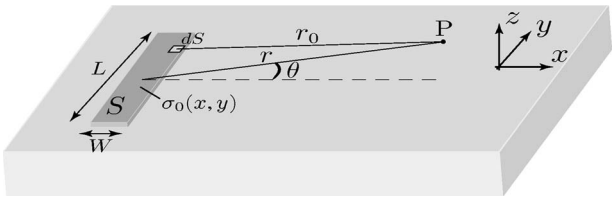


FIG. 1. Description of the problem geometry.

below will consider the normal component of the displacement on the plate surface only. If required, expressions for the transverse displacement or the displacement inside the plate thickness could be derived exactly the same way.

Assuming that a single Lamb wave is generated in the plate (this assumption can be released afterwards), the normal displacement du excited at a given point P of the plate upper surface by a small part dS with coordinates (x,y) of the actuation area can be deduced from the axisymmetric solution proposed by Ditri *et al.*[10] This can be written in a compact form as

$$du = \frac{j\sigma_0(x,y)}{4\mu}\frac{\Gamma k}{\Delta'}H_0^{(1)}(kr_0)e^{-j\omega t}dS, \tag{1}$$

where r_0 is the distance from the excitation point, k is the wave number of the considered Lamb wave, and Γ and Δ are functions depending on k and the material properties. They are representative of eigenmode-related features. Their complete expressions can be found in the literature[10] whenever needed and thus will not be reminded here. The prime $'$ denotes derivation with respect to k. $H_0^{(1)}$ represents the zeroth-order Hankel function of the first kind. For conciseness, the term $e^{-j\omega t}$ will be omitted in the next equations.

Let us now consider the case of an excitation represented by separable variable functions:

$$\sigma_0(x,y) = \sigma_{0x}(x)\sigma_{0y}(y). \tag{2}$$

Here σ_{0x} will be supposed to depend on W only and σ_{0y} on L only. In the case of an actual transducer, such an assumption is acceptable provided L and W are different enough to allow the length and width vibration modes to be uncoupled.[6]

Considering a location far enough from the source so that $r \gg L, W$ yields

$$r_0 \simeq r - x\cos\theta - y\sin\theta. \tag{3}$$

Then, by integrating Eq. (1) over the excitation area S and using the far-field asymptotical approximation of the Hankel function along with Eq. (3), the total normal displacement at a point P with polar coordinates (r,θ) can be expressed as

$$u(r,\theta) = \frac{j(1-j)\Gamma}{4\mu\sqrt{\pi}\Delta'}\sqrt{\frac{k}{r}}e^{jkr}\int_{-L/2}^{L/2}\sigma_{0y}(y)e^{-jky\sin\theta}dy \times \int_{-W/2}^{W/2}\sigma_{0x}(x)e^{-jkx\cos\theta}dx. \tag{4}$$

B. Infinite length excitation

Let us now consider the same plate, subject to an excitation defined by the same width W and the same applied stress function σ_{0x} as in the previous section. This case, however, will significantly differ from the previous one in that the length L will be considered infinite. In that case, a two-dimensional, plane-strain formulation is valid. Then the displacement at a given point, located at a distance r from the source, can be obtained by integration of a line-source solution. Such a solution can be derived from the work by Viktorov on wedge transducer excitation,[9] which had also been applied by Wilcox.[11] This yields an expression of the form

$$u_{2D}(r) = \frac{j\Gamma}{2\mu\Delta'}e^{jkr}\int_{-W/2}^{W/2}\sigma_{0x}(x)e^{-jkx}dx, \tag{5}$$

where Γ and Δ are the same as in Eqs. (1) and (4).

By introducing the modal amplitude A_m for the given Lamb mode, derived from the 2D normal mode formalism,[3,12] this displacement can also be written as

$$u_{2D}(r) = A_m u_m e^{jkr}, \tag{6}$$

where u_m is the modal displacement field (eigen vector) for the considered mode.

C. Connection of the 2D and 3D problems

Equations (4) and (5) are independently established from two geometrically different excitation cases. However, since the propagation medium (the solid plate) is supposed to be the same in both cases, it is no surprise that some eigenmode terms remain identical in both equations. This is indeed the case with Γ and Δ, as written above. Besides, the integrals over x are the same in Eqs. (4) and (5) if either the source width W is small compared to the wavelength or θ is small, or a combination of both.

In those conditions, comparison of these two equations and substitution of the above-mentioned identical terms yield the following relation between the finite-length and the 2D infinite-length solutions:

$$u(r,\theta) = \frac{(1-j)}{2\sqrt{\pi}}\sqrt{\frac{k}{r}}u_{2D}(r)\int_{-L/2}^{L/2}\sigma_{0y}(y)e^{-jky\sin\theta}dy \tag{7}$$

Then, by invoking Eq. (6), the finite-length displacement can finally be expressed as

$$u(r,\theta) = \frac{(1-j)}{2\sqrt{\pi}}\sqrt{\frac{k}{r}}A_m u_m e^{jkr}\int_{-L/2}^{L/2}\sigma_{0y}(y)e^{-jky\sin\theta}dy \tag{8}$$

In the case of an actual rectangular shaped surface-mounted transducer, it has been shown in a previous work[6] that the excitation $\sigma_{0y}(y)$ could be considered uniform over the transducer length L, provided the condition $W/L \leqslant 0.4$ is ensured. In that case, Eq. (8) can be simplified as

$$u(r,\theta) = \frac{(1-j)}{\sqrt{\pi}}\sqrt{\frac{k}{r}}A_m u_m \frac{\sin(k(L/2)\sin\theta)}{k\sin\theta}e^{jkr}. \tag{9}$$

This equation is similar to the general formulation given in Ref. 6. In this previous work, however, the directivity

relation was obtained by merely using a basic scalar diffraction model, which constituted quite a rough approximation. This resulted in rather satisfying qualitative directivity predictions but an undetermined, frequency- and mode-dependent amplitude term remained. This made direct comparison to 3D numerical modeling or experimental results impossible. On the contrary, Eq. (9) of the present paper allows the displacement waveforms to be correctly predicted, even in the multi-mode case.

The first step of the pseudo-3D modeling scheme is the treatment of the 2D plane-strain problem (assuming L infinite). This provides the modal amplitudes A_m for the considered transducer width W. It should be noted that the actual transducer thickness and piezoelectric characteristics can also be taken into account if required. A complete description of the procedure can be found in previous works.[3–5] Therefore, it will not be repeated here. Then, the third dimension is introduced by using Eq. (9). Hence, wave radiation in the (x,y) plane can be quantitatively characterized for each propagating mode. Finally, if required, summation of all mode contributions can be performed in order to obtain the total 3D displacement field in the plate.

III. APPLICATION TO A KNOWN EXCITATION

A. Problem description

In order to verify its pertinence and predicting ability, the modeling procedure will be applied to a simple case. We will consider a 3-mm-thick aluminum plate subject to a uniform pressure p_0 over the excitation area S as described in Fig. 1.

In such a simple situation, solution to the related 2D problem can be processed analytically. Details of the derivation can be found in Refs. 3 and 5. This yields

$$A_m = \frac{j\omega p_0 u_{mz}^*}{2kP_{mm}} \sin\left(\frac{kW}{2}\right) \tag{10}$$

with

$$P_{mm} = \frac{\omega}{2} \mathrm{Im}\left[\int_{-d/2}^{d/2} (u_{mx}^* \sigma_{mxx} + u_{mz}^* \sigma_{mxz}) dz\right], \tag{11}$$

where Im and * denote the imaginary part and the complex conjugate, respectively. u_m and σ_m are the modal displacement and stress tensors for the considered Lamb mode.

Though Eq. (9) has a meaning in a harmonic case only, transient excitations can be very classically dealt with using Fourier inversion.[5,13]

B. Results and comments

The first example corresponds to a two-cycle rectangular-windowed sinusoid at 150 kHz. At this frequency-thickness product and considering an out-of-plane excitation, the first antisymmetric mode A_0 is practically the only excited mode. The excitation area is defined by $W=8$ mm and $L=20$ mm. The resulting normal displacement at a distance $r=45$ mm is presented in Fig. 2(a). The dashed line represents the 2D solution, derived from the normal mode formalism. Application of Eq. (9), corresponding to the pseudo-3D solution, is represented by the solid line.

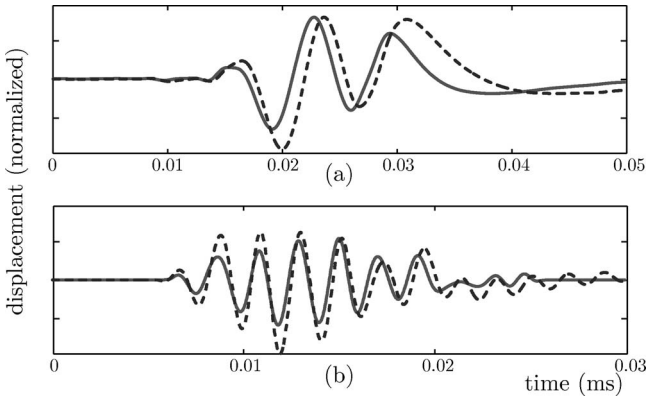


FIG. 2. Influence of 3D effects on the waveform shape (2D solution in dashed line, 3D solution in solid line). (a) Transducer size 8×20 mm, central frequency 150 kHz, $r=45$ mm, $\theta=0°$. (b) Transducer size 6×20 mm, central frequency 475 kHz, $r=28$ mm, $\theta=0°$.

A second demonstrative example is presented in Fig. 2(b). It corresponds to a multi-modal case. The excitation is a five-cycle sine-windowed sinusoid at 475 kHz and the source dimensions are $W=6$ mm and $L=20$ mm. The considered propagation distance is $r=28$ mm. All three modes: A_0, S_0 and A_1 are excited, though with different modal amplitudes. Here again, the solid and dashed lines represent the 3D and 2D results, respectively.

As shown here, taking into account the 3D behavior can lead to significant changes in the received waveforms. This can be easily interpreted from the form of Eq. (9). In addition to a mere amplitude decay proportional to $1/\sqrt{r}$ and related to the beam divergence in the (x,y) plane, the $\sqrt{k}$ factor disturbs the dispersive effects in the wave propagation. Indeed, components with low frequencies tend to be attenuated whereas higher frequency components tend to be amplified. Therefore the frequential reconstruction of the signal naturally yields modified waveforms. Besides, since S_0 and A_1 have smaller wave numbers than A_0, they appear more attenuated in the 3D case than in the 2D one. This results logically in a visible modification of the global waveform.

C. Comparison to numerical predictions

For quantitative validation purpose, both examples will be compared to results from a 3D finite element modeling (FEM). The excitation is modelized by applying normal point forces to the surface nodes in the corresponding area. The finite element mesh definition is guided by two, rather contradictory, criteria. First, the lateral dimensions of the plate have to be large enough so that no reflection pollutes the incident wave packets. Then, the elements should be numerous enough so that their sizes are smaller than a quarter of the shorter wavelength.[14] This ordinarily results in a rather cumbersome 3D mesh and, subsequently, in a huge memory and time consumption. Therefore, only a limited number of test cases can be reasonably treated in this way.

The obtained results are shown in Figs. 3(a) and 3(b).

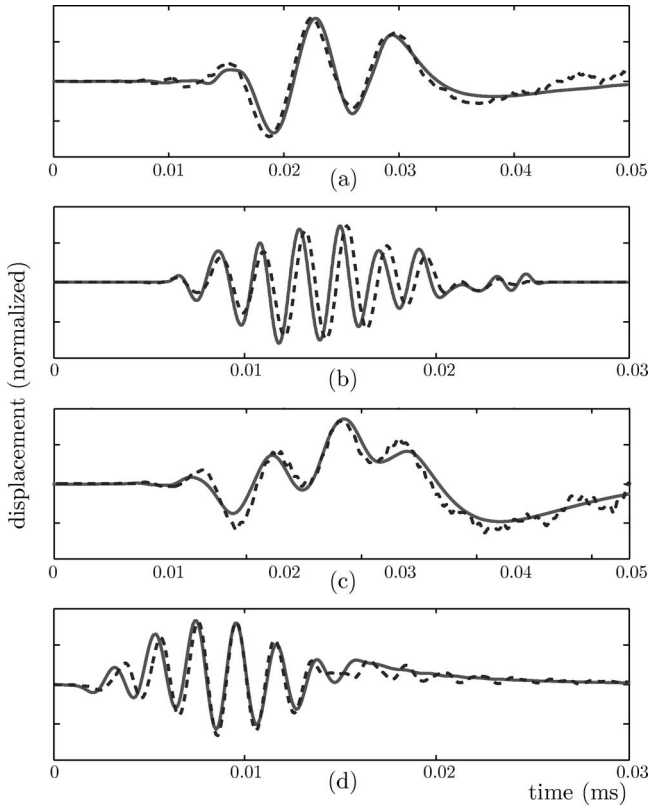


FIG. 3. Comparison of pseudo-3D solutions (solid) to 3D-FEM results (dashed). (a) Transducer size 8×20 mm, central frequency 150 kHz, $r=45$ mm, $\theta=0°$. (b) Transducer size 6×20 mm, central frequency 475 kHz, $r=28$ mm, $\theta=0°$. (c) Transducer size 8×20 mm, central frequency 150 kHz, $r=45$ mm, $\theta=50°$. (d) Transducer size 6×20 mm, central frequency 475 kHz, $r=12$ mm, $\theta=35°$.

The dashed lines correspond to the FEM results whereas the solid lines represent the same pseudo-3D results as in the previous section.

As can be seen, there is a very good agreement between the two sets of curves. Both the waveform shapes as well as the relative mode amplitudes are correctly predicted. Additionally, in order to provide validation for off-axis predictions, some results for $\theta\neq0$ are presented in Figs. 3(c) and 3(d). These results correspond to the same excitation cases as in Figs. 3(a) and 3(b), respectively. The first curve [Fig. 3(c)] is obtained at $r=45$ mm, $\theta=50°$. The second one [Fig. 3(c)] corresponds to $r=12$ mm, $\theta=35°$. These comparisons clearly demonstrate the ability of the hybrid pseudo-3D technique to provide quantitative results with a satisfying accuracy.

IV. CONCLUSION

This work was concerned with the problem of plate wave generation by a surface-mounted transducer with finite dimensions. In many practical cases, the 3D source effects cannot be neglected without leading to significantly erroneous results. As it has been verified indeed, both the waveform shapes and the mode amplitudes can be strongly affected by the finite length of the emitter.

Relying on existing works, it has been shown how the 3D characterization could be achieved by merely applying a corrective term to a 2D result. The formulation introduced allows the use of the normal mode expansion technique for solving the 2D problem, which results in a very convenient, polyvalent, and easy-to-use model. Comparison with results provided by a full-3D finite element computation have demonstrated the quantitative validity of the technique.

Besides, in previous works, a 2D hybrid FEM-normal mode technique has been successfully applied and has already been subject to adaptations to physical effects of growing complexity (anisotropy, attenuation,…). Then, the introduction of such effects into the pseudo-3D modeling could be considered without starting from the basis again.

[1]R. S. C. Monkhouse, P. D. Wilcox, and P. Cawley, "Flexible interdigital pvdf lamb wave transducers for the development of smart structures," Rev. Prog. Quant. Nondestr. Eval. **16**, 877–884 (1997).
[2]J.-B. Ihn and F.-K. Chang, "Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: I. diagnostics," Smart Mater. Struct. **13**, 609–620 (2004).
[3]E. Moulin, J. Assaad, C. Delebarre, and D. Osmont, "Modeling of lamb waves generated by integrated transducers in composite plates using a coupled finite element-normal modes expansion method," J. Acoust. Soc. Am. **107**, 87–94 (2000).
[4]S. Grondel, C. Paget, C. Delebarre, J. Assaad, and K. Levin, "Design of optimal configuration for generating a_0 lamb mode in a composite plate using piezoceramic transducers," J. Acoust. Soc. Am. **112**, 84–90 (2002).
[5]L. Duquenne, E. Moulin, J. Assaad, and S. Grondel, "Transient modeling of lamb waves generated in viscoelastic materials by surface bonded piezoelectric transducers," J. Acoust. Soc. Am. **116**, 133–141 (2004).
[6]E. Moulin, N. Bourasseau, J. Assaad, and C. Delebarre, "Directivity of integrated piezoelectric lamb wave sources," in *Proc. IEEE Ultrason. Symp.* (IEEE, Atlanta, USA, 2001), pp. 1081–1084.
[7]E. Moulin, N. Bourasseau, J. Assaad, and C. Delebarre, in "Lamb wave beam-steering for integrated health monitoring applications," in *Proc. SPIE* (SPIE, San Diego, USA, 2003), pp. 124–131.
[8]L. R. F. Rose and C. H. Wang, "Mindlin plate theory for damage detection: Source solutions," J. Acoust. Soc. Am. **116**, 154–171 (2004).
[9]I. A. Viktorov, *Rayleigh and Lamb Waves* (Plenum, New York, 1967).
[10]J. J. Ditri, A. Pilarski, B. Pavlakovic, and J. L. Rose, "Generation of guided waves in a plate by axisymmetric normal surface loading," Rev. Prog. Quant. Nondestr. Eval. **13**, 133–140 (1994).
[11]P. D. Wilcox, "Lamb wave inspection of large structures using permanently attached transducers," Ph.D. thesis, Imperial College, London, UK, 1998.
[12]B. A. Auld, *Acoustic Fields and Waves in Solids*, 2nd ed. (Krieger, Malabar, FL, 1990).
[13]L. Duquenne, E. Moulin, J. Assaad, and C. Delebarre, "Transient modeling of lamb wave generation by surface bonded piezoelectric transducers," in *Proc. SPIE Vol. 4763* (SPIE, Presqu'île de Giens, France, 2002), pp. 187–193.
[14]O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method*, 5th ed. (Butterworth-Heinemann, Oxford, MA, 2000).

Time-dependent stochastic inversion in acoustic travel-time tomography of the atmosphere

Sergey N. Vecherin
Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003

Vladimir E. Ostashev
NOAA/Earth System Research Laboratory, Boulder, Colorado 80305 and Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003

George H. Goedecke
Department of Physics, New Mexico State University, Las Cruces, New Mexico 88003

D. Keith Wilson
U.S. Army Engineer Research and Development Center, Hanover, New Hampshire 03755

Alexander G. Voronovich
NOAA/Earth System Research Laboratory, Boulder, Colorado 80305

(Received 4 August 2005; revised 3 February 2006; accepted 7 February 2006)

Stochastic inversion is a well known technique for the solution of inverse problems in tomography. It employs the idea that the propagation medium may be represented as random with a known spatial covariance function. In this paper, a generalization of the stochastic inverse for acoustic travel-time tomography of the atmosphere is developed. The atmospheric inhomogeneities are considered to be random, not only in space but also in time. This allows one to incorporate tomographic data (travel times) obtained at different times to estimate the state of the propagation medium at any given time, by using spatial-temporal covariance functions of atmospheric turbulence. This increases the amount of data without increasing the number of sources and/or receivers. A numerical simulation for two-dimensional travel-time acoustic tomography of the atmosphere is performed in which travel times between sources to receivers are calculated, given the temperature and wind velocity fields. These travel times are used as data for reconstructing the original fields using both the ordinary stochastic inversion and the proposed time-dependent stochastic inversion algorithms. The time-dependent stochastic inversion produces a good match to the specified temperature and wind velocity fields, with average errors about half those of the ordinary stochastic inverse. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2180535]

PACS number(s): 43.20.Dk, 43.28.We, 43.28.Vd [AIT] Pages: 2579–2588

I. INTRODUCTION

Acoustic tomography is widely used in physics, technology, and medicine for the remote sensing of inhomogeneous media. When applied to the atmosphere, acoustic tomography allows one to estimate (reconstruct) temperature and wind velocity within a tomographic volume or area.[1–4] The practical realization of acoustic tomography in an atmospheric boundary layer was reported in Refs. 5–10. In these tomography experiments, the sources and receivers of sound were located on masts several meters above the ground along the perimeter of a tomographic area that was a square or rectangular with side lengths of several hundred meters. The sound travel times between all pairs of sources and receivers were measured and used as input data for inverse algorithms to estimate temperature and wind velocity fields within a horizontal slice. This kind of tomography is called travel-time tomography. The inverse algorithms used were the stochastic inversion (SI) approach[5] and the simultaneous iterative reconstruction technique.[6–10] In the present paper, a generalization of the SI approach in travel-time acoustic tomography of the atmosphere is developed that allows one to effectively increase the number of data without increasing the number of sources and receivers.

The idea of travel-time acoustic tomography of the atmosphere is based on the fact that the time required for sound to propagate through a certain volume depends on the adiabatic sound speed (and hence on temperature) and wind velocity within that volume. More specifically, the travel time t_i^{tr} of the sound impulse propagation along the ith ray can be expressed as the following integral along the path L_i of this ray (e.g., Ref. 11):

$$t_i^{\mathrm{tr}} = \int_{L_i} \frac{dl}{u_g(\mathbf{R})}. \tag{1}$$

Here, $u_g(\mathbf{R})$ is the group velocity of a sound impulse, $\mathbf{R}$ is the position vector in three dimensions, and $i=1,2,\ldots,I$, where I is the number of ray paths of sound impulses simultaneously propagating through this volume from sources to receivers. It can be shown[11] that, in the presence of wind, the u_g can be expressed as $u_g=[c_{\mathrm{L}}^2(\mathbf{R})+V^2(\mathbf{R})+2c_{\mathrm{L}}(\mathbf{R})\mathbf{n}_i(\mathbf{R})\cdot\mathbf{V}(\mathbf{R})]^{1/2}$. Here $\mathbf{n}_i$ is the unit vector normal to

the wave front of a sound wave, $\mathbf{V}$ is the vector of wind velocity, and c_L is the Laplace adiabatic sound speed that relates to the acoustic virtual temperature T_{av} by

$$c_L^2 = \gamma R_a T_{av}, \tag{2}$$

where $\gamma \approx 1.41$ is the ratio of the specific heats and R_a is the universal gas constant for dry air. It is necessary to take into account the specific humidity of the air q to obtain the values of the actual thermodynamic temperature T_{th} from the acoustic virtual temperature by using the following relationship:[11] $T_{av} \cong T_{th}(1+0.511q)$. Note that the acoustic virtual temperature differs from the virtual temperature used by atmospheric scientists. When deriving Eq. (1), it is assumed that u_g does not depend on time during the propagation of sound impulses. Since in the atmosphere $c_L \gg |\mathbf{V}|$, this assumption is valid for tomographic arrays of order of several hundred meters considered in Refs. 5–10 and in Sec. IV. The goal of acoustic travel-time tomography is to estimate the fields $T_{av}(\mathbf{R})$ and $\mathbf{V}(\mathbf{R})$ within a tomographic volume, given the travel times t_i^{tr} and the locations of sources and receivers.

There are different techniques available to solve this problem. Many of them pursue the goal of finding a solution that, after substitution back into Eq. (1), yields travel times as close to the measured ones as possible. These methods are quite reliable and accurate if there are more data points (e.g., travel times) than unknown model values (e.g., total number of points at which the fields are to be reconstructed). However, this condition does not hold for atmospheric tomography unless the reconstruction is performed with low resolution.[2,6–10,12,13] In the opposite case, when the number of unknowns is greater than the number of available data points, such techniques cannot provide a unique solution, and some additional restrictions must be imposed on the sought fields. Since it is unknown in advance whether the actual fields being estimated obey such restrictions, these techniques may yield a spurious solution that perfectly matches the data but has little to do with the real fields.

In contrast, the SI approach is based on the idea that, using the available data, one seeks fields that have the minimum average deviation from the real ones. This approach also requires additional information about the sought fields, namely, the sought fields are treated as random ones with known spatial covariance functions.[1,3,5,14] Although the actual covariance functions in the turbulent atmosphere are not exactly known, they can be approximated by those corresponding to the von Kármán or Gaussian spectra of turbulence.[11,15,16] Note that other techniques that partition the fields into constant-valued grid cells implicitly assume step-like covariance functions, such that the field values are perfectly correlated within a grid cell and completely uncorrelated outside. Such functions are much less plausible than those used in the SI approach. The applicability and advantages of the SI approach in travel-time tomography of the atmosphere with high resolution are demonstrated in Refs. 3 and 5. A disadvantage of SI is that it does not use the fact that the sought fields are correlated not only in space but also in time.

In this paper, a generalization of the SI approach for acoustic travel-time tomography of the atmosphere is developed that allows one to use the data obtained at different times to reconstruct the temperature and wind velocity fields within a tomographic volume at any particular time, by using spatial-temporal covariance functions of these fields. This generalization is called time-dependent stochastic inversion (TDSI).

The general idea of using spatial-temporal covariance functions in SI is known in the literature. For example, it has been successfully used in satellite altimetry of the ocean surface, e.g., Refs. 17 and 18. In altimetry, the deviations of the ocean surface height from its average level are known at certain times and spatial points (along the satellite tracks). The problem is to estimate these deviations at other points (in space and time). Thus, this problem can be formulated as the space-time interpolation problem. There are many different techniques to interpolate the data. However, SI interpolates them in such a way that the covariance of the resulting fields holds true. The application of SI for these purposes was proposed in Ref. 19.

In contrast to satellite altimetry, travel-time tomography cannot be formulated as an interpolation problem, since the fields that are subject to estimation are unknown everywhere. Therefore, the mathematical formalism of TDSI in acoustic travel-time tomography of the atmosphere differs from that in altimetry. In the present paper, this mathematical formalism is developed. To verify that the TDSI approach improves the quality of the reconstruction, a numerical experiment was carried out for two-dimensional (2-D) travel-time acoustic tomography of a horizontal atmospheric layer.

The paper is organized as follows. In Sec. II, a general theory of 2-D TDSI is developed and a formula for the optimal stochastic inversion matrix is derived. Calculations of the covariance matrices that appear in the inversion matrix are presented in Sec. III. In Sec. IV, the numerical experiment of 2-D travel-time acoustic tomography of the atmosphere is described. Some aspects of the SI and TDSI algorithms are discussed in Sec. V. The conclusions are presented in Sec. VI.

II. THEORY

In this section, the theory of 2-D TDSI is developed. There are three steps that precede this development. First, it is taken into account that T_{av} and $\mathbf{V}$ depend not only on the spatial coordinates but also on time t. Therefore, t_i^{tr} depends on time t as well. Second, Eq. (1) is linearized due to specific conditions in the atmosphere, and, with the same degree of accuracy, the travel paths L_i are approximated by straight lines. Third, it is shown how to reconstruct the spatial mean values of the temperature and wind velocity fields within a tomographic area with the help of the conventional least square estimation. Finally, the problem for the TDSI approach in the reconstruction of temperature and wind velocity fluctuations is stated and a general solution of this problem is obtained.

A. Linearization

Let $\tilde{u}(\mathbf{r},t)$ and $\tilde{v}(\mathbf{r},t)$ be x and y components of the two-dimensional vector of wind velocity $\mathbf{V}$:

$$\mathbf{V}(\mathbf{r},t) = \tilde{u}(\mathbf{r},t)\mathbf{e}_x + \tilde{v}(\mathbf{r},t)\mathbf{e}_y. \tag{3}$$

Here t is time, a 2-D vector $\mathbf{r}$ specifies a spatial point within a tomographic area with the Cartesian coordinates (x,y), and $\mathbf{e}_x$ and $\mathbf{e}_y$ are the unit vectors along the x and y axes, respectively. The adiabatic sound speed c_L, temperature T_av, and wind velocity fields within a tomographic area at time t can be represented as sums of their spatial average values $c_0(t)$, $T_0(t)$, $u_0(t)$, and $v_0(t)$ and their fluctuations $c(\mathbf{r},t)$, $T(\mathbf{r},t)$, $u(\mathbf{r},t)$, and $v(\mathbf{r},t)$:

$$\begin{aligned} &c_\mathrm{L}(\mathbf{r},t) = c_0(t) + c(\mathbf{r},t), \quad T_\mathrm{av}(\mathbf{r},t) = T_0(t) + T(\mathbf{r},t), \\ &\tilde{u}(\mathbf{r},t) = u_0(t) + u(\mathbf{r},t), \quad \tilde{v}(\mathbf{r},t) = v_0(t) + v(\mathbf{r},t). \end{aligned} \tag{4}$$

Since in the atmosphere the absolute values of the adiabatic sound speed fluctuations and the wind components are much smaller than the spatial average value of c_0, Eq. (1) can be linearized to the first order of these fluctuations[2] and the travel paths of sound can be approximated by straight lines:[5,11]

$$\begin{aligned} t_i^\mathrm{tr}(t) = \varepsilon_i(t) + \frac{L_i}{c_0(t)}\left(1 - \frac{u_0(t)\cos\varphi_i + v_0(t)\sin\varphi_i}{c_0(t)}\right) \\ - \frac{1}{c_0^2(t)}\int_{L_i} dl\left(\frac{c_0(t)}{2T_0(t)}T(\mathbf{r},t) + u(\mathbf{r},t)\cos\varphi_i \right. \\ \left. + v(\mathbf{r},t)\sin\varphi_i\right), \end{aligned} \tag{5}$$

where $i=1,2,\ldots,I$ denotes the ray's path number, I is the total number of travel paths, L_i is the length of the ith ray path, $\mathbf{r}\in L_i$, and φ_i is the angle of the ith ray relative to the positive direction of the x axis. Note that the noise ε_i in the ith measured travel time was added to the right-hand side of Eq. (5). Furthermore, when deriving this equation the following relationship between the fluctuations of temperature and adiabatic sound speed was used: $c(\mathbf{r},t) = [c_0(t)/2T_0(t)]T(\mathbf{r},t)$. This relationship follows from Eqs. (4) and (2), with accuracy up to the first order of these fluctuations. The inverse problem, which is studied in this paper, is to reconstruct c_0, T_0, T, u, and v, given t_i^tr, L_i, and φ_i.

B. Reconstruction of mean fields

The mean values T_0, u_0, and v_0 within a tomographic area at any time can be reconstructed with the help of the least square estimation by using the travel times obtained at the same moment of time. For this purpose one should set the fluctuating parts of these fields to zero, so that the integral in Eq. (5) vanishes. Then Eq. (5) can be rewritten in matrix notation:

$$\mathbf{G}\mathbf{f} = \mathbf{b}, \tag{6}$$

where the elements of the column vector $\mathbf{b}$ are known and given by $b_i = t_i^\mathrm{tr}(t)/L_i$, the unknown column vector $\mathbf{f}$ has three elements, $f_1 = 1/c_0(t)$, $f_2 = u_0(t)/c_0^2(t)$, $f_3 = v_0(t)/c_0^2(t)$, and the matrix $\mathbf{G}$ is given by

$$\mathbf{G} = \begin{bmatrix} 1 & -\cos\varphi_1 & -\sin\varphi_1 \\ \vdots & \ddots & \vdots \\ 1 & -\cos\varphi_I & -\sin\varphi_I \end{bmatrix}. \tag{7}$$

Then, using the I elements of the vector $\mathbf{b}$ (I is assumed to be greater than 3), one should solve the overdetermined problem for the three unknowns (the elements of the vector $\mathbf{f}$) using the least square estimation:

$$\mathbf{f} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{b}, \tag{8}$$

extract from them the values of c_0, u_0, and v_0, and calculate T_0 using Eq. (2).

C. Time-dependent stochastic inversion

Once the spatial mean values of temperature, adiabatic sound speed, and wind velocity fields within a tomographic area are known, it is worthwhile to introduce the column vector of data $\boldsymbol{d}(t)$ obtained at time t with elements

$$d_i(t) = L_i[c_0(t) - u_0(t)\cos\varphi_i - v_0(t)\sin\varphi_i] - c_0^2(t)t_i^\mathrm{tr}(t) + \xi_i(t), \tag{9}$$

where noise $\xi_i(t)$ includes the errors of travel time measurements $\varepsilon_i(t)$ and errors in the estimation of T_0, c_0, u_0, and v_0. Using Eq. (5), this data vector can be expressed as $\boldsymbol{d}(t) = \boldsymbol{d}_0(t) + \boldsymbol{\xi}(t)$, where the elements of the noise-free data column vector $\boldsymbol{d}_0(t)$ are given by

$$d_{0i}(t) = \int_{L_i} dl\left(\frac{c_0(t)}{2T_0(t)}T(\mathbf{r},t) + u(\mathbf{r},t)\cos\varphi_i + v(\mathbf{r},t)\sin\varphi_i\right). \tag{10}$$

The problem now is to reconstruct the fields of fluctuations $T(\mathbf{r},t_0)$, $u(\mathbf{r},t_0)$, and $v(\mathbf{r},t_0)$ at any chosen time t_0. Thus, the column vector of models at this time moment t_0 is given by

$$\begin{aligned} \mathbf{m}(t_0) = [&T(\mathbf{r}_1,t_0);\ldots;T(\mathbf{r}_J,t_0); \\ &u(\mathbf{r}_1,t_0);\ldots;u(\mathbf{r}_J,t_0);v(\mathbf{r}_1,t_0);\ldots;v(\mathbf{r}_J,t_0)], \end{aligned} \tag{11}$$

where J is the number of spatial points within the tomographic area at which the fields are being reconstructed; here and in what follows, the semicolon between elements denotes that these elements are arranged in a column.

Let us assume that one can collect data $(N+1)$ times during the time $N\tau$ by performing sound scans of the tomographic area and forming the vector $\boldsymbol{d}$ at each scan. Suppose that the last scan was performed at the time t, which can be earlier, equal to, or later than the time t_0 at which one tries to reconstruct the models. In this paper it is assumed that these scans are performed at equal time intervals τ, although such an assumption is not necessary. The temperature and wind velocity fields in this tomographic area are changing with time, which yields different data vectors at each scan. In this case, at the time t there will be data available from $(N+1)$ scans. That is, at the moment t one can form the vector $\mathbf{d}$ of all available data:

$$\mathbf{d} = [\boldsymbol{d}(t - N\tau); \boldsymbol{d}(t - N\tau + \tau); \ldots; \boldsymbol{d}(t)]. \tag{12}$$

Here $\mathbf{d}$ is a column vector containing the $(N+1)$ column-vectors $\boldsymbol{d}(t-n\tau)$, each of length I, where $n=0,1,\ldots,N$. Therefore, the total length of the data vector $\mathbf{d}$ is $(N+1)I$.

1. Problem statement

Using the data vector $\mathbf{d}$ given by Eq. (12), it is necessary to find a linear estimation $\hat{\mathbf{m}}(t_0)$ of unknown models $\mathbf{m}(t_0)$ at time t_0. The case $t_0 > t$ corresponds to advanced estimation; the case $t_0 < t$ corresponds to posterior estimation; and $t_0 = t$ corresponds to a real time reconstruction. In all these cases, one can use available data obtained from all scans to estimate the unknown fields. If one uses only the data obtained at moment t_0 to estimate the models $\mathbf{m}(t_0)$ at the same moment, that would correspond to standard SI in atmospheric travel-time tomography.[5]

In TDSI, the estimation of the unknown fields is sought in the same form as in standard SI:

$$\hat{\mathbf{m}}(t_0) = \mathbf{A}\mathbf{d}, \tag{13}$$

where unknown elements a_{jk} of matrix $\mathbf{A}$ must be determined. Here $j=1,2,\ldots,3J$, where $3J$ is the length of models, and $k=1,2,\ldots,(N+1)I$. We now introduce the column vector of discrepancy $\boldsymbol{\epsilon}$ between the true and reconstructed fields at time t_0:

$$\boldsymbol{\epsilon}_j = \hat{m}_j(t_0) - m_j(t_0). \tag{14}$$

To find the elements a_{jk}, let us require that they give the minimum of the elements of the mean square errors vector $\langle \boldsymbol{\epsilon}^2 \rangle$:

$$\langle \epsilon_j^2 \rangle \mapsto \min_{\{a_{jk}\}} \langle \boldsymbol{\epsilon}_j^2 \rangle, \tag{15}$$

where parentheses $\langle \cdot \rangle$ denote the mathematical expectation. These mean square errors are the diagonal elements of the error covariance matrix $\mathbf{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$:

$$\mathbf{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} \equiv \langle \boldsymbol{\epsilon}\boldsymbol{\epsilon}^T \rangle, \tag{16}$$

where $\boldsymbol{\epsilon}$ is the column vector of discrepancy given by (14).

2. Solution

The matrix $\mathbf{A}$ that solves the problem stated above is given by the same formula as for ordinary SI, since the derivation of this formula does not depend on a particular structure of the models $\mathbf{m}$ and data $\mathbf{d}$ (see Appendix A):

$$\mathbf{A} = \mathbf{R}_{\mathbf{md}} \mathbf{R}_{\mathbf{dd}}^{-1}, \tag{17}$$

where $\mathbf{R}_{\mathbf{md}} \equiv \langle \mathbf{m}\mathbf{d}^T \rangle$ and $\mathbf{R}_{\mathbf{dd}} \equiv \langle \mathbf{d}\mathbf{d}^T \rangle$ are model-data and data covariance matrices.

Since the data and models have been chosen, as shown in Eqs. (12) and (11), these matrices have the following block structure:

$$\mathbf{R}_{\mathbf{md}} = [\mathbf{B}_{\boldsymbol{md}}(t_0, t - N\tau), \mathbf{B}_{\boldsymbol{md}}(t_0, t - N\tau + \tau), \ldots, \mathbf{B}_{\boldsymbol{md}}(t_0, t)], \tag{18}$$

$$\mathbf{R}_{\mathbf{dd}} = \begin{bmatrix} \mathbf{B}_{\boldsymbol{dd}}(t - N\tau, t - N\tau) & \mathbf{B}_{\boldsymbol{dd}}(t - N\tau, t - N\tau + \tau) & \cdots & \mathbf{B}_{\boldsymbol{dd}}(t - N\tau, t) \\ \mathbf{B}_{\boldsymbol{dd}}(t - N\tau + \tau, t - N\tau) & \mathbf{B}_{\boldsymbol{dd}}(t - N\tau + \tau, t - N\tau + \tau) & \cdots & \mathbf{B}_{\boldsymbol{dd}}(t - N\tau + \tau, t) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{\boldsymbol{dd}}(t, t - N\tau) & \mathbf{B}_{\boldsymbol{dd}}(t, t - N\tau + \tau) & \cdots & \mathbf{B}_{\boldsymbol{dd}}(t, t) \end{bmatrix}. \tag{19}$$

Here $\mathbf{B}_{\boldsymbol{md}}(t_1, t_2) \equiv \langle \mathbf{m}(t_1)\boldsymbol{d}^T(t_2) \rangle$ is the covariance matrix of size $[3J, I]$ between the models at time moment t_1 and data at moment t_2, $\mathbf{B}_{\boldsymbol{dd}}(t_1, t_2) \equiv \langle \boldsymbol{d}(t_1)\boldsymbol{d}^T(t_2) \rangle$ is the covariance matrix of size $[I, I]$ between data at moment t_1 and data at moment t_2, and the comma between the elements of the $\mathbf{R}_{\mathbf{md}}$ matrix denotes that the $\mathbf{B}_{\boldsymbol{md}}$ matrices are arranged in a row. Thus, the dimensions of the $\mathbf{R}_{\mathbf{md}}$ and $\mathbf{R}_{\mathbf{dd}}$ matrices are $[3J, (N+1)I]$ and $[(N+1)I, (N+1)I]$, correspondingly. Note that the $\mathbf{R}_{\mathbf{dd}}$ matrix is symmetric.

It is assumed that noise in the data $\boldsymbol{\xi}(t)$ is independent of the sought fields and the noise-free data vectors, is not correlated with itself for different time moments and for different paths, and can be described by the normal distribution $N(0, \sigma_\xi)$. Therefore, in the presence of noise,

$$\mathbf{R}_{\mathbf{md}} = \mathbf{R}_{\mathbf{md}_0}, \tag{20}$$

$$\mathbf{R}_{\mathbf{dd}} = \mathbf{R}_{\mathbf{d}_0\mathbf{d}_0} + \sigma_\xi^2 \mathbf{I}, \tag{21}$$

where $\mathbf{I}$ is the identity matrix. Since noise in the data can easily be taken into account by these formulas, further considerations will be focused on the noise-free data $\mathbf{d}_0$.

It can be shown (see Appendix B) that the error covariance matrix $\mathbf{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$ that corresponds to such an estimation of models [see Eq. (13)] with the optimal matrix $\mathbf{A}$ given by Eq. (17) is

$$\mathbf{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}} = \mathbf{R}_{\mathbf{mm}} - \mathbf{R}_{\mathbf{md}} \mathbf{R}_{\mathbf{dd}}^{-1} \mathbf{R}_{\mathbf{md}}^T, \tag{22}$$

where $\mathbf{R}_{\mathbf{mm}} \equiv \langle \mathbf{m}(t_0)\mathbf{m}^T(t_0) \rangle$ is the models covariance matrix at time t_0. The elements of the main diagonal of the $\mathbf{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}$ matrix are equal to expected mean square errors of the reconstruction at each point. They can be calculated without knowledge of the original fields. These averaged errors can differ from the errors of the reconstruction of a particular realization of the original fields.

Note that the developed mathematical formalism of TDSI is valid for statistically nonstationary and inhomogeneous random fields.

III. COVARIANCE MATRICES

Since the optimal stochastic inverse operator **A** given by Eq. (17) is determined in terms of the $\mathbf{R_{md}}$ and $\mathbf{R_{dd}}$ matrices, it is worthwhile to consider some important particular cases when these matrices can be calculated explicitly. These considerations will be carried out taking 2-D travel-time tomography of the atmosphere as an example.

A. 2-D travel-time tomography

In the case of travel-time tomography, one can use the linear relationship between $\boldsymbol{d}_0$ and the sought fields given by Eq. (10) to obtain an expression for the covariance matrix $\mathbf{B}_{\mathbf{m}\boldsymbol{d}_0}(t_1,t_2)$ between the models at time t_1 and the noise-free data at time t_2:

$$
\begin{aligned}
[\mathbf{B}_{\mathbf{m}\boldsymbol{d}_0}(t_1,t_2)]_{ji} &= \langle m_j(t_1) d_{0i}(t_2)\rangle \\
&= \int_{L_i} dl \left(\frac{c_0(t_2)}{2T_0(t_2)} \langle m_j(t_1) T(\mathbf{r},t_2)\rangle + \langle m_j(t_1) u(\mathbf{r},t_2)\rangle \cos\varphi_i + \langle m_j(t_1) v(\mathbf{r},t_2)\rangle \sin\varphi_i \right) \\
&= \begin{cases}
\displaystyle\int_{L_i} dl \left(\frac{c_0(t_2)}{2T_0(t_2)} B_{TT}(\mathbf{r}_j,t_1;\mathbf{r},t_2) \right), & \text{if } 1 \leqslant j \leqslant J, \\
\displaystyle\int_{L_i} dl (B_{uu}(\mathbf{r}_j,t_1;\mathbf{r},t_2)\cos\varphi_i + B_{uv}(\mathbf{r}_j,t_1;\mathbf{r},t_2)\sin\varphi_i), & \text{if } J+1 \leqslant j \leqslant 2J, \\
\displaystyle\int_{L_i} dl (B_{vu}(\mathbf{r}_j,t_1;\mathbf{r},t_2)\cos\varphi_i + B_{vv}(\mathbf{r}_j,t_1;\mathbf{r},t_2)\sin\varphi_i), & \text{if } 2J+1 \leqslant j \leqslant 3J,
\end{cases}
\end{aligned}
\tag{23}
$$

where $i=1,2,\ldots,I$, $j=1,2,\ldots,3J$, $\mathbf{r}\in L_i$, B_{TT}, B_{uu}, B_{vv}, B_{uv}, and B_{vu} are the spatial-temporal covariance functions of the corresponding fields marked as the subscripts, and the $\mathbf{r}_j$ are the chosen spatial points within the tomographic area at which the sought fields are reconstructed; these points stay fixed during the integration.

Similarly, an expression for the covariance matrix $\mathbf{B}_{\boldsymbol{d}_0\boldsymbol{d}_0}(t_1,t_2)$ between the noise-free data at time t_1 and time t_2 is given by

$$
\begin{aligned}
[\mathbf{B}_{\boldsymbol{d}_0\boldsymbol{d}_0}(t_1,t_2)]_{ip} = \langle d_{0i}(t_1) d_{0p}(t_2)\rangle = \int_{L_i} dl \int_{L_p} dl' \Bigg\{ & \frac{c_0(t_1)c_0(t_2)}{4T_0(t_1)T_0(t_2)} B_{TT}(\mathbf{r},t_1;\mathbf{r}',t_2) + B_{uu}(\mathbf{r},t_1;\mathbf{r}',t_2)\cos\varphi_i\cos\varphi_p \\
& + B_{vv}(\mathbf{r},t_1;\mathbf{r}',t_2)\sin\varphi_i\sin\varphi_p + B_{uv}(\mathbf{r},t_1;\mathbf{r}',t_2)\cos\varphi_i\sin\varphi_p + B_{vu}(\mathbf{r},t_1;\mathbf{r}',t_2)\cos\varphi_p\sin\varphi_i \Bigg\},
\end{aligned}
\tag{24}
$$

where $i,p=1,2,\ldots,I$, $\mathbf{r}\in L_i$, $\mathbf{r}'\in L_p$. Note that $B_{vu}(\mathbf{r},t_1;\mathbf{r}',t_2)=B_{uv}(\mathbf{r}',t_2;\mathbf{r},t_1)$. When deriving Eqs. (23) and (24), it is assumed that $B_{Tu}=B_{Tv}\equiv 0$.

Once the matrices $\mathbf{B}_{\mathbf{m}\boldsymbol{d}_0}(t_1,t_2)$ and $\mathbf{B}_{\boldsymbol{d}_0\boldsymbol{d}_0}(t_1,t_2)$ are calculated, one should use Eqs. (18) and (19) to form the $\mathbf{R}_{\mathbf{md}_0}$ and $\mathbf{R}_{\mathbf{d}_0\mathbf{d}_0}$ matrices and, then, Eq. (21) to take noise into account.

B. Stationary fields

If the temperature and wind velocity fields are statistically stationary, their covariance functions depend only on the difference of their temporal arguments:

$$
B(\mathbf{r},t_1;\mathbf{r}',t_2) = B(\mathbf{r},\mathbf{r}',\Delta t), \quad \Delta t \equiv t_2 - t_1. \tag{25}
$$

The notation B without subscripts stands for the covariance function of any two fields considered above. In this case, Eqs. (18) and (19) are modified as follows:

$$
\begin{aligned}
\mathbf{R_{md}} = [&\mathbf{B}_{\mathbf{m}\boldsymbol{d}}(t - N\tau - t_0), \mathbf{B}_{\mathbf{m}\boldsymbol{d}}(t - N\tau + \tau - t_0), \ldots, \\
&\mathbf{B}_{\mathbf{m}\boldsymbol{d}}(t - t_0)],
\end{aligned}
\tag{26}
$$

$$
\mathbf{R_{dd}} = \begin{bmatrix}
\mathbf{B}_{dd}(0) & \mathbf{B}_{dd}(\tau) & \cdots & \mathbf{B}_{dd}(N\tau) \\
\mathbf{B}_{dd}(-\tau) & \mathbf{B}_{dd}(0) & \cdots & \mathbf{B}_{dd}(N\tau - \tau) \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{B}_{dd}(-N\tau) & \mathbf{B}_{dd}(-N\tau + \tau) & \cdots & \mathbf{B}_{dd}(0)
\end{bmatrix}. \tag{27}
$$

Note that, for any Δt,

$$
\mathbf{B}_{dd}(\Delta t) = \mathbf{B}_{dd}^T(-\Delta t). \tag{28}
$$

It is sufficient to know only the first cell row of the matrix $\mathbf{R_{dd}}$ since it has the same matrices along each diagonal (a block Toeplitz-like structure). Also note that the $\mathbf{R_{dd}}$ matrix is independent of time t_0, i.e., for the given set of data it is

necessary to calculate this matrix only once to reconstruct fields at any time. Another interesting fact is that the matrices $\mathbf{B}_{dd}(0)$ on its main diagonal are equal to the covariance matrix $\mathbf{R}_{\mathbf{dd}}$ of ordinary SI.

C. Frozen turbulence

If turbulence may be considered frozen, each spatial point of the temperature and wind velocity fields is moving with a constant speed $\mathbf{U}$. In this case, the temperature field at time t_2 can be expressed in terms of the field at time t_1 by the following relationship:[20]

$$T(\mathbf{r},t_2) = T(\mathbf{r} - \mathbf{U}\,\Delta t, t_1). \tag{29}$$

The same relationship is valid for the turbulent wind velocity fields. This allows one to obtain the spatial-temporal covariance matrices, which are unknown in general, in terms of the spatial covariance matrices. The latter can be modeled by the covariance functions corresponding, for example, to the von Kármán or Gaussian spatial spectra of turbulence. For example, for the temperature field, we have

$$B_{TT}(\mathbf{r},t_1;\mathbf{r}',t_2) \equiv \langle T(\mathbf{r},t_1)T(\mathbf{r}',t_2)\rangle = \langle T(\mathbf{r},t_1)T(\mathbf{r}' - \mathbf{U}\,\Delta t, t_1)\rangle = B^s_{TT}(\mathbf{r},\mathbf{r}' - \mathbf{U}\,\Delta t), \tag{30}$$

where B^s_{TT} is the spatial covariance matrix of the temperature field. The last equality is true for statistically stationary fields. Similar equations can be derived for the spatial-temporal covariance functions of the other fields. Since frozen turbulence is a particular case of stationary fields, Eqs. (26)–(28), stay valid for this case as well. Thus, for frozen turbulence the spatial-temporal covariance matrices $\mathbf{R}_{\mathbf{md}}$ and $\mathbf{R}_{\mathbf{dd}}$ can be expressed in terms of known or inferred spatial covariance matrices.

The frozen turbulence hypothesis is widely used in meteorology and the study of wave propagation in random media, e.g., Refs. 20 and 21. This hypothesis is valid with a good accuracy for eddies in the inertial range of turbulence. However, the applicability of this hypothesis to eddies in the energy range is questionable.[22] Note that in the present paper, the frozen turbulence hypothesis is used to concretize the spatial-temporal covariance functions. TDSI can still be used if this hypothesis is invalid.

IV. NUMERICAL EXPERIMENT

In this section, a numerical experiment that demonstrates the significant improvement in the quality of temperature and wind velocity field reconstructions possible with TDSI is described. A snapshot of the temperature and wind velocity fields was created by large eddy simulation (LES),[23] which is widely used in meteorology (e.g., Refs. 24 and 25). LES numerically solves the Navier-Stokes equations for eddies large enough to be resolved by the spatial grid. The effect of smaller eddies on the resolved flow is parametrized. LES produces realistic wind velocity and temperature fields for flow processes that are well resolved by the grid.

Using the single snapshot of the LES temperature and wind velocity fields, the perfectly frozen turbulent fields with

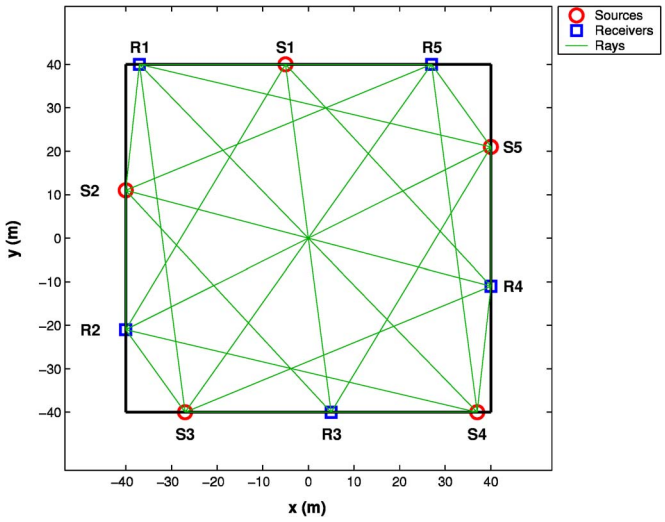


FIG. 1. (Color online) The layout of sources and receivers in the numerical experiment.

the velocity components U_x=4 m/s and U_y=0 m/s were created. The travel times were calculated each second (τ =1 s) using Eq. (5). Two-dimensional linear interpolation was employed to obtain the value of the integrand in Eq. (5) along the travel paths; then the integral was calculated numerically. These travel times were disturbed by white noise with standard deviation σ_ε=5 μs, which corresponds to relative errors of order 1% to 42% in the data $\boldsymbol{d}_0(t)$. The total number of such numerical scans was 18, and the problem was to estimate the temperature and wind velocity fields $T_{\text{av}}(\mathbf{r},t_0)$, $\tilde{u}(\mathbf{r},t_0)$, and $\tilde{v}(\mathbf{r},t_0)$ at time t_0=9 s. The time origin coincided with the first scan. The number of sources as well as receivers was 5, so that the number of rays I at each scan was 25. The tomographic area was a square of 80 by 80 meters. The layout of acoustic sources and receivers and the corresponding ray paths are shown in Fig. 1.

The reconstruction consisted of two stages. In the first stage the spatial mean values of the temperature and wind velocity fields $T_0(t_0)$, $u_0(t_0)$, and $v_0(t_0)$ within the tomographic area were reconstructed using the approach described in Sec. II B. The true and estimated spatial mean values of temperature and wind fields are presented in Table I.

One can see from Table I that the reconstruction of spatial mean fields was very accurate, the differences between true and estimated mean values was 0.14 K for the temperature field, 0.03 m/s for the x component of the wind velocity vector, and 0.01 m/s for its y component.

In the second stage, the fluctuations of the temperature and wind fields $T(\mathbf{r},t_0)$, $u(\mathbf{r},t_0)$, and $v(\mathbf{r},t_0)$ from their spatial mean values were reconstructed using either standard SI (only data at time t_0 were used, the total number of data

TABLE I. Actual and estimated spatial mean values of temperature and wind velocity fields.

Mean fields	T_0 (K)	u_0 (m/s)	v_0 (m/s)
True	301.73	3.09	1.73
Estimated	301.87	3.06	1.71

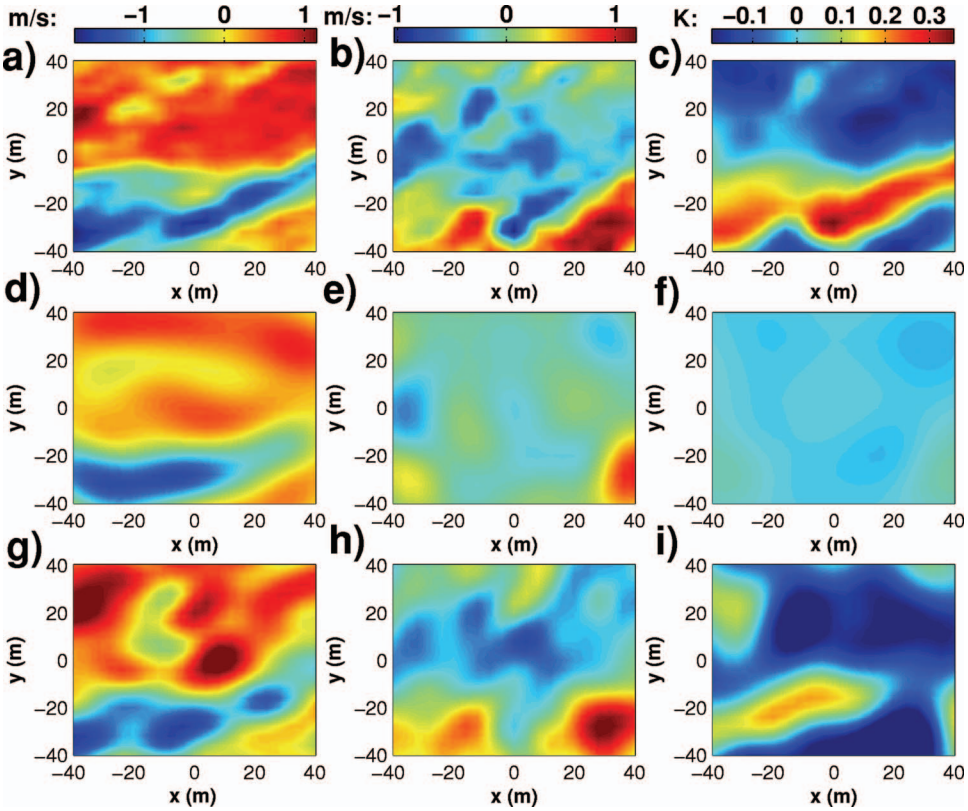


FIG. 2. The original and reconstructed temperature (K) and wind velocity (m/s) fields of fluctuations at time moment $t_0=9$ s. (a) Original u field. (b) Original v field. (c) Original T field. (d) The u field reconstructed by SI. (e) The v field reconstructed by SI. (f) The T field reconstructed by SI. (g) The u field reconstructed by TDSI. (h) The v field reconstructed by TDSI. (i) The T field reconstructed by TDSI.

points was equal to the number of rays $I=25$) or TDSI (data obtained at 18 different times were used, which resulted in 450 data points). To describe the spatial covariance of the temperature and wind velocity fields within the tomographic area, the following Gaussian covariance functions were used:[11]

$$B^s_{TT}(\mathbf{r},\mathbf{r}') = \sigma_T^2 \exp\left(-\frac{(\mathbf{r}-\mathbf{r}')^2}{l_T^2}\right), \tag{31}$$

$$B^s_{uu}(\mathbf{r},\mathbf{r}') = \sigma_u^2 \exp\left(-\frac{(\mathbf{r}-\mathbf{r}')^2}{l^2}\right)\left(1-\frac{(y-y')^2}{l^2}\right), \tag{32}$$

$$B^s_{vv}(\mathbf{r},\mathbf{r}') = \sigma_v^2 \exp\left(-\frac{(\mathbf{r}-\mathbf{r}')^2}{l^2}\right)\left(1-\frac{(x-x')^2}{l^2}\right), \tag{33}$$

$$B^s_{uv}(\mathbf{r},\mathbf{r}') = \sigma_u\sigma_v \exp\left(-\frac{(\mathbf{r}-\mathbf{r}')^2}{l^2}\right)\frac{(x-x')(y-y')}{l^2}, \tag{34}$$

where $\sigma_T, \sigma_u, \sigma_v$ are the standard deviations of the corresponding fields, l_T and l are their correlation lengths, $\mathbf{r}=(x,y)$, and $\mathbf{r}'=(x',y')$. Then, the relations similar to that given by Eq. (30) were used to obtain the spatial-temporal covariance functions of the sought fields. The noise-free matrices $\mathbf{R}_{\mathbf{md}_0}$ and $\mathbf{R}_{\mathbf{d}_0\mathbf{d}_0}$ were formed with the help of Eqs. (26) and (27), where the $\mathbf{B}_{\mathbf{m}d_0}$ and $\mathbf{B}_{d_0d_0}$ matrices were calculated using Eqs. (23) and (24); noise was taken into account by Eq. (21).

In the described numerical experiment, there were five parameters in the SI and TDSI algorithms that must be chosen: σ_T, σ_u, σ_v, l_T, and l. These parameters were estimated from the original LES fields: $\sigma_T=0.14$ K, $\sigma_u=0.72$ m/s, $\sigma_v=0.42$ m/s, and $l_T=l=15$ m.

The original and reconstructed fields were described by matrices of size 21 by 21. After all calculations were done, the fields were interpolated (2-D linear interpolation) for illustrative purposes. The fluctuating parts of the original fields are presented in Figs. 2(a)–2(c). Figures 2(d)–2(f) show the reconstructed fields using the SI approach. The reconstruction results of TDSI are given in Figs. 2(g)–2(i). As one can see, standard SI matched the contours of the sought u field fairly well, but the v and T fields were reconstructed poorly. In contrast, the time-dependent stochastic approach allowed the much more detailed and accurate reconstruction of all fields. To characterize the expected improvement of the reconstruction it is worthwhile to consider the expected errors of the reconstruction given by Eq. (22). It is convenient to normalize these mean square errors by the corresponding field variances so that their values lie in $[0, 1]$. If these normalized mean-squared errors (NMSE) are of order unity, the errors of reconstruction are of the order of variance of the original field, i.e., one has a poor reconstruction. Conversely, if NMSE are zeros, it is expected that the reconstructed and actual fields are identical. NMSE can be recalculated into root mean-square errors (RMSE). NMSE for SI are presented in Figs. 3(a)–3(c) while Figs. 3(d)–3(f) show NMSE for TDSI. To characterize the overall quality of the reconstruction, NMSE at each spatial point and corresponding RMSE were averaged over the tomographic area. These averaged values of NMSE and RMSE are presented in Table II.

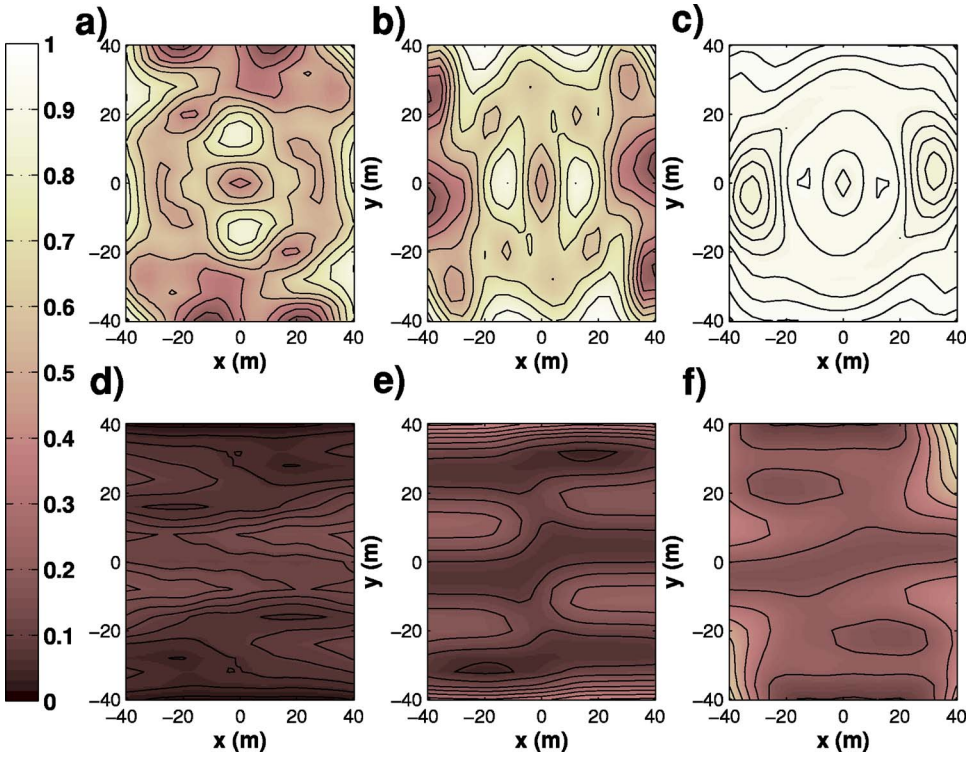


FIG. 3. (Color online) The expected normalized mean square errors. (a) The u field reconstructed by SI. (b) The v field reconstructed by SI. (c) The T field reconstructed by SI. (d) The u field reconstructed by TDSI. (e) The v field reconstructed by TDSI. (f) The T field reconstructed by TDSI.

It follows from Fig. 3 and Table II that, for TDSI, the average NMSE is about four times less for the T and v fields and seven times less for the u field than the corresponding NMSE for ordinary SI. This corresponds to a 46% reduction of the RMSE for the T field, a 52% reduction for the v fields, and a 62% reduction for the u field.

V. DISCUSSION

The idea of the stochastic approach is based on the assumption that covariance matrices of the sought fields are known. That means that in the case of the Gaussian covariance functions all five parameters (variances and correlation lengths of the sought fields) should be known in advance. In the practical implementation of travel-time tomography, however, these parameters are unknown. One can estimate these parameters by measuring the temperature and wind velocity fluctuations with the use of conventional meteorological sensors or using a turbulence similarity theory.

Furthermore, the covariance of the isotropic turbulence in the atmosphere is better described by the von Kármán spectrum rather than by the Gaussian spectrum. The latter was used in the numerical experiment because it was then possible to calculate the $\mathbf{R_{md}}$ matrix analytically and use only a single numeric integration to get the $\mathbf{R_{dd}}$ matrix. This significantly improved the accuracy and speed of the calculations. Since the actual fields were not described by the Gaussian covariance functions, the reconstruction quality improvement seems rather unexpected. However, there is a relatively simple explanation of this fact. In a certain spatial scale range, the covariance functions of atmospheric turbulence can be approximated by the Gaussian covariance functions with appropriate variances and correlation lengths.[27] Therefore, the Gaussian covariance functions can be used for SI and TDSI as approximations of the actual ones. It is expected that the use of the covariance function corresponding to the von Kármán spectrum of turbulence will improve the reconstruction.

TABLE II. Average expected errors of the reconstruction of temperature and wind velocity fields by SI and TDSI.

Fields	T		u		v	
Errors	NMSE	RMSE (K)	NMSE	RMSE (m/s)	NMSE	RMSE (m/s)
SI	0.96	0.13	0.56	0.53	0.63	0.33
TDSI	0.24	0.07	0.08	0.20	0.15	0.16

One of the advantages of the stochastic approach is that noise in the data plays, in a certain sense, a positive role by regularizing the $\mathbf{R_{dd}}$ matrix [see Eq. (21)]. Indeed, the presence of noise just adds an additional term to the main diagonal of the $\mathbf{R_{dd}}$ matrix, which improves its condition (keeps it invertible), although it does smooth the solution somewhat. This regularization is especially important for TDSI, where the condition of the matrix $\mathbf{R_{dd}}$ may be very poor and may lead to spurious solutions.

Figures 2(a)–2(i) show that the reconstruction of the temperature fluctuation field was not as good as that for the velocity fluctuations. The probable reason for this is a small effect of the temperature fluctuations on the travel times in comparison to the wind-velocity fluctuations. Indeed, the LES temperature fluctuations were in the range $[-0.15, 0.35]$ K while, for many meteorological problems, the acceptable error of temperature measurements is ± 0.3 K.[7,8] Note that to improve the reconstruction of temperature fluctuations in acoustic tomography, one can use reciprocal transmission of sound waves.[8,26]

The proposed TDSI algorithm also allows one to estimate the temporal mean values of the temperature and wind velocity fluctuations at any given spatial point. This might be important when the temporal changes of these fluctuations are negligible during the time interval $N\tau$. For this purpose, it is necessary to set all temporal arguments in the matrices $\mathbf{R_{md}}$ and $\mathbf{R_{dd}}$ equal to zero. Note that in this case the noise variance σ_ξ^2 should be enlarged, because the noise includes not only the measurement uncertainty but also the neglected variation of the original fields during the observed time interval.

VI. CONCLUSION

In this paper, a generalization of the stochastic inversion approach for tomographic problems in the atmosphere was developed. The key idea of this generalization is that, by using spatial-temporal covariance functions of temperature and wind velocity fields, one can use data obtained at different times to reconstruct these fields at any particular time. This allows one to significantly enlarge the effective amount of tomographic data, while keeping the total number of sources and/or receivers fixed. The efficiency of the developed method was demonstrated in a numerical experiment that simulated the acoustic 2-D travel-time tomography problem of the atmosphere. This numerical experiment showed a remarkable improvement of the reconstruction when one used TDSI instead of standard SI.

ACKNOWLEDGMENTS

This material is partly based upon work that was supported by the U.S. Army Research Office under Contracts No. DAAD19-03-1-0104 and No. DAAD19-03-1-0341. Also, we would like to thank E. G. Patton and P. P. Sullivan of the National Center for Atmospheric Research for providing the LES data. The LES was supported in part by the DoD High-Performance Computing Modernization Program. Finally, we would like to thank anonymous reviewers, whose comments allowed us to improve this paper.

APPENDIX A: OPTIMAL STOCHASTIC OPERATOR DERIVATION

Let $\mathbf{m}$ and $\mathbf{d}$ be two arbitrary random column-vectors of lengths M and D, correspondingly, with a known matrix of the second statistical moment $\mathbf{R_{md}} \equiv \langle \mathbf{m}\mathbf{d}^T \rangle$. Furthermore, let the matrix $\mathbf{R_{dd}} \equiv \langle \mathbf{d}\mathbf{d}^T \rangle$ also be known. Note that, in the case when $\langle \mathbf{m} \rangle = \mathbf{0}$, $\langle \mathbf{d} \rangle = \mathbf{0}$, the $\mathbf{R_{md}}$ and $\mathbf{R_{dd}}$ are the covariance matrices. The problem is to estimate the unknown vector $\mathbf{m}$ if a particular realization of the random vector $\mathbf{d}$ is given. To do this, one can seek for the estimation $\hat{\mathbf{m}}$ of the unknown vector $\mathbf{m}$ in the form

$$\hat{\mathbf{m}} = \mathbf{A}\mathbf{d}. \tag{A1}$$

Introduce the column vector of discrepancy $\boldsymbol{\epsilon}$ between the true and reconstructed vectors:

$$\epsilon_j = \hat{m}_j - m_j, \tag{A2}$$

where $j = 1, 2, \ldots, M$.

To find the elements a_{jk} of matrix $\mathbf{A}$, let us require that they give the minimum of each element of the mean square errors vector $\langle \boldsymbol{\epsilon}^2 \rangle$:

$$\langle \epsilon_j^2 \rangle \mapsto \min_{\{a_{jk}\}} \langle \epsilon_j^2 \rangle, \tag{A3}$$

where $k = 1, 2, \ldots, D$.

To satisfy the requirement (A3), one should take the partial derivative along the current element a_{ip} of the matrix $\mathbf{A}$ and make it equal to 0:

$$\begin{aligned} \frac{\partial}{\partial a_{ip}} \langle (a_{jk} d_k - m_j)^2 \rangle &= 2 \langle (a_{ik} d_k - m_i) d_p \rangle \\ &= 2 (a_{ik} \langle d_k d_p \rangle - \langle m_i d_p \rangle) \\ &= 2 (a_{ik} [\mathbf{R_{dd}}]_{kp} - [\mathbf{R_{md}}]_{ip}) = 0, \end{aligned} \tag{A4}$$

where the equality $\partial a_{jk} / \partial a_{ip} \equiv \delta_{ij} \delta_{pk}$ was taken into account, and the δ's are Kronecker's delta symbols. The last equation, after canceling the factor 2, can be written in matrix notation:

$$\mathbf{A}\mathbf{R_{dd}} - \mathbf{R_{md}} = \mathbf{0}. \tag{A5}$$

Therefore, under assumption that the inverse matrix $\mathbf{R_{dd}^{-1}}$ exists, the optimal matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{R_{md}} \mathbf{R_{dd}^{-1}}. \tag{A6}$$

APPENDIX B: ERROR COVARIANCE MATRIX DERIVATION

If a random column vector $\boldsymbol{\eta}$ relates to another random column vector $\boldsymbol{\xi}$ by a known nonrandom matrix $\mathbf{C}$ as $\boldsymbol{\eta} = \mathbf{C}\boldsymbol{\xi}$, then its covariance matrix $\mathbf{R}_{\eta\eta} \equiv \langle \boldsymbol{\eta}\boldsymbol{\eta}^T \rangle$ is expressed in terms of the covariance matrix $\mathbf{R}_{\xi\xi} \equiv \langle \boldsymbol{\xi}\boldsymbol{\xi}^T \rangle$ as

$$\mathbf{R}_{\eta\eta} = \mathbf{C}\mathbf{R}_{\xi\xi}\mathbf{C}^T. \tag{B1}$$

In the case of travel-time tomography, the column vector of discrepancy is given by $\boldsymbol{\epsilon} \equiv \hat{\mathbf{m}} - \mathbf{m}$, where $\hat{\mathbf{m}} = \mathbf{A}\mathbf{d}$. Therefore, the discrepancy can be expressed in terms of the matrix multiplication using block matrix notation:

$$\boldsymbol{\epsilon} = [\mathbf{A} - \mathbf{I}] \begin{bmatrix} \mathbf{d} \\ \mathbf{m} \end{bmatrix}, \tag{B2}$$

where $\mathbf{I}$ is the identical matrix. Using the formula (B1), one can get the expression for covariance matrix $\mathbf{R}_{\epsilon\epsilon}$:

$$\begin{aligned} \mathbf{R}_{\epsilon\epsilon} &= [\mathbf{A} - \mathbf{I}] \left\langle \begin{bmatrix} \mathbf{d} \\ \mathbf{m} \end{bmatrix} [\mathbf{d}^T \mathbf{m}^T] \right\rangle \begin{bmatrix} \mathbf{A}^T \\ -\mathbf{I} \end{bmatrix} \\ &= [\mathbf{A} - \mathbf{I}] \begin{bmatrix} \langle \mathbf{d}\mathbf{d}^T \rangle & \langle \mathbf{d}\mathbf{m}^T \rangle \\ \langle \mathbf{m}\mathbf{d}^T \rangle & \langle \mathbf{m}\mathbf{m}^T \rangle \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ -\mathbf{I} \end{bmatrix} \\ &= \mathbf{A}\mathbf{R_{dd}}\mathbf{A}^T - \mathbf{A}\mathbf{R_{md}^T} - \mathbf{R_{md}}\mathbf{A}^T + \mathbf{R_{mm}}. \end{aligned} \tag{B3}$$

Note that this formula for the covariance matrix of the discrepancy vector is valid for any matrix $\mathbf{A}$. Since the optimal matrix $\mathbf{A}$ of the stochastic inversion and time-dependent stochastic inversion is given by $\mathbf{A} = \mathbf{R_{md}}\mathbf{R_{dd}^{-1}}$, the $\mathbf{R}_{\epsilon\epsilon}$ matrix takes the form

$$\mathbf{R}_{\epsilon\epsilon} = \mathbf{R}_{\mathbf{mm}} - \mathbf{R}_{\mathbf{md}}\mathbf{R}_{\mathbf{dd}}^{-1}\mathbf{R}_{\mathbf{md}}^{T}. \quad \text{(B4)}$$

Note that the diagonal elements of the error covariance matrix $\mathbf{R}_{\epsilon\epsilon}$ represent the mean square errors of the estimation $\hat{\mathbf{m}}$.

[1]D. K. Wilson, A. Ziemann, V. E. Ostashev, and A. G. Voronovich, "An overview of acoustic travel-time tomography in the atmosphere and its potential applications," Acta. Acust. Acust. **87**, 721–730 (2001).

[2]S. N. Vecherin, V. E. Ostashev, D. K. Wilson, A. G. Voronovich, G. H. Goedecke, S. L. Collier, J. M. Noble, and D. Ligon, "Forward and inverse problems of acoustic tomography of the atmosphere," *Proc. 11th Intern. Symp. on Long Range Sound Propagation*, Fairlee, VT, 2004.

[3]D. K. Wilson, V. E. Ostashev, S. N. Vecherin, A. G. Voronovich, S. L. Collier, and J. M. Noble, "Assessment of acoustic travel-time tomography of the atmospheric surface layer," *Proceedings of AMS Symposium on Boundary Layers and Turbulence*, Portland, ME, 2004.

[4]S. L. Collier, D. A. Ligon, J. M. Noble, E. Patton, P. Sullivan, and V. E. Ostashev, "Acoustic tomographic array simulation," *Proc. 11th Intern. Symp. on Long Range Sound Propagation*, Fairlee, VT, 2004.

[5]D. K. Wilson and D. W. Thomson, "Acoustic tomographic monitoring of the atmospheric surface layer," J. Atmos. Ocean. Technol. **11**, 751–769 (1994).

[6]K. Arnold, A. Ziemann, and A. Raabe, "Acoustic tomography inside the atmospheric boundary layer," Phys. Chem. Earth, Part B **24**, 133–137 (1999).

[7]A. Ziemann, K. Arnold, and A. Raabe, "Acoustic tomography in the atmospheric surface layer," Ann. Geophys. **17**, 139–148 (1999).

[8]A. Ziemann, K. Arnold, and A. Raabe, "Acoustic travel-time tomography—a method for remote sensing of the atmospheric surface layer," Meteorol. Atmos. Phys. **71**, 43–51 (1999).

[9]A. Ziemann, K. Arnold, and A. Raabe, "Acoustic tomography as a method to identify small-scale land surface characteristics," Acta. Acust. Acust. **87**, 731–737 (2001).

[10]K. Arnold, A. Ziemann, and A. Raabe, "Tomographic monitoring of wind and temperature at different heights above the ground," Acta. Acust. Acust. **87**, 703–708 (2001).

[11]V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN SPON, London, 1997).

[12]P. Holstein, A. Raabe, R. Müller, M. Barth, D. Mackenzie, and E. Starke, "Acoustic tomography on the basis of travel-time measurement," Meas. Sci. Technol. **15**, 1420–1428 (2004).

[13]M. Barth, A. Raabe, P. Holstein, R. Muller, A. Ziemann, K. Arnold, D. Mackenzie, E. Starke, and M. Seliger, "Acoustic travel-time tomography as a tool to investigate temperature distributions on different spatial scales," *12th International Symposium on Acoustic Remote Sensing and Associated Techniques of the Atmosphere and Oceans*, Cambridge, UK, 2004.

[14]K. Aki and P. G. Richards, *Quantitative Seismology. Theory and Methods* (Freeman, San Francisco, 1980).

[15]V. E. Ostashev, B. Brähler, V. Mellert, and G. H. Goedecke, "Coherence functions of plane and spherical waves in a turbulent medium with the von Kármán spectrum of medium inhomogeneities," J. Acoust. Soc. Am. **104**, 727–737 (1998).

[16]V. E. Ostashev, V. Mellert, R. Wandelt, and F. Gerdes, "Propagation of sound in a turbulent medium I. Plane waves," J. Acoust. Soc. Am. **102**, 2561–2570 (1997).

[17]J. Gilson, D. Roemmich, B. D. Cornuelle, and L.-L. Fu, "Relationship of TOPEX/Poseidon altimetric height to steric height and circulation in the North Pacific," J. Geophys. Res. **103**, 965 (1998).

[18]N. Ducet, P. Y. Le Traon, and G. Reverdin, "Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and -2," J. Geophys. Res. **105**, 498 (2000).

[19]F. P. Bretherton, R. E. Davis, and C. B. Fandry, "A technique for objective analysis and design of oceanographic experiments applied to MODE-73," Deep-Sea Res. **23**, 559–582 (1976).

[20]J. Hinze, *Turbulence* (McGraw-Hill, New York, 1975).

[21]V. I. Tatarskii, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Keter, Jerusalem, 1971).

[22]J. C. Kaimal and J. J. Finnigan, *Atmospheric Boundary Layer Flows: Their Structure and Measurement* (Oxford University Press, New York, 1994).

[23]V. E. Ostashev, S. N. Vecherin, G. H. Goedecke, D. K. Wilson, A. G. Voronovich, and E. G. Patton, "Numerical simulation of acoustic tomography of the atmosphere," J. Acoust. Soc. Am. **117**, 2532 (2005).

[24]A. Andren, A. R. Brown, J. Graf, P. J. Mason, C. H. Moeng, F. T. M. Nieuwstadt, and U. Schumann, "Large-eddy simulation of a neutrally stratified boundary layer: A comparison of four computer codes," Q. J. R. Meteorol. Soc. **120**, 1457–1484 (1994).

[25]R. B. Stull, *An Introduction to Boundary Layer Meteorology* (Kluwer, Dordrecht, 1988).

[26]W. H. Munk, P. Worcester, and C. Wunsch, *Ocean Acoustic Tomography* (Cambridge University Press, New York, 1995).

[27]D. K. Wilson, J. G. Brasseur, and K. E. Gilbert, "Acoustic scattering and the spectrum of atmospheric turbulence," J. Acoust. Soc. Am. **105**, 30–34 (1999).

Fast calculation system specialized for head-related transfer function based on boundary element method

Makoto Otani[a)]
Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Yoshida-Honmachi, Sakyou-ku, Kyoto, Kyoto, 606-8501, Japan

Shiro Ise
Department of Urban and Environment Engineering, Graduate School of Engineering, Kyoto University, Yoshida-Honmachi, Sakyou-ku, Kyoto, Kyoto, 606-8501, Japan

(Received 26 May 2005; revised 8 December 2005; accepted 7 March 2006)

Recently, development of a numerical calculation of the head-related transfer function (HRTF) has been conducted using a computer model of a human head and the boundary element method. The reciprocity theorem is incorporated into the computational process in order to shorten the computational time, which is otherwise very long. On the other hand, another fast HRTF calculation method for any source position, which is realized by calculating factors independent of the source position in advance, has been suggested by the authors. Using this algorithm, the HRTF for any source position can be obtained in a few seconds with a common PC. The resulting HRTFs are more precise and are calculated faster than those by using the reciprocity theorem. However, speeding the process up even further is required in order to respond to a head movement and rotation or to moving sources during binaural sound reproduction. In this paper, a faster calculation method by incorporating a time domain operation into the authors' previous algorithm is proposed. Additionally, the new formulation, which eliminates the extra computational time in the preprocess, is proposed. This method is shown to be faster than the previous ones, but there are some discrepancies at higher frequencies.
 [DOI: 10.1121/1.2191608]

PACS number(s): 43.20.Fn, 43.64.Ha, 43.66.Qp [LLT] Pages: 2589–2598

I. INTRODUCTION

In order to reproduce sound accurately for binaural detection, the head-related transfer function (HRTF) is necessary to synthesize sound signals at each ear. HRTF databases, which are usually measured in an anechoic chamber, are published[1,2] and are widely used. These HRTF databases, however, only provide HRTFs for measured source positions. Thus, much work has been done to interpolate the existing HRTFs[3] in order to generate HRTFs that are not available in the database. In addition, HRTF calculation has been studied using analyses treating the head as a sphere or spheroid.[4] Recently, a large-scale numerical simulation has been shown to be a practical means of simulating acoustical phenomena around the human head. Such research has used a computer model of a human head and the boundary element method (BEM). Katz measured the geometry of the human head by using a 3D laser scanner in order to analyze an accurate model of the head by the BEM.[5] Subsequently, he measured the acoustic impedances of human skin and hair[6] to incorporate these results into the HRTF simulation.[7] Kahana *et al.* developed a computer model of the KEMAR dummy head and calculated the HRTF based on this model.[8] Finally, by constructing a computer model of the head including ear canals and eardrums, Walsh *et al.* confirmed numerically that a sound pressure is accurately reproduced on an eardrum if the sound pressure is modeled correctly at the entrance to the ear canal.[9]

The main problem with BEM-based calculation of the HRTF is its computational cost, especially with respect to computational time. In this method, element length needs to be less than $\frac{1}{4}-\frac{1}{6}$ of the wavelength of the sound that is modeled. Therefore, an accurate head model with a very finely discretized boundary is required to obtain accurate results. It has been shown that calculated HRTFs agree well with measured HRTFs if the geometry of the head and external ears is accurately reproduced in a computer model and its boundary elements satisfy the condition mentioned above.[7,8,10] However, for high frequencies which are within the audible range of humans, very long computational time is required because of the large number of elements. A goal of the work described in this paper is to speed up BEM-based HRTF calculation and, consequently, to develop a useful tool for calculating the HRTF for any source position more quickly than conventional methods, while maintaining the accuracy of the calculation.

The performance and resources of computers are constantly improving. However, computing costs required for BEM-based HRTF calculation are still large. In the HRTF calculation, the number of sound sources is one and that of receivers is two, that is, each ear. However, a sound source configuration is infinitely variable whereas the two receiver positions are constant. In order to save computational time

[a)]Electronic mail: otani@ae-gatel.archi.kyoto-u.ac.jp

involved in repetitive calculation for many source configurations, two solutions have been proposed. The first is a frequently used method that defines multiple source positions before the calculation and obtains results for each source position simultaneously. This method is realized by applying the lower-upper (LU) decomposition to a coefficient matrix and by solving simultaneous equations sequentially with changing factors dependent on the source position, i.e., the right-hand side of the linear system. One disadvantage to this method is that the LU decomposed matrices must be preserved in order to restart the back substitution process for source configurations, which were not initially defined. In the second method, the reciprocity theorem is invoked, which interchanges source and receiver positions.[5,8,10] In this method, changing a source position can be interpreted as changing a receiver position. Therefore, results for multiple source configurations can be obtained simultaneously. In addition, by reusing surface pressures obtained in the process of calculation, it becomes possible to calculate results for any source position, even those that are not defined before the calculation. However, it is necessary to be careful about possible errors introduced by the reciprocity theorem. That is, applying the theorem to the BEM could give rise to different results from those obtained with the normal BEM. Kahana *et al.* defined the boundary elements of the external ears to be finer than other elements in order to improve the performance of the method using the reciprocity theorem.[8] This method is referred to as the "reciprocity-theorem method" in this paper.

As described above, previous solutions have had some disadvantages. Therefore, another solution has been suggested by the authors.[11] In order to calculate the HRTF quickly for any source position without the reciprocity theorem, it may be preferable to divide the computational processes of the BEM into two steps, one independent of source position and the other dependent upon it. In this paper, these parts are referred to as the "preprocess" and "postprocess," respectively. The preprocess is calculated as the inverse of the coefficient matrix, which is independent of source position. For the postprocess, the HRTF can be calculated very quickly between any source position and each ear by using the results of the preprocess with maintaining the accuracy of the calculation. Unless the boundary conditions of the head model change, the preprocess, which takes a much longer time to calculate than the postprocess, does not need to be performed again. The postprocess of the authors' previous method calculates the HRTF more quickly for any source position compared to the postprocess of the reciprocity-theorem method, which involves more computational processes. Additionally, although the reciprocity-theorem method and the conventional method are, in theory, identical, the reciprocity-theorem method has the possibility of numerical errors. However, the results of the authors' previous method and those of the conventional BEM are virtually identical both in theory and in numerical analysis.[11] This method is referred to as the "source-independent method" in this paper.

In previous work, as mentioned above, to execute all the processes independently of the source position in the preprocess enables rapid calculation of the HRTF for any source position in the postprocess. Although this method is faster than the reciprocity-theorem method, it does take more than a few seconds. Because this computational time turns out to be mainly due to the frequency-domain operation involved in the postprocess, the current paper describes an enhancement of postprocess speed by executing it in the time domain instead of the frequency domain. Additionally, an improved formulation, which reduces the preprocess computational time, is described. The main disadvantage of the source-independent method is that the preprocess requires a lot more time than the reciprocity-theorem method. However, this new formulation enables the preprocess to be executed as fast as the conventional BEM and the reciprocity-theorem method. In this paper, an algorithm, which implements these improvements, is proposed and its computational time and precision are examined.

II. A FAST HRTF CALCULATION BASED ON THE BEM

In this section, an outline of the source-independent method previously proposed by the authors[11] is described. The most time-consuming process in the BEM is the calculation of surface pressures by solving simultaneous equations written as

$$\left(\tfrac{1}{2}\mathbf{I}_M + \mathbf{G}_n + j\omega\rho\mathbf{G}\mathbf{Y}\right)\hat{\mathbf{P}} = g\mathbf{A}, \tag{1}$$

where

$$\mathbf{I}_M = \mathrm{diag}(1,1,1,\ldots,1) \in \mathbf{C}^{M\times M},$$

$$\mathbf{G} = \left[\int\int_{S_j} G(\mathbf{r}|\mathbf{r}_j)\,dS\right] \in \mathbf{C}^{M\times M},$$

$$\mathbf{G}_n = \left[\int\int_{S_j} \frac{\partial G(\mathbf{r}|\mathbf{r}_j)}{\partial n}\,dS\right] \in \mathbf{C}^{M\times M},$$

$$g = [G(\mathbf{r}'_k|\mathbf{r}_j)] \in \mathbf{C}^{M\times N},$$

$$\mathbf{A} = [q_1, q_2, q_3, \ldots, q_N]^T \in \mathbf{C}^N,$$

$$\mathbf{Y} = \mathrm{diag}(y_n(\mathbf{r}_1), y_n(\mathbf{r}_2), y_n(\mathbf{r}_3), \ldots, y_n(\mathbf{r}_M)) \in \mathbf{C}^{M\times M},$$

M is the number of boundary elements or nodes, N is the number of sources, S_j is the jth surface element ($j=1,\ldots,M$), $\mathbf{r}'_k$ is the kth source position ($k=1,\ldots,N$), $\mathbf{r}_j$ is the center of S_j, $G(\mathbf{s}|\mathbf{r})$ is the Green function between $\mathbf{s}$ and $\mathbf{r}$, q_k is the volume velocity of the kth source, ω is the angular frequency, ρ is the air density, n is the normal vector, y_n is the acoustic admittance in the normal direction, and all the boundary elements are given acoustic admittances $\mathbf{Y}$.

Furthermore, since $g\mathbf{A}$ depends on the source position, the back substitution process of Eq. (1) must be restarted whenever the source position changes. Thus, the total time necessary for calculation is very long. However, if the calculation is specialized to the HRTF, boundary conditions and receiver positions near entrances to the ear canals are constant, enabling an enhancement of calculation speed. In the

conventional BEM, the sound pressure at a point $\mathbf{s}$ can be calculated by solving the simultaneous equations as written in Eq. (1) and given as

$$P(\mathbf{s},\omega) = g_s(\omega) - (\mathbf{G}_{ns} + j\omega\rho\mathbf{G}_s\mathbf{Y})\hat{\mathbf{P}},$$

where

$$g_s = [G(\mathbf{r}'_k|\mathbf{s})] \in \mathbf{C}^N,$$

$$\mathbf{G}_s = \left[\int \int_{\mathbf{S}j} G(\mathbf{r}_j|\mathbf{s})\, ds \right] \in \mathbf{C}^M,$$

$$\mathbf{G}_{ns} = \left[\int \int_{\mathbf{S}_j} \frac{\partial G(\mathbf{r}_j|\mathbf{s})}{\partial n}\, ds \right] \in \mathbf{C}^M,$$

and, since the calculation is specialized to the HRTF, $\mathbf{A}=1$ because the sound source number is one and its strength is unity. On the other hand, in the source-independent method, $\mathbf{Q}$, which is derived from the inverse of the coefficient matrix and is independent of source position, can be obtained by

$$\mathbf{Q} = (\mathbf{G}_{ns} + j\omega\rho\mathbf{G}_s\mathbf{Y})\left(\tfrac{1}{2}\mathbf{I}_M + \mathbf{G}_n + j\omega\rho\mathbf{G}\mathbf{Y}\right)^{-1}. \quad (2)$$

Calculating $\mathbf{Q}$ by Eq. (2) requires more time than solving Eq. (1), since the computation of an inverse of the coefficient matrix includes M back substitution processes, which is an $O(M^2)$ process, whereas solving the simultaneous equations includes N back substitution processes. Note that $N=1$ when the calculation is specialized for the HRTF and that $N=2$ when the reciprocity theorem is applied.

By substituting $\mathbf{Q}, P(\mathbf{s},\omega)$ can be obtained by

$$P(\mathbf{s},\omega) = g_s(\omega) - \mathbf{Q}_g(\omega). \quad (3)$$

Here, $\mathbf{Q}(\in C^{1\times M})$ depends on a frequency. In Eq. (3), $g_s(\omega)$ represents a direct sound from the sound source and $\mathbf{Q}_g(\omega)$ describes indirect sounds from each element. Then, $\mathbf{Q}(\in C^{1\times M})$ can be interpreted as the coefficient of $g(\omega)$ that represents the transfer function between each element and the source position.

In the conventional BEM, changing the source configuration requires a recalculation of $\hat{\mathbf{P}}$ and a back substitution process in Eq. (1), which is an $O(M^2)$ process. However, once $\mathbf{Q}$ is calculated in this method, $\mathbf{Q}$ does not need to be recalculated, which allows the total time necessary for the HRTF calculation to be greatly reduced since assessing $P(\mathbf{s},\omega)$ in Eq. (3) is only an $O(M)$ process.

Here, the process in which $\mathbf{Q}$ is calculated for each frequency is termed the preprocess, whereas the process that computes $P(\mathbf{s},\omega)$ at any receiver position from $\mathbf{Q}$ using Eq. (3) is referred to as the postprocess. The postprocess includes the inverse Fourier transform of the frequency response at $\mathbf{s}$ in order to obtain an impulse response.

As for the method using the LU decomposition, sound pressures for multiple source configurations can be obtained easily by changing the right-hand side of Eq. (1), which includes factors dependent upon source position, and by calculating each of these equations sequentially. In this method, it is necessary to preserve the coefficient matrices in order to define a source position after the calculation and calculate a sound pressure for such source configuration. Such calculations require an excessively large amount of hard-drive space, however, making them impractical for actual use. For example, computation for 10 000 boundary elements with double-precision arithmetic would require about 1.6 GB of hard-drive storage space for each frequency calculated. If the HRTF is calculated up to 12 kHz at 43-Hz increments, total hard drive space wouid be approximately 450 GB.

If the reciprocity theorem is used and positions of the source and receiver position are swapped, the source position adjacent to the entrance to the ear canal remains constant and $\hat{\mathbf{P}}$ in Eq. (2) can be reused even if the receiver position changes. Therefore, the HRTF for any source position can be obtained after the calculation if the surface pressures $\hat{\mathbf{P}}$ are preserved. Unlike in the case of the LU decomposition, only the vector needs to be preserved. Thus, it is easy to calculate a sound pressure for a source position defined after the calculation, which is equivalent to calculating the sound pressure at the receiver position when the reciprocity theorem is invoked.

Compared to the computational time required for the reciprocity-theorem method, the preprocess in the source-independent method requires more time to execute since computing the inverse of the coefficient matrix requires more back substitution processes than solving the simultaneous equations in Eq. (1). However, the postprocess in the source-independent method is significantly faster than for the reciprocity-theorem method. This is because the postprocess of the reciprocity-theorem method uses integrals that are not included in the postprocess of the source-independent method.

III. SPEEDING UP THE PREPROCESS

As described in the preceding section, the main disadvantage of the source-independent method is the computational time required for the preprocess, since the inverse of the coefficient matrix in Eq. (2) takes much longer to compute. However, there is an alternative way to obtain $\mathbf{Q}$ as follows. Equation (2) can be transformed into

$$\mathbf{Q}^T = \left(\left(\tfrac{1}{2}\mathbf{I}_M + \mathbf{G}_n + j\omega\rho\mathbf{G}\mathbf{Y}\right)^{-1}\right)^T (\mathbf{G}_{ns} + j\omega\rho\mathbf{G}_s\mathbf{Y})^T,$$

where T denotes the transposition of a matrix, and, subsequently,

$$\mathbf{Q}^T = \left(\left(\tfrac{1}{2}\mathbf{I}_M + \mathbf{G}_n + j\omega\rho\mathbf{G}\mathbf{Y}\right)^T\right)^{-1} (\mathbf{G}_{ns} + j\omega\rho\mathbf{G}_s\mathbf{Y})^T.$$

Therefore, $\mathbf{Q}$ can also be obtained by solving simultaneous equations written as

$$\left(\tfrac{1}{2}\mathbf{I}_M + \mathbf{G}_n + j\omega\rho\mathbf{G}\mathbf{Y}\right)^T \mathbf{Q}^T = (\mathbf{G}_{ns} + j\omega\rho\mathbf{G}_s\mathbf{Y})^T. \quad (4)$$

$\mathbf{Q}$ can be computed as fast as assessing $\hat{\mathbf{P}}$ in Eq. (1) by using Eq. (4) instead of Eq. (2). Thus, it is possible to eliminate the main disadvantage of the source-independent method, that is, a long computational time in the preprocess.

IV. SPEEDING UP THE POSTPROCESS

As described in Sec. II, it is possible to obtain the HRTF quickly for an arbitrary source position based on the BEM.

However, the time required for calculation is still more than a few seconds. Therefore, the authors developed an improved method, which enables the enhancement of postprocessing speed. The first step is to compute the inverse Fourier transform of $\mathbf{Q}(\mathcal{F}(\mathbf{Q})^{-1})$ for each frequency prior to the postprocess, so that it can be executed in the time domain. The second step is to replace the convolution of $\mathcal{F}(\mathbf{Q})^{-1}$ and $\mathcal{F}(g)^{-1}$ with adjustments for the delay and amplitude.

A. Computation in the time domain

$\mathbf{Q}$ in terms of the angular frequency ω can be represented as

$$\mathbf{Q} = [Q_1(\omega)Q_2(\omega)\cdots Q_M(\omega)],$$

where $Q_j(\omega)(j=1,\ldots,M)$ are sound pressures at receiver positions attributed to the sound waves radiating from the jth boundary element. By computing the inverse Fourier transform of $Q_j(\omega)$, an impulse response can be written as

$$\mathbf{q}_j(t) = \mathcal{F}(Q_j(\omega))^{-1}, \tag{5}$$

where t denotes a continuous time.

By adding a new process, which computes and preserves the time domain signals $\mathbf{q}_j(t)$, to the preprocess, the impulse response $\mathbf{p}(t)$ at a receiver position $\mathbf{s}$ can be obtained by the postprocess as

$$\mathbf{p}(t) = \mathbf{g}_s(t) - \sum_{j=1}^{M} \mathbf{q}_j(t) * \mathbf{g}_j(t), \tag{6}$$

where * denotes convolution, and where $\mathbf{g}_s(t)$ and $\mathbf{g}_j(t)$ are time-domain signals obtained by computing the inverse Fourier transform of transfer functions for direct sounds from a sound source to a receiver position, and indirect sounds from a sound source to the jth boundary element, respectively. If $\mathbf{r}'$ is defined as a source position and $\mathbf{r}_j$ as a center of jth boundary element, $\mathbf{g}_s(t)$ and $\mathbf{g}_j(t)$ can be derived from the Green function as

$$g_s(\omega) = \frac{e^{-jk|\mathbf{r}'-\mathbf{s}|}}{4\pi|\mathbf{r}'-\mathbf{s}|}$$

and

$$g_j(\omega) = \frac{e^{-jk|\mathbf{r}'-\mathbf{r}_j|}}{4\pi|\mathbf{r}'-\mathbf{r}_j|},$$

where k is the wave number.

B. High-speed operation

As described in Sec. IV A, the postprocess can be computed in the time domain. However, the calculation of Eq. (6) includes a convolution in the time domain and, therefore, takes longer than the source-independent method. However, $\mathbf{g}_j(t)$ in Eq. (6) corresponds to a delta function with a delay and a decay and can be given as

$$\mathbf{g}_j(t) = \frac{1}{4\pi|\mathbf{r}'-\mathbf{r}_j|}\delta\left(t - \frac{|\mathbf{r}'-\mathbf{r}_j|}{c}\right), \tag{7}$$

where c is the sound velocity and $\delta(t)$ is a delta function. Therefore, the convolution appearing on the right-hand side of Eq. (6) can be replaced with an operation that multiplies the amplitude of $\mathbf{q}_j(t)$ by $1/(4\pi|\mathbf{r}'-\mathbf{r}_j|)$ and gives $|\mathbf{r}'-\mathbf{r}_j|/c$ seconds delay to $\mathbf{q}_j(t)$, expressed as

$$\mathbf{p}(t) = \mathbf{g}_s(t) - \sum_{j=1}^{M} \frac{1}{4\pi|\mathbf{r}'-\mathbf{r}_j|}\mathbf{q}_j\left(t - \frac{|\mathbf{r}'-\mathbf{r}_j|}{c}\right). \tag{8}$$

As a result, the postprocessing speed is increased. Figure 1 shows a flowchart comparing the source-independent method and the proposed method.

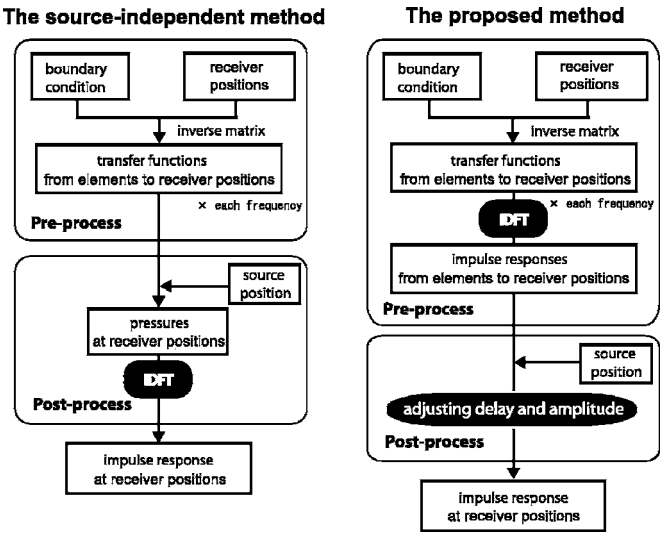


FIG. 1. Flow chart comparing the source-independent method and the newly proposed method. An inverse Fourier transformation is included in the preprocess of the proposed method.

When Eq. (8) is applied to a discrete time system in a practical calculation, Eq. (8) is replaced with the following equation:

$$\mathbf{p}[n] = \mathbf{g}_s[n] - \sum_{j=1}^{M} \frac{1}{4\pi|\mathbf{r}-\mathbf{r}_j|}\mathbf{q}_j\left[n - \mathrm{round}\left(\frac{|\mathbf{r}-\mathbf{r}_j|}{c}S\right)\right], \tag{9}$$

where $S(=\Delta f L)$ is the sampling frequency (Δf is a frequency interval, L is the length of $\mathbf{p}[n]$) and round (x) denotes an integer near x. Thus, the $O(L^2)$ process required by the convolution is reduced to an $O(L)$ process.

The $|\mathbf{r}'-\mathbf{r}_j|/c$ seconds delay in Eq. (8) is approximated by the delay of round $(|\mathbf{r}'-\mathbf{r}_j|S/c)$ samples in Eq. (9). Therefore, using Eq. (9) introduces computational errors not elicited with the source-independent method, because delay can only be expressed as an integral multiple of one sample ($1/S$ seconds). To reduce these computational errors, an oversampling is applied to $\mathbf{q}_j[n]$ in the preprocess. If sampling frequency becomes T times larger, it is possible to represent a delay as integral multiples of $1/(TS)$ seconds, thus limiting the potential computational errors.

V. COMPUTER MODEL

In this section, a procedure for developing a computer model of a dummy head is described. First, the geometry of the head and torso simulator B&K4128C was measured at about 500 000 points using a 3D laser scanner (Cyberware

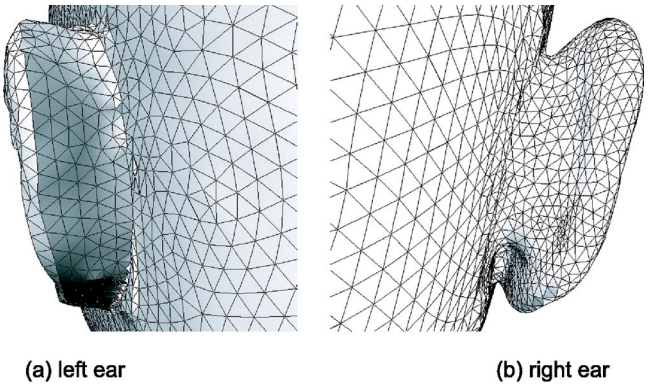


FIG. 2. (a) Backside of the left external ear scanned by a laser scanner. The area behind the ear is filled. (b) Backside of the right external ear scanned by a micro-CT scanner. Compared with the left ear scanned by a laser scanner, the right ear is modeled more precisely than the left ear.

3030RGB/I). Then, the data relating to the head were extracted and triangular elements were generated from these data. Surfaces that were blocked by others, such as the backside of the external ear, could not be measured by the laser scanner. Such surfaces were thus modeled using data obtained with a micro-CT (computerized tomography) scanner. The right ear of the model head obtained by the laser scanner was replaced by the one measured using the micro-CT scanner, but the left ear was not, permitting comparison between the two different ear models. Figures 2(a) and 2(b) show the left and right ears, respectively. Whereas no information is available about the backside of the left ear, which is simply filled in continuously, the right ear is modeled more accurately due to the incorporation of the micro-CT data.

The resulting computer model has too many surface elements to be applied to the BEM computation. Therefore, the model's surface is remeshed by a 3D mesh generator[12] in order to construct a tractable model for the BEM. Figure 3 shows the head model of B&K 4128C, including 28 000 triangular elements and 14 000 nodes. The maximum element

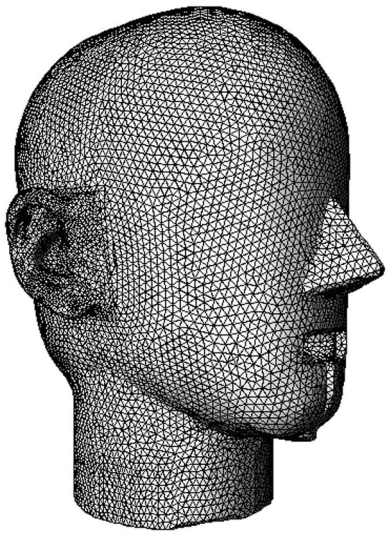


FIG. 3. Computer model of a head of B&K 4128C. This model consists of 28 000 elements and 14 000 nodes, and a maximum boundary-element length of 5.64 mm, corresponding to a frequency limitation of 10 kHz for six elements per wavelength and 15 kHz for four elements.

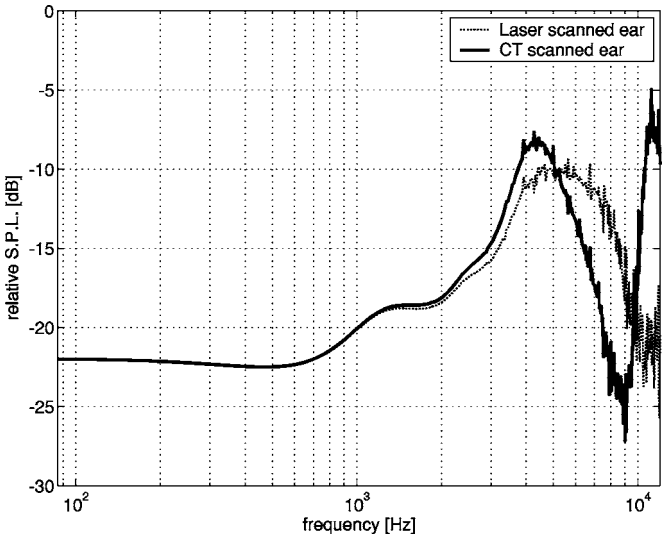


FIG. 4. Comparison between micro-CT scanned ear model and laser scanned ear model. A sound source is located 1 m in front of the head.

length is 5.64 mm. Allowing five elements per wavelength, this model is valid up to 12 000 Hz. In this model, the ear canals are blocked at their entrances. Thus, the receiver points are located near the blocked entrances, not on the eardrums. Hammershøi *et al.* showed, and Walsh *et al.* confirmed numerically, that HRTFs even at blocked ear canals contain complete spatial information.[9,13]

The HRTFs demonstrated in this paper are those calculated at the right ear, which was scanned by a micro-CT scanner. HRTFs calculated with the same micro-CT scanned ear model agree well with actual measurements below 10 kHz.[10] When comparing the HRTFs obtained for the right (micro-CT scanned) and the left (laser scanned) ear in response to a point source in front of the head, it was found that the results are consistent below 4 kHz. However, significant discrepancies are observed at higher frequencies as shown in Fig. 4, especially at frequencies where the notches appear. The first notch appears at 9 kHz for the right ear, but the first notch appears above 10 kHz for the left ear. The location of this dip provides important spectral cues for localizing a sound source on a median plane.[14] Thus, the HRTFs calculated at the left ear, i.e., a laser-scanned ear, may not provide precise spectral cues, especially for localizing sound sources on a median plane. These discrepancies are caused by inaccurate modeling of the left ear using a laser scanner. These results indicate that accurate modeling of the ear is crucial for the HRTF calculation.

VI. CALCULATION TIME AND ACCURACY

Table I shows the calculation time required using the conventional BEM, the reciprocity-theorem method, the source-independent method, and the currently proposed method with no oversampling and twofold, fourfold, eightfold, and 16-fold oversampling. The frequency range of calculation is 43–12 015 Hz at intervals of 43 Hz. All boundary elements were assumed to be rigid. Sampling frequency was 44.1 kHz and the length of $\mathbf{p}[n]$ and $\mathbf{q}_j[n]$ were both 1024. An Itanium2 1-GHz machine was used for the preprocess

TABLE I. Calculation time, required for the HRTF calculation for single receiver position and single source position using the conventional BEM, the reciprocity-theorem method, the source-independent method, and the proposed method with and without oversampling. Except for the conventional BEM, the computational processes are divided into the preprocess and postprocess.

	Preprocess (h)	Postprocess/source position (s)
Conventional BEM	420	—
Reciprocity-theorem method	420	239.94
Source-independent method	814	4.44
Proposed method without oversampling	420	0.08
Proposed method with "×2" oversampling	420	0.13
Proposed method with "×4" oversampling	420	0.20
Proposed method with "×8" oversampling	420	0.36
Proposed method with "×16" oversampling	420	0.61

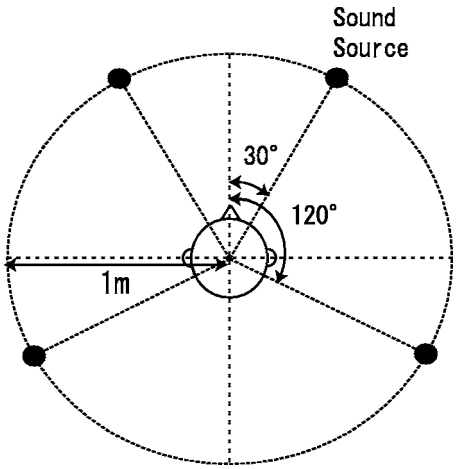


FIG. 5. The HRTF at the right ear is calculated. A sound source is located at 1-m distance from the midpoint of both ears with 30, 120, 240, and 330 deg azimuth.

and a Xeon 2.8-GHz machine was used for the postprocess. All computations were executed in double-precision arithmetic. The conventional BEM takes about 420 h to calculate the HRTF for a single source position. Applying the LU decomposition to the coefficient matrix allows the surface pressures of previously defined source configurations to be simultaneously solved. However, the LU decomposed matrices must be preserved when the source configurations are not previously defined, which requires an enormous amount of hard drive storage, making them impractical for actual use, as described in Sec. II. Katz, Kahana *et al.*, and Kakuhari *et al.* incorporated the reciprocity theorem into the BEM.[5,8,10] A calculation for multiple source configurations is thus converted into one for multiple receiver positions by replacing a receiver position with a source position. As a result, the surface pressures remain constant even when a source position changes. In addition, if surface pressures are preserved, it is possible to calculate quickly the HRTFs for any source position as the source-independent method does. Using the reciprocity-theorem method, the calculation required about 420 h for the preprocess and 239.94 s for the postprocess. On the other hand, using the source-independent method the postprocess took 4.44 s, but the preprocess took 814 h, which is significantly longer than the time required for either the conventional BEM or the reciprocity-theorem method. As stated in Sec. II, the difference in computational time required by the postprocess between these methods arises mainly from the use of integrals of the Green function and its normal derivative, which are included in the postprocess of the reciprocity-theorem method. Finally, the preprocess of the proposed method requires 420 h due to an improved formulation, which is stated in Sec. III, and the postprocess of proposed method requires 0.08 s without oversampling, 0.36 s with eightfold oversampling, and 0.61 s with 16-fold oversampling.

Next, the accuracy of the proposed method was assessed. The results of the source-independent method, which are identical to those of the conventional BEM,[11] are the benchmark for assessing the accuracy of the currently proposed method. The HRTF at the right ear, which was modeled using data from a micro-CT scanner, of the head with 28 000 triangular elements from 43 to 12 015 Hz in 43-Hz intervals was used. A sound source was located 1 m away from the midpoint of both ears at 30, 120, 240, and 330 deg azimuth. Figure 5 demonstrates the locations of sound sources and a head. Figures 6(a)–6(d) show the frequency responses obtained using the source-independent method and the newly proposed method for 30-, 120-, 240-, and 330-deg-azimuth sounds, respectively. At higher frequencies, there are numerous discrepancies between the two methods. The frequency responses obtained using the proposed method include many steep peaks and dips that appear locally and, consequently, the spectral slopes, the major peaks, and notches are incorrectly reproduced. There are fewer steep peaks and dips at frequencies below 4 kHz, but the curves of the proposed method deviate from those of the source-independent method. Thus, the proposed method without oversampling does not provide correct HRTFs.

Figures 7(a)–7(d) show the corresponding results obtained with an eightfold oversampling. Note that the ranges of the vertical axes in Fig. 7 are smaller than those in Fig. 6. Although the steep peaks and dips at higher frequencies cause many discrepancies in the results obtained by the proposed method with eightfold oversampling, there are fewer peaks and dips and their respective heights and depths are much smaller at lower frequencies. Deviations from the curves by the source-independent method at lower frequencies, as demonstrated in Fig. 6, cannot be seen in these figures. Hence, the results are better with respect to the non-oversampled method. Figures 8(a)–8(d) show the corresponding results with 16-fold oversampling, which compared to eightfold oversampling suppresses the discrepancies at higher frequencies. Thus, it is concluded that the proposed method provides more accurate results.

The root mean square (rms) errors of the newly proposed method were also calculated in the frequency and time domains and compared with the results to those from the source-independent method. The rms errors in both domains are expressed as

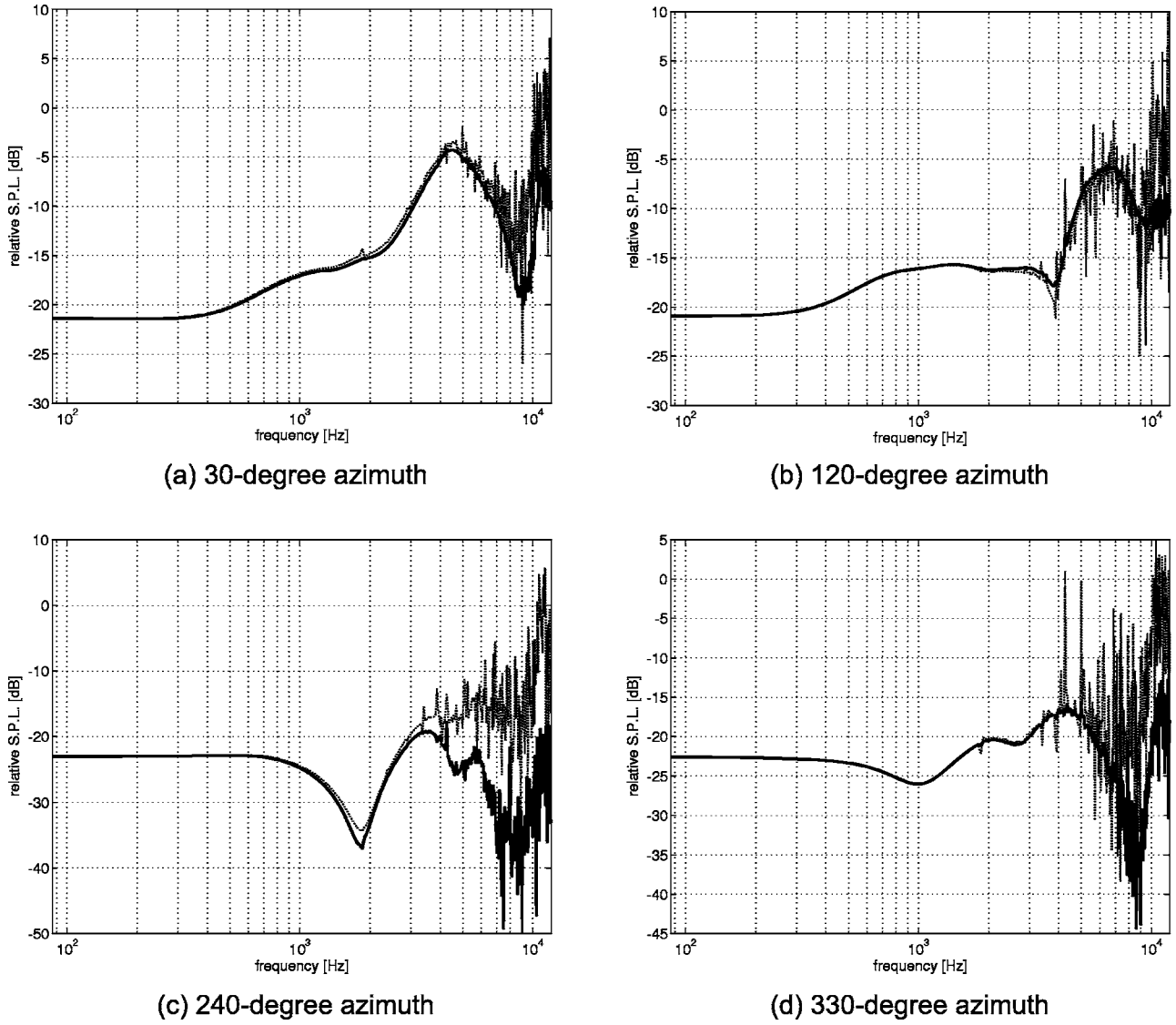


FIG. 6. Comparison in the frequency domain between the source-independent method and the newly proposed method without oversampling at (a) 30 deg azimuth, (b) 120 deg azimuth, (c) 240 deg azimuth, and (d) 330 deg azimuth. The solid line and the dotted line show the results obtained by the source-independent method and the newly proposed method, respectively.

$$\mathrm{rms}_{\mathrm{freq}} = \sqrt{\frac{1}{K}\sum_{h=1}^{K} (|P(\mathbf{s},\omega_h)| - |\hat{P}(\mathbf{s},\omega_h)|)^2}$$

and

$$\mathrm{rms}_{\mathrm{time}} = \sqrt{\frac{1}{L}\sum_{n=1}^{L} (\mathbf{p}[n] - \hat{\mathbf{p}}[n])^2},$$

respectively, where K is the number of frequencies calculated in the postprocess. $\omega_h(=\Delta f \times 2\pi h)$ is the angular frequency. $P(\mathbf{s},\omega_h)$ and $\hat{P}(\mathbf{s},\omega_h)$ are the frequency responses obtained using the source-independent method and the method proposed in this paper, respectively. $\mathbf{p}[n]$ and $\hat{\mathbf{p}}[n]$ are impulse responses obtained by these two methods. Figures 9(a) and 9(b) show the rms errors in the frequency and time domains, respectively, "×1," "×2," "×4," "×8," and "×16" in the figure denote the fold of oversampling. "×1" indicates that oversampling is not applied. As shown in Figs. 9(a) and 9(b), the value of rms error becomes smaller in each source direction for both domains with the more degrees of oversampling. Therefore, it is concluded that the accuracy of the proposed method increases with more degrees of oversampling. The rms error values are nearly the same for the four source positions, especially with eight and 16-fold oversampling, indicating that the total absolute error does not significantly vary with a source position. However, as seen in Figs. 7 and 8, the results at the 240- and 330-deg azimuths differ more from the source-independentmethod than those at other source positions. Since the source is located at the opposite side of the listening ear, the overall power of the HRTF is smaller at the 240 and 330-deg azimuths than at the 30- and 120-deg azimuths. Therefore, the relative errors are larger at the 240- and 330-deg azimuths than at other source positions.

As shown in Figs. 6–8, some of the steep peaks and dips obtained by the proposed method appear at frequencies that also have steep peaks and dips in the source-independent

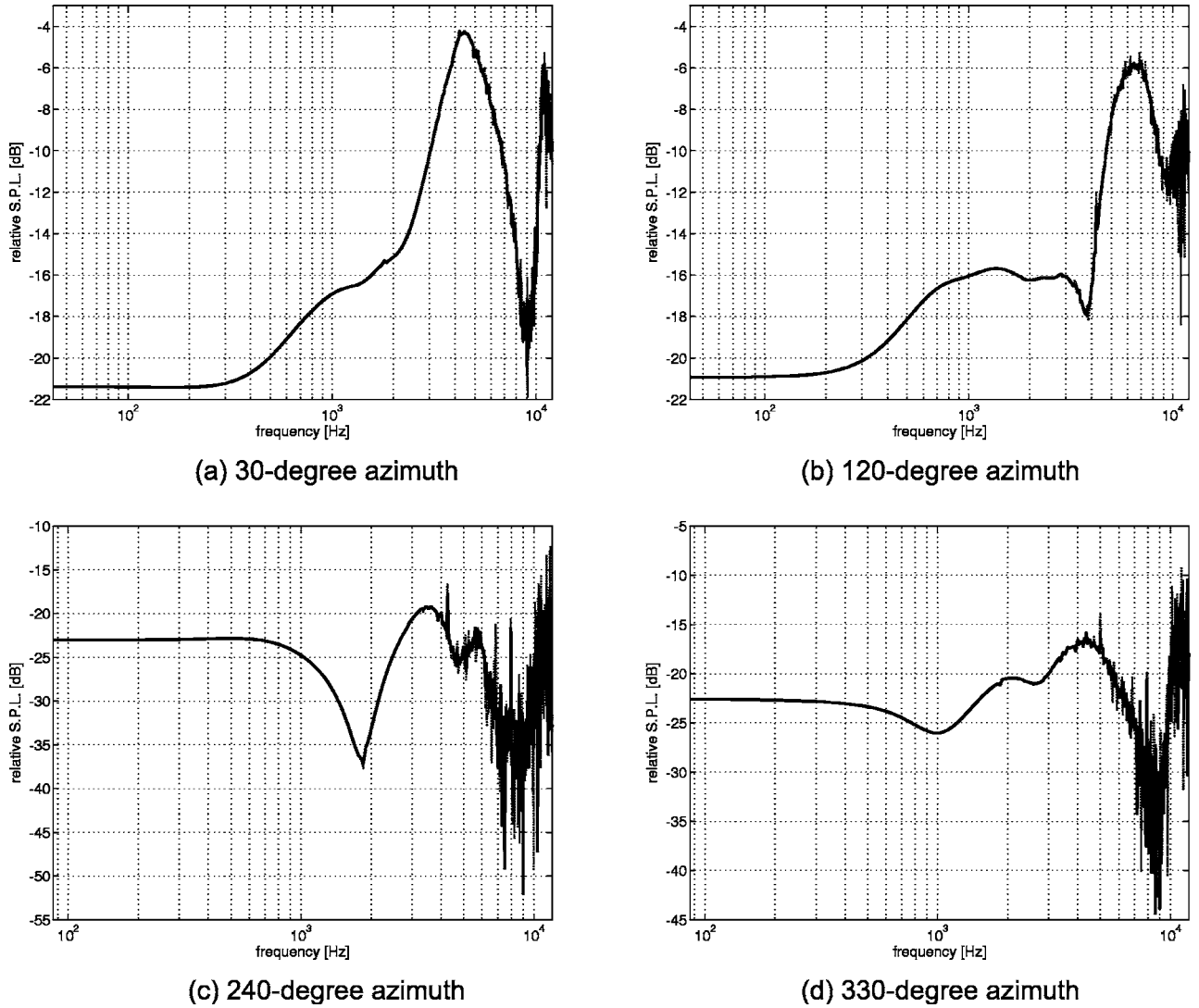


FIG. 7. Comparison in the frequency domain between the source-independent method and the newly proposed method with eightfold oversampling at (a) 30 deg azimuth, (b) 120 deg azimuth, (c) 240 deg azimuth, and (d) 330 deg azimuth. The solid line and the dotted line show the results obtained by the source-independent method and the newly proposed method with eightfold oversampling, respectively.

method. The approximations used in the proposed method may amplify these peaks and dips. As the azimuth increases, the results obtained using the source-independent method include more peaks and dips at higher frequencies. Consequently, increasing the azimuth in the proposed method increases the number of unwanted steep peaks and dips, i.e., the differences from the source-independent method increases.

As stated above, the main features of a frequency response of the HRTF such as spectral slopes, major peaks, and notches, which may be crucial for localizing sound sources,[14] are reproduced below 12 kHz by the proposed method with eightfold oversampling, which enables faster calculations than using the source-independent method. Furthermore, 16-fold oversampling enables the proposed method to give more accurate results with fewer peaks and dips at higher frequencies. However, if more accuracy of the overall source direction and/or higher frequencies are needed, then a higher degree of oversampling should be applied, which will increase the computational time of the proposed method.

VII. CONCLUSION

This paper provides two improvements upon the source-independent method previously proposed by the authors,[11] which is used to calculate the HRTF based on the BEM. A new formulation makes the preprocess faster, which may eliminate the main disadvantage of the source-independent method. Subsequently, the speed of the postprocess has been enhanced by performing a major operation in the time domain rather than in the frequency domain. The HRTFs are calculated based on a computer model of the head with 28 000 elements generated by a 3D laser scanner and a micro-CT scanner. Comparing the calculated HRTFs based on a micro-CT scanned ear and laser-scanned ear, which has an inaccurate shape, reveals the differences in the locations of peaks and notches and shows that accurate modeling of the external ears is significant for numerical simulation of the HRTFs.

Compared to previous methods, the postprocess of the proposed method with 16-fold oversampling is faster than

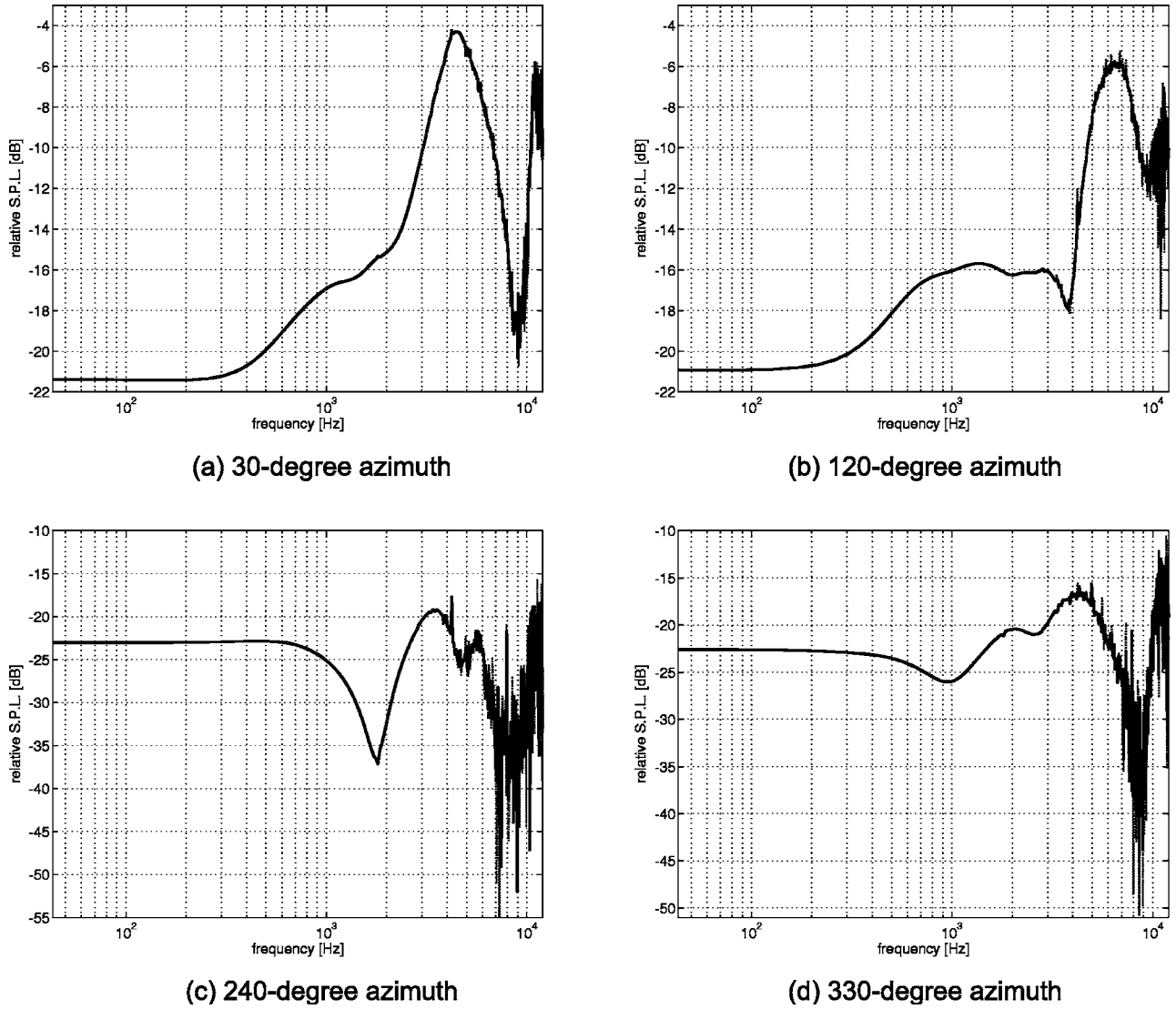


FIG. 8. Comparison in the frequency domain between the source-independent method and the newly proposed method with 16-fold oversampling at (a) 30 deg azimuth, (b) 120 deg azimuth, (c) 240 deg azimuth, and (d) 330 deg azimuth. The solid line and the dotted line show the results obtained by the source-independent method and the newly proposed method with 16-fold oversampling, respectively.

that of the source-independent method. The proposed method, which was compared to the source-independent method since the source-independent method is identical to the conventional BEM method, is faster and has improved accuracy when the preprocess is oversampled and can reproduce the main features of the HRTFs, such as spectral slopes, major peaks, and notches, with eightfold oversampling. Furthermore, the proposed method with 16-fold oversampling suppresses more discrepancies at higher frequencies and gives more accurate results. However, there are still steep peaks and dips at higher frequencies. Thus, if a higher precision is required, then a higher degree of oversampling should be applied.

The calculation method described in this paper is a simple but effective way to speed up boundary-element simulation of the HRTFs. In addition to an improved preprocess in speed, the postprocess is optimized to calculate the HRTF for source configurations defined *a posteriori*. The HRTFs generated using this algorithm can thus be applied to various virtual auditory displays, in order to present accu-

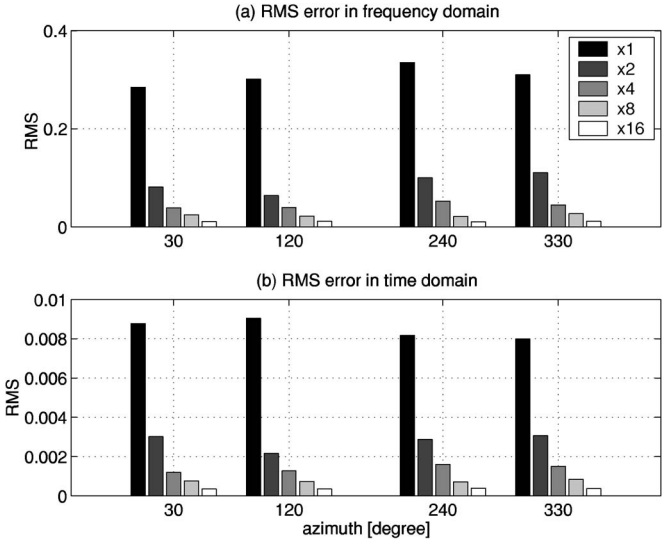


FIG. 9. Root mean square (rms) error in (a) frequency domain and (b) time domain of the proposed method compared with the source-independent method for four azimuthal locations. The error associated with the proposed method can be reduced if oversampling is applied.

rately localized stimuli directly to a listener's ears. At present, the calculation time may still be too long to allow its use in systems that require HRTFs for many source positions to be calculated in real time. However, compared to the traditional BEM, the proposed method can reduce the overall computational time necessary to generate a HRTF database based on a BEM simulation. Additionally, the major limitation of the current HRTF databases is the limited number of HRTFs for predetermined source configurations. Therefore, when the HRTFs for source configurations not included in the database are needed, the HRTFs for these source configurations must be interpolated. Since the proposed method is faster, it can be used to generate the HRTF quickly for required source configuration without interpolation using HRTFs from the database.

ACKNOWLEDGMENTS

The computer model of the head used in this work was developed in conjunction with the National Institute of Advanced Industrial Science and Technology (AIST) and Matsushita Electric Industrial Co., Ltd. The authors are grateful for their assistance.

[1]W. G. Gardner and K. D. Martin, "HRTF Measurement of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908 (1995).
[2]S. Takane, D. Arai, T. Miyajima, K. Watanabe, Y. Suzuki, and T. Sone, "A database of Head-Related Transfer Functions in whole directions on upper hemisphere," J. Acoust. Soc. Jpn. (E) **23**, 160–162 (2002).
[3]T. Nishino, S. Kajita, K. Takeda, and F. Itakura, "Interpolation of the head related transfer function on the horizontal plane" (in Japanese), J. Acoust. Soc. Jpn. (E) **55**, 91–99 (1999).
[4]K. Sugiyama, T. Sakaguchi, S. Aoki, and I. Kinoshita, "Calculation of acoustic coefficients between two ears using spheroids" (in Japanese), J. Acoust. Soc. Jpn. **51**(2), 117–122 (1995).
[5]B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function.I.Rigid model calculation," J. Acoust. Soc. Am. **110**, 2440–2448 (2001).
[6]B. F. G. Katz, "Acoustic absorption measurement of human hair and skin within the audible frequency range," J. Acoust. Soc. Am. **108**, 2238–2242 (2001).
[7]B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparison to real measurements," J. Acoust. Soc. Am. **110**, 2449–2455 (2001).
[8]Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi, "Boundary Element Simulation of HRTFs and Sound Fields Produced by Virtual Acoustic Imaging," Audio Engineering Society 105th Convention, preprint 4817 (1998).
[9]T. Walsh, L. Demkowicz, and R. Charles, "Boundary element modeling of the external human auditory system," J. Acoust. Soc. Am. **115**, 1033–1043 (2004).
[10]I. Kakuhari, K. Abe, and K. Terai, "The verification of head-related transfer function(HRTF) analysis method used by accurate 3-D model" (in Japanese), *Proc. Autumn Meet. Acoust. Soc. Jpn.* (2003), pp. 513–514.
[11]M. Otani and S. Ise, "A fast calculation method of the head-related transfer functions for multiple source points based on the boundary element method," Acoust. Sci. & Tech. 259–266 (2003).
[12]J. Schoeberl, "NETGEN - An advancing front 2D/3D-mesh generator based on abstract rules," Comput. Vis. Sci. **1**, 41–52 (1997).
[13]D. Hammershøi and H. Møller, "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427 (1996).
[14]J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834 (1974).

Iterative simulation of elastic wave scattering in arbitrary dispersions of spherical particles[a)]

Timothy E. Doyle
Department of Physics, Utah State University, Logan, Utah 84322-4415

(Received 15 September 2005; revised 6 February 2006; accepted 15 February 2006)

A numerical modeling approach was developed to simulate the propagation of shear and longitudinal waves in arbitrary, dense dispersions of spherical particles. The scattering interactions were modeled with vector multipole functions and boundary condition solutions for each particle. Multiple scattering was simulated by translating the scattered wave fields from one particle to another with the use of translational addition theorems, summing the multiple-scattering contributions, and recalculating the scattering using an iterative method. The approach can simulate 3D dispersions with a variety of particle sizes, compositions, and volume fractions. To test the model, spectra and wave field images were generated from ordered and disordered microstructures containing up to several thousand particles. The model predicted wave propagation phenomena such as refractive focusing and mode conversion. The iterations converged for many particle configurations, but did not converge or only partially converged for certain conditions, specifically large particle dispersions (>100 particles) at short wavelengths ($\lambda \sim$ particle diameter). Incorporating viscoelastic damping into the matrix properties reduced these numerical instabilities. The model is currently constrained by these convergence limitations and by the computation of sufficiently high multipole order for large numbers of particles. The theory and initial results for the model are presented. [DOI: 10.1121/1.2184989]

PACS number(s): 43.20.Gp, 43.35.Cg [YHB] Pages: 2599–2610

I. INTRODUCTION

The propagation of elastic waves through particles dispersed in an elastic or viscoelastic solid is relevant to many applications, including the nondestructive evaluation of particulate composites, the design of advanced acoustic materials, the ultrasonic characterization of biological tissues, and the remote sensing of geologic media. Predicting elastic wave behavior in such systems with statistical approaches is difficult, however, due to multiple scattering between particles, mode conversion at particle surfaces, lack of periodicity in the dispersion, and heterogeneity of particle sizes and compositions.[1–3] Nonempirical, physics-based models are therefore desired to simulate elastic wave propagation in dense, random dispersions of particles and predict the resultant macroscopic wave propagation properties.

This paper presents the theory and results of a numerical model developed to address these goals. The model uses vector multipole functions, boundary condition solutions, translational addition theorems, and direct iteration to solve the multiple-scattering problem for an arbitrary 3D distribution of spheres in an elastic or viscoelastic matrix. The model computes the spatial distributions and frequency spectra of elastic wave fields for a simulated particle configuration. From these results macroscopic properties such as attenuation and velocity can be predicted for a wide variety of particulate systems.

Elastic wave propagation in heterogeneous materials, particularly those with polycrystalline structures, has been theoretically addressed by perturbation methods,[4–8] radiative transfer theory,[9,10] and the Dyson equation.[11–14] For dispersions with spherical, spheroidal, or cylindrical particles, however, multipole expansions provide a means for directly simulating elastic wave scattering and computing macroscopic properties. Several multipole techniques have been developed for acoustic scattering in particulate systems, notably by Peterson and Ström,[15] Koc and Chew,[16] and most recently Gumerov and Duraiswami.[17,18] These models have focused on the transfer matrix (T-matrix) and fast multipole methods, but have been limited to longitudinal wave propagation and have not addressed the more complex problem of shear wave propagation and mode conversion. Elastic wave propagation in particulate systems has been addressed with multipole methods by Liu *et al.*,[19] however, who modeled elastic wave scattering in ordered lattices of particles by using a formulism based on periodic scatterers in a planar layer.

The T-matrix method has been favored for scattering models since it imposes an algebraic structure on the multiple scattering interactions that is accessible to many numeric methods.[15] Both fast multipole and periodic lattice models generally incorporate the T-matrix approach to solve the multiple-scattering problem.[16,18–20] Although T-matrix approaches have been formulated for elastic wave scattering,[21–24] very little work has been reported on simulating full elastic wave propagation in random particulate systems (i.e., with longitudinal and shear waves in both particles

[a)]Portions of this work were presented in "Iterative Simulation of Elastic Wave Scattering in Particle-Filled Materials," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 24, edited by D. O. Thompson and D. E. Chimenti (American Institute of Physics, Melville, New York, 2005), pp. 1121–1128.

and matrix). This is in contrast to electromagnetic scattering, where random particulate media have been extensively modeled with use of the T-matrix method.[20]

The direct iterative method is an alternative to the T-matrix method and simulates wave scattering in particle systems by iteratively recomputing the multiple-scattering contributions at each particle until convergence of the wave fields. It has been applied to electromagnetic scattering in small collections of spherical particles, where it is also known as the order-of-scattering method.[25–30] The direct iterative method has remained relatively unexplored for simulating elastic wave propagation, however, or for modeling fully random 3-D particle systems with large numbers of particles (>100 particles), high particle concentrations (>20%), and mixed particle sizes and properties.

The direct iterative method may offer advantages over a T-matrix formulism for modeling elastic waves in particulate systems. Specifically, the direct iterative method is computationally simpler than the T-matrix approach and avoids the problems of numerically solving large systems of linear equations for large numbers of particles. Although these linear systems are themselves often solved using iterative methods, the direct iterative approach circumvents the need to explicitly formulate and solve the multiple scattering in a matrix formulism. Scattering interactions that include shear waves and mode conversion can therefore be more readily modeled since they do not have to be incorporated into a T matrix.

An additional feature of the direct iterative method stems from the step-by-step computation of each successive scattering order (for example, zeroth order single scattering from the pre-iteration computations, first-order multiple scattering from the first iteration, and so forth). This ability provides greater access in the model to different levels of the scattering process, and allows the effects of multiple scattering to be quantified with respect to single scattering. It also provides a useful diagnostic of the model's performance and accuracy and allows certain interactions to be individually monitored and manipulated for computational efficiency (e.g., different convergence rates for different wave modes). Such access is difficult in the T-matrix approach, where all orders of multiple scattering are solved simultaneously with solution of the linear system.

The present research extends the work of Gumerov and Duraiswami,[17,18] Liu *et al.*,[19] and other investigators by modeling full elastic waves in arbitrary dispersions of spherical particles. The objectives were to develop the theory, demonstrate the feasibility, and explore the potential of a direct iterative approach for elastic wave propagation in particulate systems. An iterative multipole model was developed that computes the propagation of waves by using single-particle scattering solutions, addition theorems to translate the scattered fields from one particle to another, and iteration to arrive at a convergent solution for the multiply scattered fields. Simulations were performed on a personal computer for particulate systems of up to several thousand particles and particle volume fractions up to 50%. Initial results are presented for example simulations, and the capabilities and deficiencies of the current model are summarized.

II. THEORY

A. Vector multipole functions

Elastic waves in solids derive from the Navier equation for linear, homogeneous materials:

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = (\lambda + 2\mu)\,\nabla\,(\nabla \cdot \mathbf{u}) - \mu\,\nabla \times (\nabla \times \mathbf{u}), \tag{1}$$

where λ and μ are Lamé's constants, ρ is the density, and $\mathbf{u}$ is the displacement vector. Equation (1) can be recast as two separate vector Helmholtz equations by splitting $\mathbf{u}$ into longitudinal ($\mathbf{u}_L$) and transverse ($\mathbf{u}_S$) parts and imposing a harmonic time dependency of the form $e^{-i\omega t}$:

$$(\nabla^2 + k_L^2)\mathbf{u}_L = 0, \tag{2}$$

$$(\nabla^2 + k_S^2)\mathbf{u}_S = 0, \tag{3}$$

where k_L and k_S denote the longitudinal and shear wave vectors, respectively.

Solutions to Eqs. (2) and (3) in spherical coordinates are the vector multipole functions, also referred to as vector spherical wave functions. Since they are solutions to the Helmholtz equation, the vector multipole functions can be generated from a scalar potential Φ and vector potential $\boldsymbol{\Psi}$ defined as

$$\Phi = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} z_n(kr) Y_{nm}(\theta, \varphi), \tag{4}$$

$$\boldsymbol{\Psi} = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} z_n(kr) \mathbf{Y}_{nm}^{n}(\theta, \varphi). \tag{5}$$

The radial function $z_n(kr)$ represents either a spherical Bessel function $[j_n(kr)]$ or a spherical Hankel function of the first or second kind $[h_n^{(1)}(kr)$ or $h_n^{(2)}(kr)]$. The vector spherical harmonic $\mathbf{Y}_{nm}^{l}(\theta, \varphi)$ is a pure-orbital vector spherical harmonic and is defined with Clebsch-Gordan coefficients, scalar spherical harmonics, and helicity basis vectors as follows:[31–33]

$$\mathbf{Y}_{nm}^{l}(\theta, \varphi) = C_{l,m+1,1,-1}^{n,m} Y_{l,m+1} \mathbf{e}_{-1} + C_{l,m,1,0}^{n,m} Y_{l,m} \mathbf{e}_0 + C_{l,m-1,1,1}^{n,m} Y_{l,m-1} \mathbf{e}_{+1}. \tag{6}$$

The notation used is from Ref. 33. The vector multipole functions are derived by taking the gradient of the scalar potential, the curl of the vector potential, and the double curl of the vector potential in the following manner:

$$\mathbf{U} = \frac{1}{k}\,\nabla\,\Phi, \tag{7}$$

$$\mathbf{V} = \frac{1}{k}\,\nabla \times \boldsymbol{\Psi}, \tag{8}$$

$$\mathbf{W} = \frac{1}{k^2}\,\nabla \times (\nabla \times \boldsymbol{\Psi}). \tag{9}$$

The resulting functions are complete and orthogonal multipole expansions for $n=0$ to ∞ and $m=-n$ to $+n$:

$$\mathbf{U} = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \left(\sqrt{\frac{n}{2n+1}} z_{n-1}(kr) \mathbf{Y}_{nm}^{n-1}(\theta,\varphi) + \sqrt{\frac{n+1}{2n+1}} z_{n+1}(kr) \mathbf{Y}_{nm}^{n+1}(\theta,\varphi) \right), \tag{10}$$

$$\mathbf{V} = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} \left(i \sqrt{\frac{n+1}{2n+1}} z_{n-1}(kr) \mathbf{Y}_{nm}^{n-1}(\theta,\varphi) - i \sqrt{\frac{n}{2n+1}} z_{n+1}(kr) \mathbf{Y}_{nm}^{n+1}(\theta,\varphi) \right), \tag{11}$$

$$\mathbf{W} = \sum_{n=0}^{\infty} \sum_{m=-n}^{+n} z_n(kr) \mathbf{Y}_{nm}^{n}(\theta,\varphi). \tag{12}$$

The **U**, **V**, and **W** vector multipoles are similar to the vector spherical wave functions of Stratton;[34] the vector multipoles of Rose,[35] Edmonds,[36] and Greiner and Maruhn;[31] and the vector spherical harmonic of Jackson.[37] **U** represents the longitudinal displacement field. **V** and **W** are transverse fields that can represent either elastic shear waves or electromagnetic waves. The shear fields are designated by their electromagnetic counterparts since the horizontal and vertical conventions normally used to differentiate the two types of shear fields are ambiguous in a spherical coordinate system. **V** and **W** therefore correspond to the shear-electric and shear-magnetic displacement fields, respectively.

The initial waves incident on the particles are modeled as plane waves. The spherical expansion for the longitudinal plane wave is derived from the gradient of the partial wave expansion for scalar fields:

$$\begin{aligned} \mathbf{e}_z e^{ikz} &= \frac{1}{ik} \nabla e^{ikz} \\ &= \frac{1}{ik} \nabla \left[\sum_{n=0}^{\infty} i^n \sqrt{4\pi(2n+1)} j_n(kr) Y_{n,0}(\theta,\varphi) \right] \\ &= \sum_{n=0}^{\infty} i^{n-1} \sqrt{4\pi(2n+1)} \mathbf{U}_{n,0}. \end{aligned} \tag{13}$$

At least three different partial wave expansions have been presented in the literature for transverse vector planes waves, including those by Stratton,[34] Jackson,[37] and Greiner and Maruhn.[31] Grandy further derived expansions for plane-polarized waves from Jackson's circularly polarized waves.[38] Mathematical analysis and numerical testing, however, showed that only the expansions of Greiner and Maruhn were correct and converged properly. Adapting their expressions, the partial wave expansions for shear waves propagating in the z direction and polarized in the x and y directions are the following:

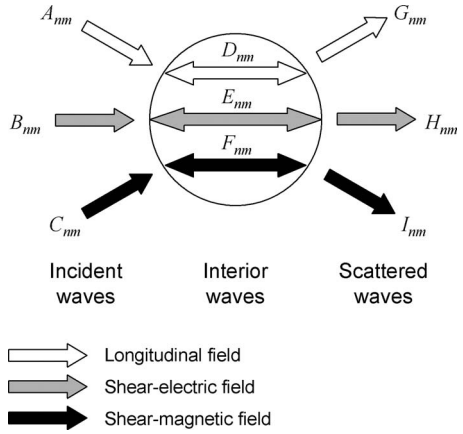


FIG. 1. Diagram of incident, interior, and scattered elastic waves for single-particle scattering, with associated amplitude coefficients for the (n,m) multipole moment.

$$\mathbf{e}_x e^{ikz} = \sum_n i^n \sqrt{\pi(2n+1)} [\mathbf{W}_{n,+1} + \mathbf{V}_{n,+1} + \mathbf{W}_{n,-1} - \mathbf{V}_{n,-1}], \tag{14}$$

$$\mathbf{e}_y e^{ikz} = (-i) \sum_n i^n \sqrt{\pi(2n+1)} [\mathbf{W}_{n,+1} + \mathbf{V}_{n,+1} - \mathbf{W}_{n,-1} + \mathbf{V}_{n,-1}]. \tag{15}$$

B. Boundary condition solutions

Simulation of multiple scattering in a dispersion of spherical particles first requires scattering solutions for the individual spheres. For scattering from a single solid sphere in a solid matrix, there will be an incoming incident field, a refracted interior field, and an outgoing scattered field for each of the **U**, **V**, and **W** wave fields. Each wave field component will also have an associated amplitude coefficient. Figure 1 shows the relationship between each of the field components and the coefficients.

Single-sphere scattering has been solved numerous times in the literature, but often not in the most general form. These solutions model the incident fields as plane waves with only a longitudinal or shear component.[39–44] Although these conditions are often sufficient for dilute suspensions, the incident fields for each particle in multiple-scattering computations will not in general be plane waves, but a combination of the initial plane waves and the scattered waves from other particles. Also, the incident fields will be a combination of both longitudinal and shear waves. Single-sphere scattering solutions were therefore required for arbitrary incident fields comprised of both longitudinal and shear waves.

Given an arbitrary incident field, the amplitude coefficients of the interior and scattered fields are found by solving the boundary conditions on the surface of the sphere. The boundary conditions provide a set of six linear equations for the six unknown coefficients. Three of these equations are obtained from continuity of the displacements:

$$\mathbf{u}^{\text{incident}} + \mathbf{u}^{\text{scattered}} = \mathbf{u}^{\text{interior}}. \tag{16}$$

Orthogonality conditions for the vector spherical harmonics are used to eliminate the angular dependence from the boundary conditions:

$$\int_0^{2\pi} d\varphi \int_0^{\pi} d(\cos\theta) \mathbf{Y}_{NM}^{L} {}^{*}(\theta,\varphi) \mathbf{Y}_{nm}^{l}(\theta,\varphi) = \delta_{Ll}\delta_{Nn}\delta_{Mm}. \tag{17}$$

The multipole fields [Eqs. (10)–(12)] and associated amplitude coefficients are first substituted into Eq. (16). The resulting equation is then multiplied with $\mathbf{Y}_{NM}^{N-1}{}^{*}(\theta,\varphi)$, $\mathbf{Y}_{NM}^{N}{}^{*}(\theta,\varphi)$, and $\mathbf{Y}_{NM}^{N+1}{}^{*}(\theta,\varphi)$ to obtain three separate equations. Finally, the equations are integrated over θ and φ to obtain the first three coefficient equations.

The other three equations are derived from continuity of the stresses, where σ is the stress tensor:

$$\sigma^{\text{incident}} + \sigma^{\text{scattered}} = \sigma^{\text{interior}}. \tag{18}$$

Equations for continuity of the radial stresses σ_{rr} and tangential stresses $\sigma_{r\theta}$ and $\sigma_{r\varphi}$ are solved separately. The stresses are related to the displacements by the following:

$$\sigma_{rr} = \lambda(\nabla \cdot \mathbf{u}) + 2\mu \frac{\partial u_r}{\partial r}, \tag{19}$$

$$\sigma_{r\theta} = \mu \cdot \left(\frac{1}{r}\frac{\partial u_r}{\partial \theta} + \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r} \right), \tag{20}$$

$$\sigma_{r\varphi} = \mu \cdot \left(\frac{1}{r\sin\theta}\frac{\partial u_r}{\partial \varphi} - \frac{u_\varphi}{r} + \frac{\partial u_\varphi}{\partial r} \right). \tag{21}$$

Since Eqs. (19)–(21) contain only scalar components of the displacement, the angular dependence is removed using orthogonality of the scalar spherical harmonics:

$$\int_0^{2\pi} d\varphi \int_0^{\pi} d(\cos\theta) Y_{NM}{}^{*}(\theta,\varphi) Y_{nm}(\theta,\varphi) = \delta_{Nn}\delta_{Mm}. \tag{22}$$

To solve the σ_{rr}, $\sigma_{r\theta}$, and $\sigma_{r\varphi}$ boundary conditions, Eqs. (19)–(21) are substituted into Eq. (18) with the appropriate scalar components and amplitude coefficients for the multipole fields. This yields three equations that are then multiplied with $Y_{NM}{}^{*}(\theta,\varphi)$ and integrated over θ and φ to obtain the last three coefficient equations.

The resulting six equations have the following matrix form that relates the six unknown coefficients to the three known coefficients:

$$\begin{pmatrix} \eta_{1j} & -\eta_{1h} & \eta_{2j} & -\eta_{2h} & 0 & 0 \\ \eta_{6j} & -\eta_{6h} & \eta_{7j} & -\eta_{7h} & 0 & 0 \\ \eta_{4j} & -\eta_{4h} & \eta_{5j} & -\eta_{5h} & 0 & 0 \\ \eta_{9j} & -\eta_{9h} & \eta_{10j} & -\eta_{10h} & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta_{3j} & -\eta_{3h} \\ 0 & 0 & 0 & 0 & \eta_{8j} & -\eta_{8h} \end{pmatrix} \begin{pmatrix} D_{NM} \\ G_{NM} \\ E_{NM} \\ H_{NM} \\ F_{NM} \\ I_{NM} \end{pmatrix} = \begin{pmatrix} A_{NM}\eta_{1g} + B_{NM}\eta_{2g} \\ A_{NM}\eta_{6g} + B_{NM}\eta_{7g} \\ A_{NM}\eta_{4g} + B_{NM}\eta_{5g} \\ A_{NM}\eta_{9g} + B_{NM}\eta_{10g} \\ C_{NM}\eta_{3g} \\ C_{NM}\eta_{8g} \end{pmatrix}. \tag{23}$$

Lines 1, 3, and 5 in Eq. (23) derive from the displacement boundary conditions [Eq. (16)]. Line 2 derives from the radial stress boundary condition (σ_{rr}), and lines 4 and 6 derive from the tangential stress boundary conditions ($\sigma_{r\varphi}$ and $\sigma_{r\theta}$). The η matrix elements are functions of the spherical radial functions [denoted by the subscript $j = j_N(ka)$, $h = h_N^{(1)}(ka)$, or $g = h_N^{(2)}(ka)$], multipole order N, longitudinal wave vector k_L, shear wave vector k_S, and sphere radius a.[45] The η matrix elements are given in the Appendix. Since the coefficient matrix separates into a 4×4 matrix and 2×2 matrix, the shear-magnetic fields are decoupled from the longitudinal and shear-electric fields and do not participate in mode conversion. In the scattering model the scattering amplitudes are obtained for each particle from algebraic solutions for the 2×2 and 4×4 matrices using Cramer's rule.

For monopole ($N=0$) terms, many of the η matrix elements in Eq. (23) vanish, resulting in a singular matrix that cannot be solved using Cramer's rule. This is due to the fact that the lowest moment possible for shear waves is the dipole ($N=1$) term. Therefore, a different matrix equation is solved for the longitudinal monopole fields. For monopole scattering, a system of only two equations is required for solution of the two unknown coefficients (the interior and scattered longitudinal field coefficients). In matrix form they are

$$\begin{pmatrix} \eta_{1j} & -\eta_{1h} \\ \eta_{6j} & -\eta_{6h} \end{pmatrix} \begin{pmatrix} D_{00} \\ G_{00} \end{pmatrix} = \begin{pmatrix} A_{00}\eta_{1g} \\ A_{00}\eta_{6g} \end{pmatrix}. \tag{24}$$

The first line is from the displacement boundary condition [Eq. (16)], and the second line is from the radial stress boundary condition (σ_{rr}). There are no tangential stresses for the monopole moment. Similar modifications to Eq. (23) are required for scattering from a fluid-filled sphere in a solid matrix and from a solid sphere in a fluid matrix. The 4×4 submatrix in Eq. (23) reduces to a 3×3 matrix for these cases.

C. Translation of scattered fields

In multiple scattering of elastic waves, the scattered waves from each particle impinge on the other particles and modify their incident wave fields. Since the multipole functions are specific to each particle's coordinate system, the

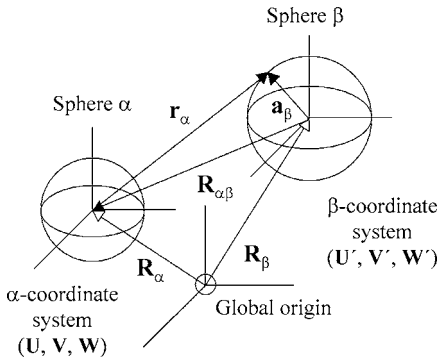


FIG. 2. Position ($\mathbf{R}_\alpha$ and $\mathbf{R}_\beta$) and translation ($\mathbf{r}_\alpha$) vectors for addition theorems with respect to local and global coordinates for spheres α and β.

vector multipole functions for each particle must be transformed into the coordinate systems of the other particles in order to update the incident fields. These transformations are the following:

$$\mathbf{U}'_{\nu\mu} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Q_{\nu\mu}^{nm} \mathbf{U}_{nm}, \tag{25}$$

$$\mathbf{V}'_{\nu\mu} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (S_{\nu\mu}^{nm} \mathbf{V}_{nm} + T_{\nu\mu}^{nm} \mathbf{W}_{nm}), \tag{26}$$

$$\mathbf{W}'_{\nu\mu} = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} (S_{\nu\mu}^{nm} \mathbf{W}_{nm} + T_{\nu\mu}^{nm} \mathbf{V}_{nm}), \tag{27}$$

where $\mathbf{U}_{nm}$, $\mathbf{V}_{nm}$, and $\mathbf{W}_{nm}$ are wave fields scattered from the originating sphere (α) and are therefore in sphere α's coordinate system (r_α, θ_α, and φ_α). $\mathbf{U}'_{\nu\mu}$, $\mathbf{V}'_{\nu\mu}$, and $\mathbf{W}'_{\nu\mu}$ are the same wave fields incident on a second sphere (β) and are therefore in sphere β's coordinate system (r_β, θ_β, and φ_β). Equations (25)–(27) transform the wave fields with expansions containing the original multipole functions and translation coefficients ($Q_{\nu\mu}^{nm}$, $S_{\nu\mu}^{nm}$, and $T_{\nu\mu}^{nm}$) derived from the translational addition theorems.

Figure 2 shows the geometric relationship for the field translations between the two spheres. The position of sphere α with respect to sphere β is $\mathbf{R}_{\alpha\beta}=\mathbf{R}_\alpha-\mathbf{R}_\beta$, where $\mathbf{R}_\alpha$ and $\mathbf{R}_\beta$ are the global position vectors for the two spheres. The translation coefficients are only valid on the surface of sphere $\beta(r_\beta=a_\beta)$, and the center of sphere α must lie outside of sphere β. Although addition theorems can also be derived for spheres embedded within larger spheres ($a_\beta>R_{\alpha\beta}$), this work only considered spheres external to each other and nonoverlapping ($a_\beta<R_{\alpha\beta}$).

The scalar addition theorem is sufficient for translating the longitudinal field since the longitudinal vector multipole function arises from a scalar potential [Eq. (7)]. The scalar addition theorem has been published extensively[33,46–49] and provides the $Q_{\nu\mu}^{nm}$ translation coefficients:

$$Q_{\nu,\mu}^{n,m} = \sqrt{4\pi} \sum_{p=0}^{\infty} \left\{ i^{\nu-p-n} h_p^{(1)}(kR_{\alpha\beta}) Y_{p,m-\mu}(\Theta_{\alpha\beta}, \Phi_{\alpha\beta}) \times C_{\nu,0,p,0}^{n,0} C_{\nu,\mu,p,m-\mu}^{n,m} \sqrt{\frac{(2\nu+1)(2p+1)}{(2n+1)}} \right\}. \tag{28}$$

Vector addition theorems are required to translate the shear fields since the transverse vector multipoles arise from a vector potential [Eqs. (8) and (9)]. Cruzan was one of the earliest to derive expressions for the $S_{\nu\mu}^{nm}$ and $T_{\nu\mu}^{nm}$ translation coefficients by directly transforming the coordinates in the vector multipole functions and applying various identities and relationships to arrive at an analytical solution.[50] Another approach is to expand the vector multipoles in the α coordinate system as a series comprised of expansion coefficients and vector spherical harmonics in the β coordinate system. The expression can then be integrated in a manner similar to a Fourier series to determine the translation coefficients.[45] The results produce a set of solutions for the translation coefficients $S_{\nu\mu}^{nm}$ and $T_{\nu\mu}^{nm}$, the most direct of which are the following:

$$S_{\nu\mu}^{nm} = Z_{\nu,\nu,\mu}^{n,n,m}, \tag{29}$$

$$T_{\nu\mu}^{nm} = -i\sqrt{\frac{2\nu+1}{\nu+1}} Z_{\nu-1,\nu,\mu}^{n,n,m} = i\sqrt{\frac{2\nu+1}{\nu}} Z_{\nu+1,\nu,\mu}^{n,n,m} = i\sqrt{\frac{2n+1}{n+1}} Z_{\nu,\nu,\mu}^{n-1,n,m} = -i\sqrt{\frac{2n+1}{n}} Z_{\nu,\nu,\mu}^{n+1,n,m}, \tag{30}$$

where

$$Z_{\lambda,\nu,\mu}^{l,n,m}(\mathbf{R}_{\alpha\beta}) = \sum_{p=|l-\lambda|}^{l+\lambda} \left\{ i^{\lambda-l-p} h_p^{(1)}(kR_{\alpha\beta}) Y_{p,m-\mu}(\Theta_{\alpha\beta}, \Phi_{\alpha\beta}) \times C_{\lambda,0,p,0}^{l,0} \sqrt{\frac{4\pi(2\lambda+1)(2p+1)}{2l+1}} \times \sum_{\tau=-1}^{1} C_{l,m-\tau,1,\tau}^{n,m} C_{\lambda,\mu-\tau,1,\tau}^{\nu,\mu} C_{\lambda,\mu-\tau,p,m-\mu}^{l,m-\tau} \right\}. \tag{31}$$

Numerical testing of Eqs. (29)–(31) showed they were numerically equivalent to those of Cruzan.[50] Recurrence formulas have additionally been derived for both the scalar and vector addition theorems, and significantly reduce the number of operations for computing the translation coefficients.[51–53] These formulas were expanded for elastic wave scattering and implemented in the iterative multipole model.

D. Multiple-scattering computations

The iterative multipole model first calculates the scattered wave fields for each particle in the system due to an initial plane wave. The scattered wave fields are then translated between all particle pairs and summed at each particle. The new incident wave field (initial plane wave+scattered wave fields) is then used to compute revised scattered wave fields from each particle. This process is repeated iteratively

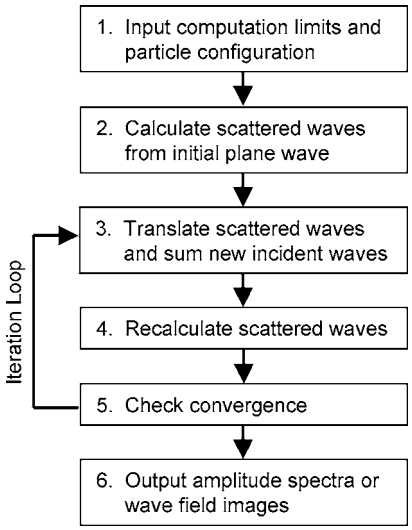


FIG. 3. Flow diagram of computation steps performed in the elastic wave scattering model.

until the scattered wave fields converge (i.e., no change in wave field amplitudes between consecutive iterations). Each iteration represents a successive order of scattering (first iteration=first-order multiple scattering, second iteration =second-order multiple scattering, etc.). Figure 3 illustrates the computation process.

Computation limits include the maximum multipole expansion order to compute ($n_{\max}$) and the convergence criteria (precision limit) for stopping the iterations. Computer algorithms for the model were written, debugged, and compiled in FORTRAN 90. Simulations were performed on personal desktop computers with up to 2.0 GB RAM and either a 1.7- or 3.2-GHz processor. Computation times varied from a few minutes to 36 h.

The model can be configured to provide wave field images, amplitude spectra, or elastic wave velocity. Wave velocity was computed using a Fourier transform method that modified the spectral input of the initial plane wave with appropriate wavelet functions [Fig. 4(a)], calculated the interactions of the modified wave in the simulated particle pack, and applied an inverse Fourier transform to the resultant spectra to simulate a wave pulse in the time domain. The time delay between this pulse and one with the simulated particle pack absent [Fig. 4(b)] provided the effective wave velocity. This procedure can also provide pulse attenuation and is equivalent to simulating the frequency characteristics of a transmitting ultrasonic transducer. Complex wave vectors were also implemented in the algorithm to model dispersions with viscoelastic or intrinsic attenuation properties.

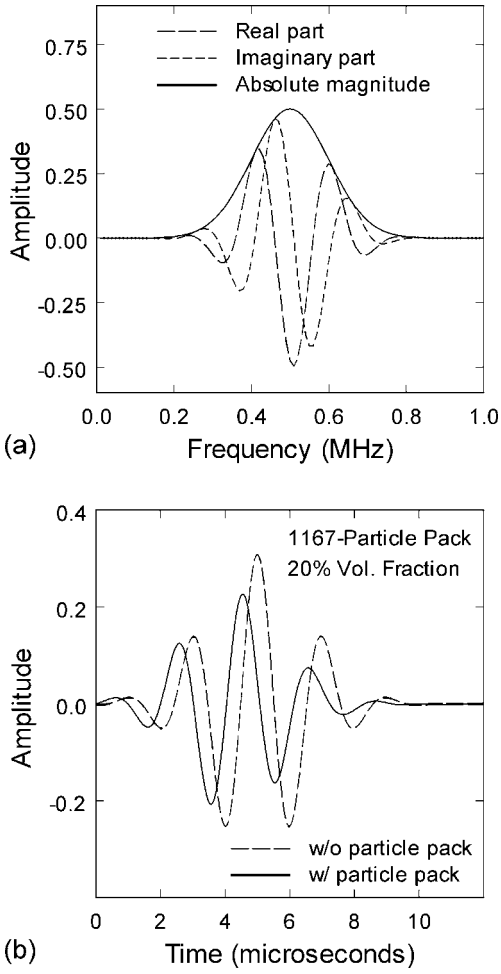


FIG. 4. Modulation of the initial plane wave spectrum using wavelet functions (a) and a time-delay comparison of resultant pulses for an 1167-particle dispersion (b).

III. RESULTS

A. Simulations of 2 to 16 particles

The scattering interactions between 2 and 16 particles were simulated to test the ability of the model to reproduce wave propagation phenomena. The particles were modeled as quartz spheres in a matrix of water ice since the elastic properties of quartz and ice are sufficiently different to clearly show scattering phenomena such as reflection, but not too different to preclude other types of wave propagation such as refraction through particles.

Figure 5 shows images of the scattered wave fields resulting from a longitudinal wave incident on two spheres aligned along the direction of wave propagation (from left to right in the images in the $+z$ direction). The incident wave is not superimposed on the scattered wave fields in order to emphasize the scattering mechanisms, and the gray scale is proportional to the scattered wave field amplitude. The images display both refractive focusing (forward scattering) of the longitudinal wave, Fig. 5(a), and mode conversion of the longitudinal wave to shear waves, Fig. 5(b).

Since most of the scattered wave behavior in Fig. 5 arises from single-sphere scattering, single-scattering computations were compared to the multiple-scattering computations to ascertain differences between the two solutions and determine the effects of multiple scattering. The single-scattering computations superimpose the single-scattering solutions from each of the spheres to calculate the effective field. Figure 6(a) shows the z component of the total displacement field (initial plane wave+multiple-scattered waves) for the two particles. Figure 6(b) is the difference image between the multiply scattered waves shown in Fig.

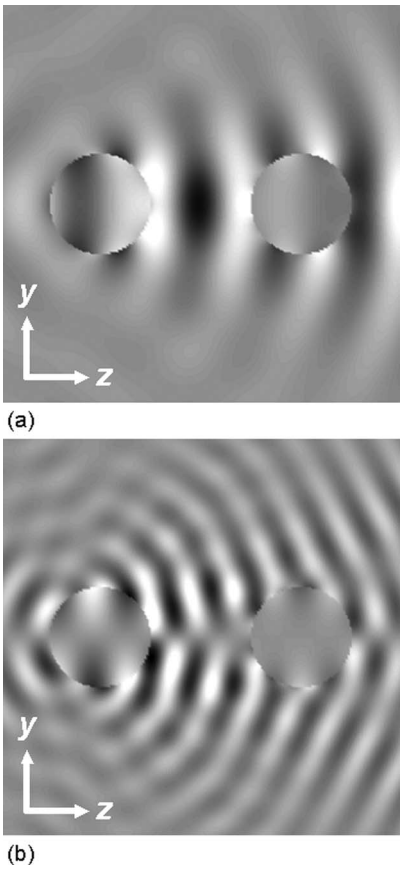


FIG. 5. Wave field images for a 5-MHz longitudinal wave incident on a pair of 1.0-mm-diam quartz spheres in ice: (a) scattered longitudinal wave (z component) displaying focusing and (b) scattered shear wave (y component) displaying mode conversion from longitudinal wave. Wave propagation in these and subsequent images is in the $+z$ direction.

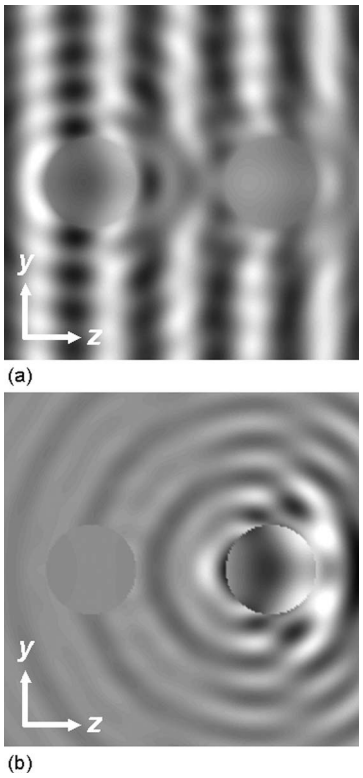


FIG. 6. Wave field images for a 5-MHz longitudinal wave incident on a pair of 1.0-mm-diam quartz spheres in ice: (a) total displacement field (z component) and (b) displacement differences (z component) between multiple-scattering and single-scattering computations.

6(a) and the wave fields generated by the single-scattering solutions. The difference plot highlights the spatial regions where multiple scattering has the largest contribution to the wave fields, and clearly shows the primary effect of multiple scattering in the two-sphere configuration is the shielding of the second (right) sphere by the first (left) sphere.

Figure 7 displays the propagation of a 1.0-MHz [Fig. 7(a)] and 2.0-MHz [Fig. 7(b)] shear plane wave through a random configuration of 16 quartz particles lying in a plane. Here the images show the x component of the total shear wave field (incident+scattered wave fields). The greater attenuation exhibited for the 2.0-MHz waves can be attributed to the fact that Mie scattering is inversely proportional to the wavelength. Figure 7 additionally shows distortion of the wave front from the initial plane-wave geometry for the 2.0-MHz image. The distortion arises from disorder in both particle size and position.

The results from simulations of small particle configurations demonstrated the model's ability to reproduce basic wave propagation behavior such as refraction, mode conversion, shielding, and wave-front distortion at a qualitative level. All of the above simulations were calculated to a maximum multipole order of $n_{\max}=12$.

B. Simulations of dispersions with 10^2–10^3 particles

Wave field images, amplitude spectra, and ultrasonic velocities were computed for larger 3-D particle packs (100–

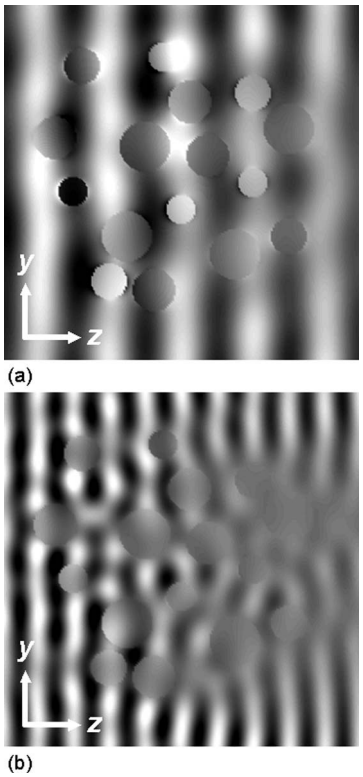


FIG. 7. Total (scattered+incident) shear wave x component for an incident shear wave at 1.0 MHz (a) and 2.0 MHz (b) for a disordered configuration of quartz particles in ice (diameter=0.8–1.4 mm).

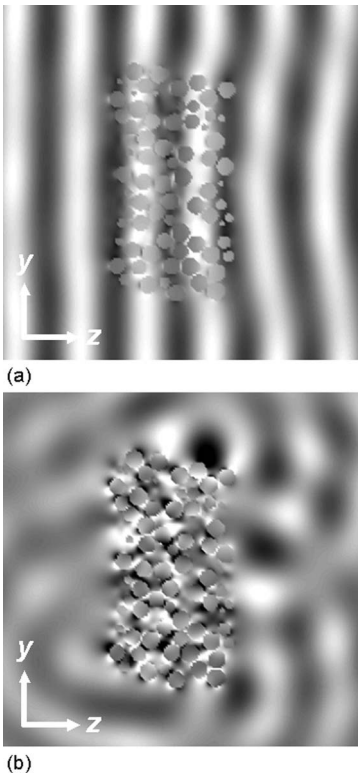


FIG. 8. Wave field images of the z component (a) and y component (b) of displacement resulting from an incident 2.0-MHz longitudinal wave in a cylindrical packing of 200-μm-diameter NaCl particles in rubber. Images are of a random packing of 736 particles at 50% volume fraction.

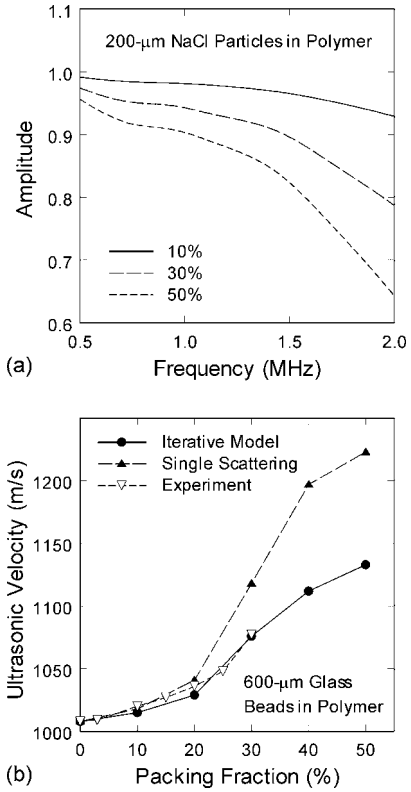


FIG. 9. Amplitude spectra and velocity results for large particle dispersions in a viscoelastic polymer matrix: (a) longitudinal wave spectra for 200-μm NaCl particles; and (b) comparison of longitudinal wave velocity at 500-kHz center frequency for multiple-scattering computations, single-scattering computations, and experimental measurements for 600-μm glass beads.

3000 particles at 10%–50% volume fraction) to test the model's ability to simulate material microstructures. The particle dispersions were configured into a cylindrical disk with the initial plane wave incident on one of the cylinder's end faces. For the simulation of spectra and velocities, the wave fields were evaluated on a planar array of points with a circular shape. This array simulated the face of an ultrasonic transducer by averaging the wave field amplitudes over a defined area. This averaging eliminated local interference effects that arose when the fields were evaluated at a single point. The simulations were computed to a maximum multipole order of $n_{\max}=6$ to make the computations tractable for a desktop computer. An effective interaction distance was also used that excluded less significant interactions between widely separated particles.

Wave field images were generated for both periodic and random arrangements of 200-μm diameter NaCl particles in a viscoelastic polymer matrix (rubber). Figure 8 displays wave field images for a random packing of 736 particles at 50% volume fraction. The image plane in Fig. 8 slices through the center of the cylinder along the cylinder's axis, and the incident wave is included in order to show the refraction of the elastic wave field by the random particle dispersion [Fig. 8(a)]. Since the incident longitudinal field is wholly in the z direction, it does not show up in the y-component image [Fig. 8(b)]. However, both single and multiple scattering redirect a portion of the wave field energy into the y-direction (i.e., refraction), thus producing y-component waves.

Figure 9 displays amplitude spectra and velocity results for large particle dispersions in a viscoelastic polymer matrix. Amplitude spectra for 200-μm NaCl particles [Fig. 9(a)] showed increases in attenuation (decreases in amplitude) with particle volume fraction and frequency, a result qualitatively consistent with experimental observations. Longitudinal wave velocities were additionally calculated for random dispersions of 600-μm glass beads and compared to velocities from both single-scattering computations and ultrasonic measurements of laboratory specimens [Fig. 9(b)]. The results from the iterative model show an increase in effective velocity with increasing packing density and are consistent with the experimental data up to 30%. In comparison, the single-scattering model agrees with the experimental data up to only 20% volume fraction. Although the experimental data only goes to 30% volume fraction, the iterative model appears to deviate from the trend of the experimental velocities above 30%, indicating the computations may not be accurate for dense packings. This deficiency can be attributed to a lack of iterative and addition theorem convergence (an $n_{\max}=6$ value was used for the simulations) resulting in increasingly larger errors for the multiple scattering. The particle packs in the velocity simulations varied from 148 particles (10% volume fraction) to 736 particles (50% volume fraction).

C. Convergence

The convergence of the iterative multipole model has three components: Addition theorem convergence, single-

scattering convergence, and iterative convergence. Although convergence of all three components is necessary for computational fidelity, they are not strongly coupled to one another, and convergence of any one component is not directly dependent on the convergence of the other two. Therefore, the properties of each must be considered separately.

1. Addition theorem convergence

The convergence of the addition theorems was numerically tested by translating a quadrupole moment ($n=2$, $m=1$) for all three elastic fields from one sphere (α) to a point on another sphere (β) and comparing the translated fields to the untranslated fields at that point (see Fig. 2). Convergence was evaluated as a function of $n_{\max}$, frequency, sphere size, sphere separation, orientation of the two spheres, and orientation of the evaluation point on sphere β.[45] The size and distance scale for sphere β was in the millimeter range, and the frequencies were in the 0.10–10.0-MHz range. Since the addition theorems couple the $\mathbf{V}$ and $\mathbf{W}$ fields [Eqs. (26) and (27)], the sum of the two transverse fields have to be compared instead of the two fields individually (i.e., $\mathbf{V}+\mathbf{W}$ is compared to $\mathbf{V}'+\mathbf{W}'$, rather than $\mathbf{V}$ to $\mathbf{V}'$ and $\mathbf{W}$ to $\mathbf{W}'$).

Following the coordinate conventions in Fig. 2, Fig. 10 displays convergence plots for sphere β positioned at $\Theta_{\alpha\beta}=34°$, $\Phi_{\alpha\beta}=295°$, and with four separation distances ($R_{\alpha\beta}=1.0$, 2.0, 3.0, and 4.0 mm). The evaluation point on sphere β was positioned at $\theta_\beta=163°$, $\varphi_\beta=320°$, and $a_\beta=0.5$ mm. Results are displayed for 0.10 MHz [Fig. 10(a)] and 5.0 MHz [Fig. 10(b)]. At low frequencies [Fig. 10(a)], the field translations converge more rapidly with $n_{\max}$ for greater sphere separations. However, as the frequency increases the convergence curves shift and eventually overlap at 5.0 MHz [Fig. 10(b)]. The curves continue to shift to higher $n_{\max}$ as frequency increases, indicating slower convergence of the addition theorem expansions.

The convergence curves in Fig. 10 indicate the degree of error present in the field translations for a specified $n_{\max}$, frequency, and particle separation. For example, the convergence curve for $R_{\alpha\beta}=2$ in Fig. 10(b) would correspond to the two-sphere simulations in Figs. 5 and 6. For this case, computations at $n_{\max}=6$ would result in field translation errors on the order of 10%, whereas computations at $n_{\max}=12$ would produce errors on the order of only 1%.

2. Single-scattering convergence

In multipole-based optical scattering models for single scatterers, the convergence of the scattered wave amplitudes as a function of frequency is dependent on the particle diameter d and $n_{\max}$. Simulation results verify the relationship between $n_{\max}$, d, and convergence frequency $\nu_{\max}$ (the highest frequency at which the wave fields converge) is proportional for elastic wave scattering as well. This relationship has been found to be approximately

$$\nu_{\max} \cong \frac{9}{10}\frac{c}{\pi d} n_{\max}. \tag{32}$$

where c is the longitudinal or shear wave velocity in the matrix. As an example, for longitudinal wave scattering from

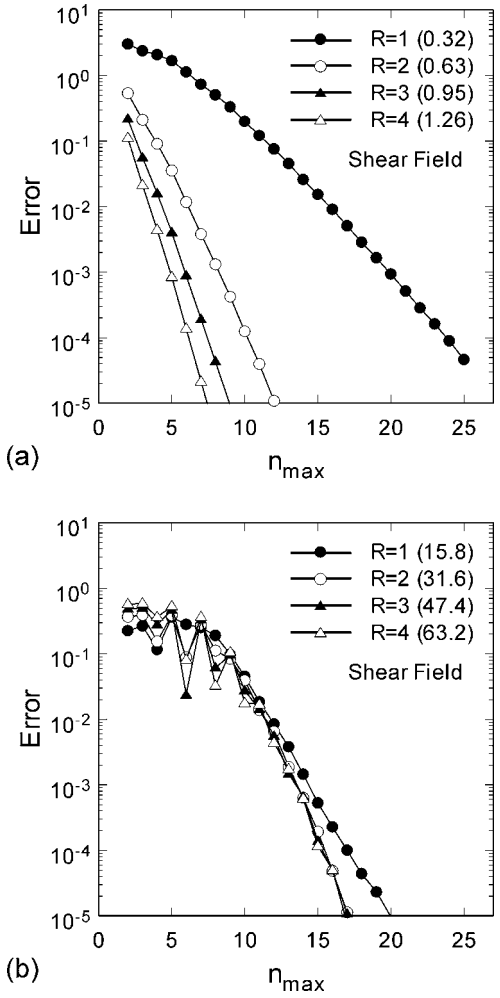


FIG. 10. Addition theorem convergence for translation of 0.10-MHz (a) and 5.0-MHz (b) shear fields between two 1.0-mm-diam quartz spheres in ice. $R=R_{\alpha\beta}$=center-to-center distance between spheres in mm. Numbers in parentheses are $kR_{\alpha\beta}$ values.

1-mm-diam quartz particles in ice ($c_L=3.98\times10^5$ cm/s) and for $n_{\max}=4$, the maximum frequency at which the spectrum converges is $\nu_{\max}\approx4.6$ MHz. Therefore the elastic wave spectra will be convergent for frequencies from 0 to 4.6 MHz, but higher $n_{\max}$ values will be required for convergence at higher frequencies. Lack of single-scattering convergence has been found to contribute significant error to multiple-scattering results in simulations of ordered and disordered clusters of eight particles.[45]

3. Iterative convergence

Iterative convergence assures that the back-and-forth scattering interactions modeled with each iteration step converge to stable wave field amplitudes. This is physically required since each successive multiple scattering should produce a diminishing contribution to the total field, eventually resulting in the final field configuration. Simulations of small particle packings (2–16 particles) have shown that all three elastic fields converge to within computational limits (10^{-16} for double precision) for arbitrary structures and a wide range of frequencies.

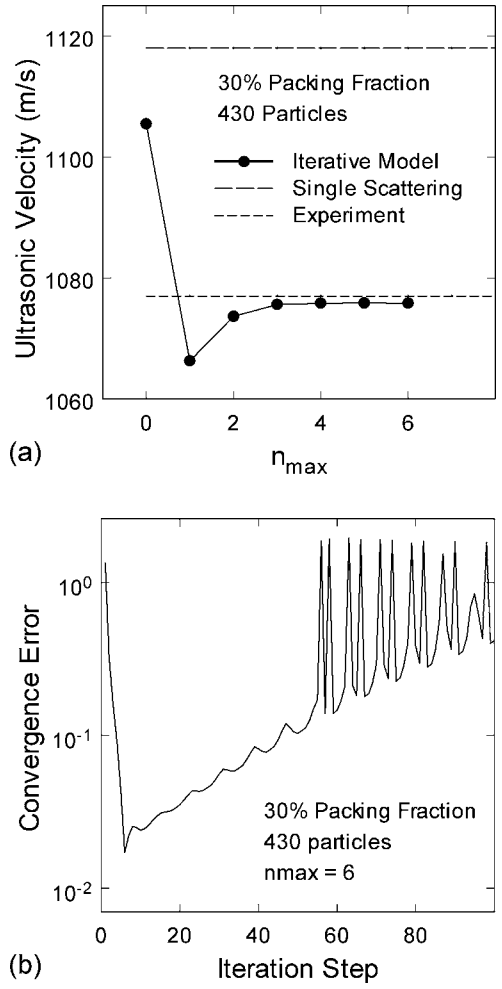


FIG. 11. Convergence behavior for 600-μm glass beads in polymer, 30% volume fraction: (a) predicted ultrasonic velocity as a function of n_{max} and (b) convergence error as a function of iteration step for the longitudinal field at 600 kHz.

Simulations of large particle packings (>100 particles) have revealed, however, that the elastic wave computations are numerically unstable and do not converge or only partially converge for short wavelengths ($\lambda \sim$ particle diameter). The shear waves contribute most significantly to the instability due to their shorter wavelengths in materials as compared to longitudinal waves. Damping the interactions with attenuation factors of a magnitude common in viscoelastic polymers has been found to reduce the instability. At short wavelengths the iterative model therefore appears better suited for particles embedded in viscoelastic materials (for example, Figs. 8 and 9) than in highly elastic materials. Another approach to resolving the nonconvergence is to relax the convergence criteria since the iterations for large particle packings often partially converge to errors less than 10^{-1} before instability causes the iterations to diverge. Iterative convergence has also been found to be strongly dependent on particle volume fraction, but only weakly dependent on n_{max}.

Figure 11 displays the convergence behavior of the velocity simulation results shown in Fig. 9(b) (the 600-μm glass bead/polymer system) for 30% volume fraction. The convergence of the predicted ultrasonic velocity as a function of n_{max}, Fig. 11(a), demonstrates the computed values converge rapidly by $n_{max}=6$ and closely approach the experimental value. Also, the $n_{max}=0$ simulation (monopole-only scattering) is closer in agreement with the single-scattering model than the higher n_{max} simulations, thereby showing the importance of the dipole, quadrupole, and octupole terms in multiple scattering.

Figure 11(b) presents the convergence error as a function of iteration step for the 30% volume fraction simulation. The convergence error is the fractional difference in the average computed wave field amplitude between iteration steps and is shown in Fig. 11(b) for the longitudinal field at 600 kHz and $n_{max}=6$. Extensive testing revealed that the velocity simulations for this system were only marginally stable, converging for the first five to ten iteration steps but then diverging as in Fig. 11(b). Additionally, the lowest error in the iterations varied with frequency from 10^{-1} to 10^{-6}. In this situation, defining the convergence criterion as the point at which the iterations reach a fixed error value such as 10^{-6} does not work, and the model fails to converge. Rather, limiting the number of iterations to a fixed value (for example, five iterations) allows results to be obtained from the model when the minimum convergence error varies across the frequency range. Although this produces a variable convergence error for the computed fields, it permits a maximum of frequency steps to be computed with an overall minimum error. (A more sophisticated approach would be to program the model to detect when such a minimum has passed and to use the results from the minimum error iteration step.)

The results from Fig. 11 illustrate the instability of the iterative model for particle packs greater than a few hundred particles, but also show how the instabilities do not preclude the acquisition of multiple scattering results. If the convergence error as a function of iteration step has a minimum with an error less the 10^{-1} as in Fig. 11(b), then simulation results can be obtained as in Fig. 11(a) that display reasonable convergence properties with respect to n_{max} and experimental data.

IV. DISCUSSION

The results to date indicate the iterative multipole model reproduces realistic wave propagation behavior at a qualitative level. Phenomena such as mode conversion, refraction, wave front distortion, and scattering-induced attenuation have been observed in both wave field images and spectra. The model can also simulate large numbers of particles (10^2–10^3) and random structures directly for viscoelastic materials. Random particle packings of up to 3000 particles and 50% particle volume fraction have also been tested. The iterations are numerically stable and converge for small particle configurations, but become unstable for large particle configurations. Simulations at long wavelengths or with strong attenuation of shear waves can improve the stability of the numerical process.

A common approach to validating numerical-based multiple-scattering models is to compare the results for a small cluster of spheres to other numerical approaches such as the boundary element method.[17] This approach is feasible for clusters of two or three spheres, but becomes computa-

tionally intractable for large particle packings with more than a few dozen particles. Alternatively, model results can be compared to experimental data. This approach has been used extensively for evaluating classical formulations where multiple scattering is treated statistically.[1,2] Analogous experimental validation of the iterative multipole method was therefore implemented as a test for the model. Comparisons between multiple-scattering and single-scattering computations were additionally used to quantify the magnitude of multiple scattering in a particulate system and to establish the limits of validity for the computations.

The results indicate that both addition theorem and iterative convergence are the current limiting factors for the iterative multipole approach implemented on a personal computer. Due to the virtual memory requirements involved with the addition theorems, the maximum computed multipole order has so far been limited to $n_{\max}=6$ for large dispersions (10^2-10^3 particles). Convergence studies of the addition theorems (Fig. 10) and model comparisons with experimental data [Figs. 9(b) and 11(a)] indicate low $n_{\max}$ values ($n_{\max}\leqslant 6$) are sufficient for particle dispersions up to 30% particle volume fraction but not for denser packings. Consequently, simulations performed to date with high particle densities have exhibited significant error due to inadequate convergence of both the field translations and iterations.

The results also indicate dilute dispersions ($\leqslant 20\%$) are not sufficient to validate numerical models for multiple scattering since single scattering will dominate over multiple scattering in such systems. In such cases the models may produce accurate results for a dilute particle dispersion although the multiple scattering may be insufficiently or incorrectly calculated. As shown in Fig. 9(b), a single-scattering model shows good agreement with experimental results for volume fractions up to 20%. Validation with dense particle systems is therefore necessary for accurately evaluating a model's performance. Since iterative convergence, addition theorem convergence, and single-scattering convergence are not strongly coupled, care must also be taken in interpreting the accuracy of a multiple-scattering model with regards to any single convergence component.

The iterative or order-of-scattering method appears to be a direct and accessible approach for modeling elastic wave scattering in arbitrary dispersions of spherical particles. Computational modeling of multiple scattering at moderate volume fractions (30%) and particle numbers (~400 particles) has been demonstrated and verified against experimental data. It is a simple and highly generalized approach that can be applied to dispersions of spherical scatterers of almost any configuration and composition. Although the model currently has significant computational and convergence limitations, it is likely methods can be found or developed to resolve these issues. Such methods include implementation of the model on larger computer systems and development of more advanced scattering algorithms to resolve the efficiency and stability constraints.

V. CONCLUSIONS

An iterative multipole method has been developed to model arbitrary configurations of spherical, elastic particles in either an elastic or viscoelastic matrix and has been implemented on a personal computer to test its functioning and performance. Initial results for systems of two to several hundred particles show that the model predicts many aspects of elastic wave behavior at the qualitative level. The model is also quantitatively accurate for simulating multiple scattering in dispersions with moderate particle densities ($\leqslant 30\%$). At higher particle densities, the model predictions deviate due to limitations in computing sufficiently high multipole orders ($n_{\max}=7-16$) for convergence of the addition theorems. The current model is also numerically unstable for large particle configurations in elastic materials at short wavelengths ($\lambda\sim$ particle diameter), but is partially to fully stable for either elastic materials at long wavelengths or viscoelastic materials.

ACKNOWLEDGMENTS

This work was supported by ATK Thiokol and the Air Force Research Laboratory under Contract No. F04611-98-C-0005. The author thanks I. Lee Davis, Lee H. Pearson, J. R. Dennison, R. Scott Hyde, Gregory A. Ruderman, Brent L. Carruth, and Keith H. Warnick for their assistance.

APPENDIX

The η matrix elements from the elastic wave boundary condition matrix, Eq. (23), are presented below. The matrix elements are functions of the multipole order N, longitudinal wave vector k_L, shear wave vector k_S, and the sphere radius a. The spherical radial functions are denoted by $z_n(kr)$ and vary according to the radial function designated by the subscripts in Eq. (23) as $j=j_n(kr)$, $h=h_n^{(1)}(kr)$, or $g=h_n^{(2)}(kr)$:

$$\eta_{1z}=N\frac{z_N(k_La)}{k_La}-z_{N+1}(k_La), \tag{A1}$$

$$\eta_{2z}=i\sqrt{(N)(N+1)}\frac{z_N(k_Sa)}{k_Sa}, \tag{A2}$$

$$\eta_{3z}=z_N(k_Sa), \tag{A3}$$

$$\eta_{4z}=\frac{z_N(k_La)}{k_La}, \tag{A4}$$

$$\eta_{5z}=\frac{i}{\sqrt{(N)(N+1)}}\left[(N+1)\frac{z_N(k_Sa)}{k_Sa}-z_{N+1}(k_Sa)\right], \tag{A5}$$

$$\eta_{6z}=-\lambda k_Lz_N(k_La)+2\mu k_L\left\{\left[\frac{(N-1)(N)}{(k_La)^2}-1\right]z_N(k_La)+2\frac{z_{N+1}(k_La)}{k_La}\right\}, \tag{A6}$$

$$\eta_{7z}=2i\mu k_S\sqrt{(N)(N+1)}\left[(N-1)\frac{z_N(k_Sa)}{(k_Sa)^2}-\frac{z_{N+1}(k_Sa)}{k_Sa}\right], \tag{A7}$$

$$\eta_{8z} = \mu k_S \left[(N-1)\frac{z_N(k_S a)}{k_S a} - z_{N+1}(k_S a) \right], \tag{A8}$$

$$\eta_{9z} = 2\mu k_L \left[(N-1)\frac{z_N(k_L a)}{(k_L a)^2} - \frac{z_{N+1}(k_L a)}{k_L a} \right], \tag{A9}$$

$$\eta_{10z} = \frac{i\mu k_S}{\sqrt{(N)(N+1)}} \left[\left[\frac{2(N^2-1)}{(k_S a)^2} - 1 \right] z_N(k_S a) + 2\frac{z_{N+1}(k_S a)}{k_S a} \right]. \tag{A10}$$

[1] L. W. Anson and R. C. Chivers, "Ultrasonic velocity in suspensions of solids in solids—a comparison of theory and experiment," J. Phys. D **26**, 1506–1575 (1993).
[2] R. E. Challis, J. S. Tebbutt, and A. K. Holmes, "Equivalence between three scattering formulations for ultrasonic wave propagation in particulate mixtures," J. Phys. D **31**, 3481–3497 (1998).
[3] R. E. Baddour, M. D. Sherar, J. W. Hunt, G. J. Czarnota, and M. C. Kolios, "High-frequency ultrasound scattering from microspheres and single cells," J. Acoust. Soc. Am. **117**, 934–943 (2005).
[4] F. E. Stanke and G. S. Kino, "A unified theory for elastic wave propagation in polycrystalline materials," J. Acoust. Soc. Am. **75**, 665–681 (1984).
[5] S. Hirsekorn, "The scattering of ultrasonic waves by polycrystals," J. Acoust. Soc. Am. **72**, 1021–1031 (1982).
[6] S. Hirsekorn, "The scattering of ultrasonic waves by polycrystals. II. shear waves," J. Acoust. Soc. Am. **73**, 1160–1163 (1983).
[7] S. Hirsekorn, "The scattering of ultrasonic waves in polycrystalline materials with texture," J. Acoust. Soc. Am. **77**, 832–843 (1985).
[8] S. Hirsekorn, "Directional dependence of ultrasonic propagation in textured polycrystals," J. Acoust. Soc. Am. **79**, 1269–1279 (1986).
[9] J. A. Turner and R. L. Weaver, "Time dependence of multiply scattered diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **97**, 2639–2644 (1994).
[10] J. A. Turner and R. L. Weaver, "Ultrasonic radiative transfer in polycrystalline media: Effects of a fluid-solid interface," J. Acoust. Soc. Am. **98**, 2801–2808 (1995).
[11] R. L. Weaver, "Diffusivity of ultrasound in polycrystals," J. Mech. Phys. Solids **38**, 55–86 (1990).
[12] J. A. Turner and R. L. Weaver, "Radiative transfer and multiple scattering of diffuse ultrasound in polycrystalline media," J. Acoust. Soc. Am. **96**, 3675–3683 (1994).
[13] J. A. Turner, "Elastic wave propagation and scattering in heterogeneous, anisotropic media: Textured polycrystalline materials," J. Acoust. Soc. Am. **106**, 541–552 (1999).
[14] J. A. Turner and P. Anugonda, "Scattering of elastic waves in heterogeneous media with local isotropy," J. Acoust. Soc. Am. **109**, 1787–1795 (2001).
[15] B. Peterson and S. Ström, "Matrix formulation of acoustic scattering from an arbitrary number of scatterers," J. Acoust. Soc. Am. **56**, 771–780 (1974).
[16] S. Koc and W. C. Chew, "Calculation of acoustical scattering from a cluster of scatterers," J. Acoust. Soc. Am. **103**, 721–734 (1998).
[17] N. A. Gumerov and R. Duraiswami, "Computation of scattering from *N* spheres using multipole reexpansion," J. Acoust. Soc. Am. **112**, 2688–2701 (2002).
[18] N. A. Gumerov and R. Duraiswami, "Computation of scattering from clusters of spheres using the fast multipole method," J. Acoust. Soc. Am. **117**, 1744–1761 (2005).
[19] Z. Liu, C. T. Chan, P. Sheng, A. L. Goertzen, and J. H. Page, "Elastic wave scattering by periodic structures of spherical objects: theory and experiment," Phys. Rev. B **62**, 2446–2457 (2000).
[20] M. I. Mishchenko, G. Videen, V. A. Babenko, N. G. Khlebtsov, and T. Wriedt, "T-matrix theory of electromagnetic scattering by particles and its applications: a comprehensive reference database," J. Quant. Spectrosc. Radiat. Transf. **88**, 357–406 (2004).
[21] P. C. Waterman, "Matrix theory of elastic wave scattering," J. Acoust. Soc. Am. **60**, 567–580 (1976).
[22] Y.-H. Pao, "Betti's identity and transition matrix for elastic waves," J. Acoust. Soc. Am. **64**, 302–310 (1978).
[23] A. Boström, "Multiple scattering of elastic waves by bounded obstacles," J. Acoust. Soc. Am. **67**, 399–413 (1980).
[24] A. Boström, "Scattering by a smooth elastic obstacle," J. Acoust. Soc. Am. **67**, 1904–1913 (1980).
[25] I. L. Davis, "Electric field in an arbitrary random pack of spherical particles," J. Appl. Phys. **67**, 955–964 (1990).
[26] A-K. Hamid, I. R. Ciric, and M. Hamid, "Iterative solution of the scattering by an arbitrary configuration of conducting or dielectric spheres," IEE Proc., Part H: Microwaves, Antennas Propag. **138**, 565–572 (1991).
[27] A-K. Hamid, I. R. Ciric, and M. Hamid, "Multiple scattering by a linear array of conducting spheres," Can. J. Phys. **68**, 1157–1168 (1990).
[28] A-K. Hamid, I. R. Ciric, and M. Hamid, "Electromagnetic scattering by an arbitrary configuration of dielectric spheres," Can. J. Phys. **68**, 1419–1428 (1990).
[29] K. A. Fuller and G. W. Kattawar, "Consummate solution to the problem of classical electromagnetic scattering by an ensemble of spheres. II. Clusters of arbitrary configuration," Opt. Lett. **13**, 1063–1065 (1988).
[30] D. W. Mackowski, "Analysis of radiative scattering for multiple sphere configurations," Proc. R. Soc. London, Ser. A **433**, 599–614 (1991).
[31] W. Greiner and J. A. Maruhn, *Nuclear Models* (Springer-Verlag, Berlin, 1996), pp. 77–85.
[32] K. S. Thorne, "Multipole expansions of gravitational radiation," Rev. Mod. Phys. **52**, 299–339 (1980).
[33] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988), Chaps. 5, 7, and 8.
[34] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), Chap. 7.
[35] M. E. Rose, *Elementary Theory of Angular Momentum* (Wiley, New York, 1957), pp. 131–139.
[36] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, NJ, 1957), pp. 81–85.
[37] J. D. Jackson, *Classical Electrodynamics*, 2nd ed. (Wiley, New York, 1975), pp. 767–769.
[38] W. T. Grandy, Jr., *Scattering of Waves from Large Spheres* (Cambridge U.P., Cambridge, UK, 2000), pp. 67–69.
[39] C. F. Ying and R. Truell, "Scattering of a plane longitudinal wave by a spherical obstacle in an isotropically elastic solid," J. Appl. Phys. **27**, 1086–1097 (1956).
[40] L. Knopoff, "Scattering of shear waves by spherical obstacles," Geophysics **24**, 209–219 (1959).
[41] L. Knopoff, "Scattering of compression waves by spherical obstacles," Geophysics **24**, 30–39 (1959).
[42] N. G. Einspruch, E. J. Witterholt, and R. Truell, "Scattering of a plane transverse wave by a spherical obstacle in an elastic medium," J. Appl. Phys. **31**, 806–818 (1960).
[43] D. L. Jain and R. P. Kanwal, "Scattering of P and S waves by spherical inclusions and cavities," J. Sound Vib. **57**, 171–202 (1978).
[44] M. K. Hinders, "Plane-elastic-wave scattering from an elastic sphere," Nuovo Cimento Soc. Ital. Fis., B **106B**, 799–817 (1991).
[45] T. E. Doyle, "Computational scattering models for elastic and electromagnetic waves in particulate media," Ph.D. dissertation, Utah State University, Logan, UT, 2004.
[46] B. Friedman and J. Russek, "Addition theorems for spherical waves," Q. Appl. Math. **12**, 13–23 (1954).
[47] M. E. Rose, "The electrostatic interaction of two arbitrary charge distributions," J. Math. Phys. (Cambridge, Mass.) **37**, 215–222 (1958).
[48] S. Stein, "Addition theorems for spherical wave functions," Q. Appl. Math. **19**, 15–24 (1961).
[49] R. A. Sack, "Three-dimensional addition theorem for arbitrary functions involving expansions in spherical harmonics," J. Math. Phys. **5**, 252–259 (1964).
[50] O. R. Cruzan, "Translational addition theorems for spherical vector wave functions," Q. Appl. Math. **20**, 33–40 (1962).
[51] W. C. Chew, "Recurrence relations for three-dimensional scalar vector addition theorem," J. Electromagn. Waves Appl. **6**, 133–142 (1992).
[52] W. C. Chew and Y. M. Wang, "Efficient ways to compute the vector addition theorem," J. Electromagn. Waves Appl. **7**, 651–665 (1993).
[53] D. W. Mackowski, "Calculation of total cross sections of multiple-sphere clusters," J. Opt. Soc. Am. A **11**, 2851–2861 (1994).

Paranasal sinus ventilation by humming

Svante Granqvist and Johan Sundberg[a)]
Department of Speech Music Hearing, KTH, Stockholm, Sweden

Jon O. Lundberg and Eddie Weitzberg
Department of Physiology and Pharmacology, Karolinska Institute, Stockholm, Sweden

(Received 21 June 2005; revised 16 February 2006; accepted 17 February 2006)

The gas nitric oxide (NO) is generated in the human paranasal sinuses and can be measured in nasally exhaled air. During humming, a marked increase in exhaled NO content has been observed. The acoustic phenomenon responsible for this evacuation of NO gas from the sinuses was analyzed. A tube model was constructed with a syringe containing NO gas attached radially. This tube was excited with an air stream modulated by a sine wave. Increased evacuation was observed whenever the syringe was not located at a pressure node of the exciting sine wave. A computer model of the system showed a good matching of observed pressure versus frequency data in the syringe resonator. The results thus suggest that the alternating pressure in the nasal cavity forces the air plug in the ostium of the paranasal sinus resonators to vibrate, thus expelling from the cavity NO gas, which is transported to free air by the exhalatory air stream. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188887]

PACS number(s): 43.20.Ks, 43.70.-h, 43.70.Aj, 43.70.Bk [DOS] Pages: 2611–2617

I. INTRODUCTION

The maxillary sinuses, located in the skull bone just below the eyes (Fig. 1), contain an enzyme that produces large amounts of nitric oxide (NO) gas. NO is thought to play an important role in the airway immune defense system (Lundberg *et al.*, 1995).

The volume of the sinus cavities is on the order of a few cubic centimeters and they are connected to the nasal cavity via narrow bony passages, the sinus ostia (Knops *et al.*, 1993; Aust and Drettner, 1974). Hence, acoustically they can be regarded as Helmholtz resonators.

Recent investigations have shown that sinus NO is efficiently evacuated during the production of nasal murmur or humming (Weitzberg and Lundberg, 2002); great increases of the NO content in the air exhaled have been measured during humming, as illustrated in Fig. 2.

It can be hypothesized that the sound field present in the nasal tract during humming is responsible for the evacuation of the sinus NO gas. The aim of the present experiments was to test this hypothesis.

II. METHOD

Nasal murmur is associated with a compound nasal air stream containing a static and an alternating component. An air stream from a pressure tank was conducted through a tube. An alternating pressure component was added to this air stream. The alternating component was obtained from a loudspeaker enclosed in a 2 l box, attached to the tube by means of a T-joint. Three experiments were carried out with this compound airflow.

In Experiment A, the alternating pressure component consisted of a 20 s sine sweep from 800 down to 50 Hz. This compound airflow was fed to one nostril through an olive-shaped mouthpiece while the subjects held their breath and kept the velopharyngeal port closed. A rapid-response chemiluminescence system (Aerocrine AB, Stockholm, Sweden) was used to monitor the NO gas content in the air stream through the other nostril. This system was attached to another olive-shaped mouthpiece that completely sealed off the subject's nostril. The signal output from this device was connected to a computer-based system (Aerocrine NO system; Aerocrine AB), which yielded an instant on-screen display of flow, pressure, NO concentration, and NO output. The data were also recorded to computer files.

In Experiment B the air stream was passed through a tube, $L=700$ mm, inner diameter 20 mm, modeling the nasal tract, see Fig. 3. The compound airflow was fed to one end of the 700-mm-long tube (location marked "A" in the figure). The size of the T-joint was a few cm from the end of the tube. A syringe, modeling a maxillary sinus, ID 28.7 mm and with volume variable between 2.7 and 60 ml, was inserted radially into the tube. The diameter of the syringe tip, representing the ostium, was 1.7 mm and its length 9 mm. The Helmholtz resonance frequency of the syringe was varied by choosing volumes ranging between 40 and 2.7 ml. To estimate the resulting Helmholtz frequency the transfer function of the system, including the tube and the syringe, was recorded with the microphone at the free end of the tube. The frequency of the zero thus observed was accepted as the Helmholtz frequency of the syringe. These frequencies are listed in Table I.

The syringe was filled with NO gas 10 ppm (parts per million, AGA AB, Stockholm, Sweden). This low concentration does not affect the speed of sound appreciably. The NO content of the output gas was measured downstream of the syringe by means of the same chemiluminescence system, attached to a thin catheter that was inserted into the tube.

[a)]Electronic mail: johan.sundberg@speech.kth.se

FIG. 1. Coronal section of a male showing the location of the maxillary sinuses in the skull bone.

Sound was recorded by an electret microphone, diameter 7 mm, length 16 mm (AV-JEFE TCM 110) attached to the microphone input of the computer.

In Experiment C a computer model was developed in order to examine the acoustic flow in the syringe neck, see Fig. 4. This model was compared with a tube model similar to that used in Experiment B, although without flow meter, without the connection to the pressure tank (marked "A" in Fig. 3), and without the constriction corresponding to the gas meter. The computer model included a loudspeaker model, two cavities (C_{v1} and C_{v2}), a lumped-element model of the tube, a radiation impedance (M_{al} and R_{al}), and a Helmholtz resonator. The Helmholtz resonator was modeled as a series resonator (C_{ah}, R_{ah}, M_{ah}). The tube was modeled by a number of sections (typically 26) each consisting of an acoustic mass, a compliance and two resistances ($M_{at}, C_{at}, R_{atp}, R_{ats}$). However, the model did not include the static flow T-joint.

III. RESULTS

It seemed reasonable to assume that the NO evacuation resulted from a driving of the airplug in the ostia of the

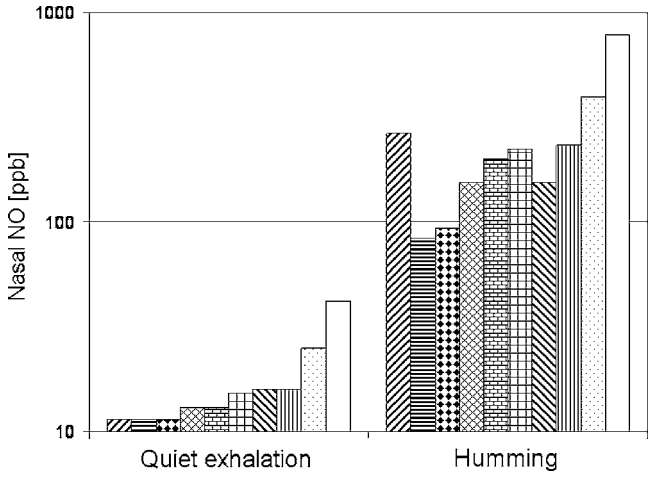


FIG. 2. NO content in parts per billion (ppb) in the air exhaled during silent exhalation and during humming (left and right columns). Column colors refer to subjects. Data from Weizberg and Lundberg, 2002.

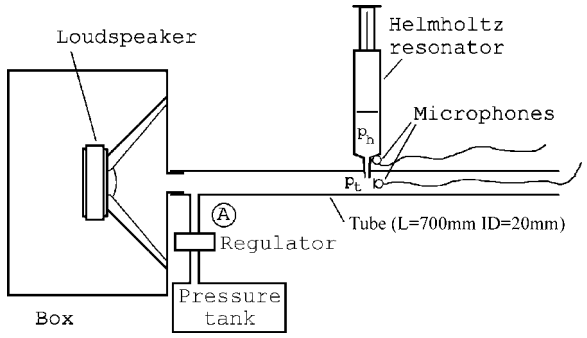


FIG. 3. Tube model of the nasal system, including the excitation system, used in Experiment B. One of the maxillary sinuses is represented by a syringe forming a Helmholtz resonator. The sound pressures in the Helmholtz resonator p_h and in the tube p_t were recorded by the microphones.

maxillary sinuses into and out of the resonator. It could further be assumed that this driving was caused by the sound wave in the nasal tract. If this assumption is correct, one might expect that for a constant excitation pressure the evacuation would reach a maximum at the Helmholtz resonance frequency, so the effect should vary with the frequency of the alternating component of the airflow. Also, the excitation of the airplug in the neck of the syringe should be influenced by the pressure amplitude at the neck location in the tube, which depends both on the standing wave pattern and the Helmholtz resonator's loading of the tube. If the neck is located in a pressure node, no evacuation should occur.

Experiment A, where the compound airflow was fed to a nostril, was carried out with two males and two females. As can be seen in Fig. 5 the NO content varied substantially during the sweep and peaked at different frequencies in these two subjects.

The lowest resonance frequencies of the entire tube system used in Experiment B with the compound airflow were measured by sine sweep in the absence of the syringe resonator. They were found to be about 250, 530, and 750 Hz, so the second resonance had a pressure node near $\frac{1}{2}$ of the wavelength, i.e., near $\frac{1}{2} \cdot c/530 = 0.33$ m from the tube end, and an antinode near $\frac{1}{4}$ of the wavelength, i.e., near $\frac{1}{4} \cdot c/530 = 0.16$ m from the tube end. A Helmholtz resonator located at a pressure node will not affect the corresponding tube resonance frequency, but in other locations the pressure distribution along the tube will be altered.

Figure 6 illustrates the relevance of the location of the resonator relative to the standing wave pattern of the tube. In the left graph the resonator was placed at the node of the second resonance, and in the right it was placed off this

TABLE I. Volumes and estimated Helmholtz frequencies of the syringe used in Experiment B.

Volume (ml)	Frequency (Hz)
40	185
20	250
10	305
5	360
2.7	475

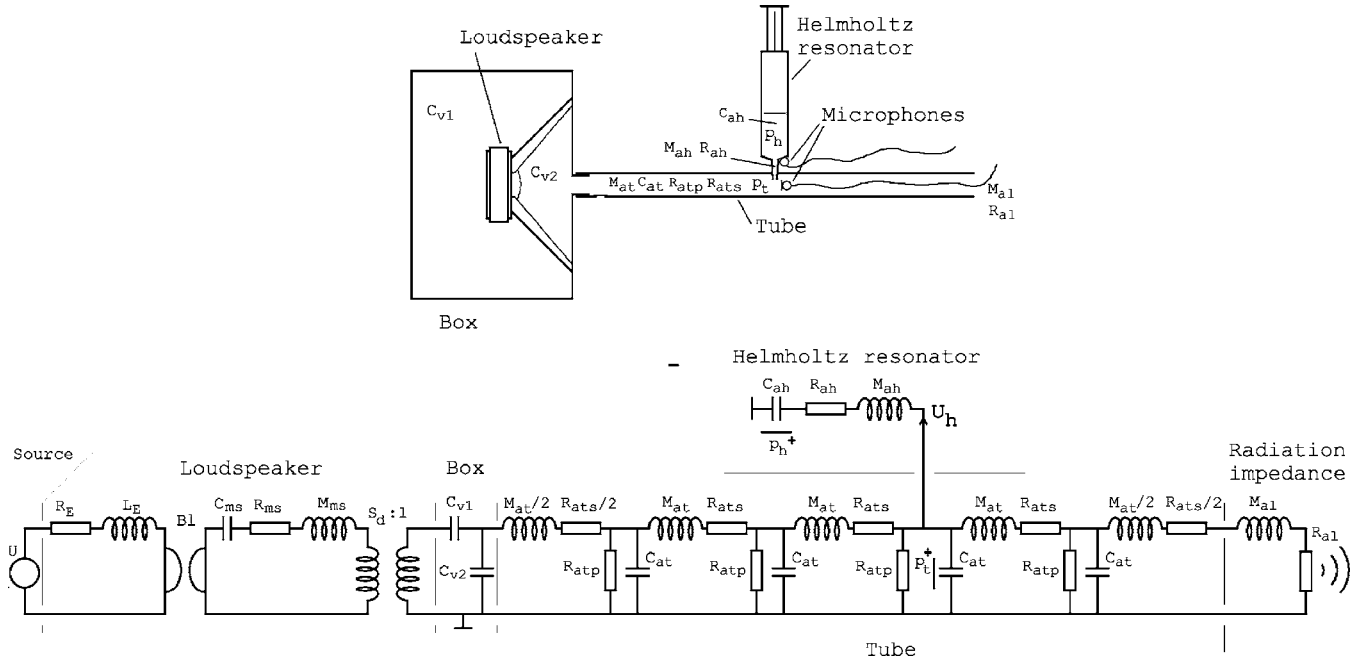


FIG. 4. Tube model and its electrical analog (upper and lower figures). The tube model included a classic loudspeaker model, two cavities (C_{v1} and C_{v2}), a lumped-element model of the tube, a radiation impedance (M_{al} and R_{al}), and a number of sections, each consisting of an acoustic mass, a compliance and two resistances ($M_{at}, C_{at}, R_{atp}, R_{ats}$). The Helmholtz resonator was modeled as a series resonator (C_{ah}, R_{ah}, M_{ah}). In the electrical analog only four tube sections are shown. The flow U_h is of prime interest for the ventilation of NO gas from the Helmholtz cavity.

node, approximately at $L/4$, i.e., about 0.17 m from the free end of the tube. In both cases the excitation frequency was manually varied according to the sequence F2, F1, F1, F2, remaining at each frequency for about 5 s and with an interval of at least 1 s between excitations. The volume of the resonator was chosen such that the resonance appeared at 183 Hz. The occurrence of NO gas did not exceed the background level when the resonator was located at the F2 node and approximately 10 ppb (parts per billion) in the other case. The evacuation was quite efficient at 249 Hz.

The relevance of the Helmholtz resonance is illustrated in Fig. 7 showing the evacuation of the NO gas at 249 and 530 Hz, i.e., at the first and second tube resonance. The evacuation was more efficient when the Helmholtz frequency was 250 Hz than when it was tuned to 305 Hz.

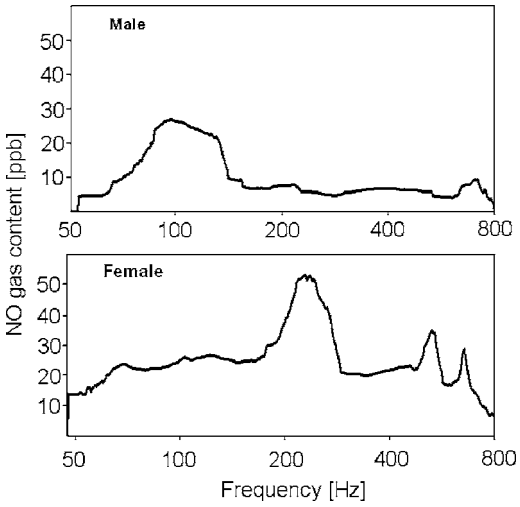


FIG. 5. NO-gas content in ppb obtained from one nostril of a male and a female subject, when a sine sweep from 800 down to 50 Hz plus a dc airflow was fed to the other nostril while the subjects held their breath and kept their velopharyngeal port closed.

These observations illustrate the complex acoustic conditions of the system. To gain a better understanding of this system, co-author S.G. developed the computer model, described earlier (see Fig. 4). This model was tested in Experiment C.

The inclusion in the model of the loudspeaker and the cavities of the box were essential for a good fit with the data observed in the tube model. The loudspeaker and the box were simulated as lumped elements (see, e.g., Small, 1972) using parameters from the manufacturer's datasheet, except for the voice coil inductance, which was considered lossy according to Leach (2002). The values of the lossy inductance were obtained by matching the model against the electrical impedance curve supplied by the manufacturer.

The compliances C_{v1} and C_{v2} were calculated as

$$C_{v1} = \frac{V_1}{\rho_0 c^2}, \tag{1}$$

$$C_{v2} = \frac{V_2}{\rho_0 c^2}, \tag{2}$$

where V_1 and V_2 are the volumes of the cavities, ρ_0 is the density of air (1.2 kg/m^3), and c is the sound velocity in air (345 m/s). The lumped elements of the tube were calculated as

$$M_{at} = \frac{L}{n\rho_0 S_t}, \tag{3}$$

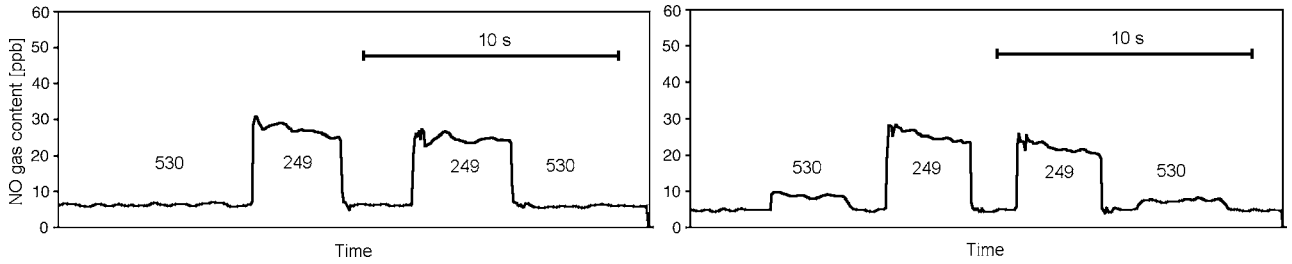


FIG. 6. NO-gas content in ppb observed in the tube model when it was excited with sinewaves at 530, 249, 249, and 530 Hz corresponding to the two lowest resonances in the tube in the absence of the Helmholtz resonator. In the left graph the resonator, tuned to 183 Hz with a 40 ml volume, was placed at the node of the second resonance (530 Hz), and in the right it was placed off this node.

$$C_{at} = \frac{S_t L}{n\rho_0 c^2}, \tag{4}$$

$$R_{atp} = \sqrt{\frac{M_{at}}{C_{at}}} \cdot Q_{\text{tubep}} \cdot n, \tag{5}$$

$$R_{ats} = \frac{\sqrt{M_{at}/C_{at}}}{n \cdot Q_{\text{tubes}}}, \tag{6}$$

where S_t is the cross-sectional area of the tube, L is the total length of the tube, Q_{tubes} and Q_{tubep} are values describing the losses, and n is the number of tube sections.

The Helmholtz resonator was modeled by entering the volume of the syringe cavity V_h, the Helmholtz frequency f_h, and a Q value Q_h. The software also optionally provided a possibility to lock the value of M_{ah} and recalculate f_h when V_h was changed [Eq. (10)],

$$C_{ah} = \frac{V_h}{\rho_0 c^2}, \tag{7}$$

$$M_{ah} = \frac{1}{C_{ah}(2\pi f_h)^2}, \tag{8}$$

$$R_{ah} = \frac{\sqrt{M_{ah}/C_{ah}}}{Q_h}, \tag{9}$$

$$f_h = \frac{1}{2\pi\sqrt{M_{ah}C_{ah}}}. \tag{10}$$

The radiation impedance was modeled as the low-frequency behavior of a baffled piston:

$$R_{al} = \frac{\rho_0}{2\pi c}\omega^2, \tag{11}$$

$$M_{at} = \frac{8\rho_0}{3\pi r_t}, \tag{12}$$

where ω is the angular frequency and r_t is the radius of the tube. The use of the baffled piston impedance is an approximation; the real tube end was not baffled. However, the error caused by this approximation can be assumed to be small, since the radiation impedance of both a baffled piston and an unbaffled tube is small as compared to the impedance of the tube (Beranek, 1954).

The results obtained from this computer model were compared to those obtained from a simplified tube model, without the static flow T-joint and the gas meter. Moreover, for practical reasons a different syringe was used from that used in Experiment B.

Figure 8 shows the agreement between the two models with respect to the pressure driving the resonator, and the pressure inside the resonator cavity. The testing was performed for resonator volumes V_h of 5, 10, 20, 40, and 60 ml. The associated Helmholtz resonance frequencies are listed in Table II. The results show that the computer model was capable of predicting the behavior of the physical model up to about 2 kHz. Some of the parameters had to be manually adjusted in order to make the curves fit. This was particularly true for the resistive losses, which are hard to estimate from the geometry of the model. However, once a good match was achieved, all parameters were kept constant, except for the V_h parameter which was set to the values above-listed.

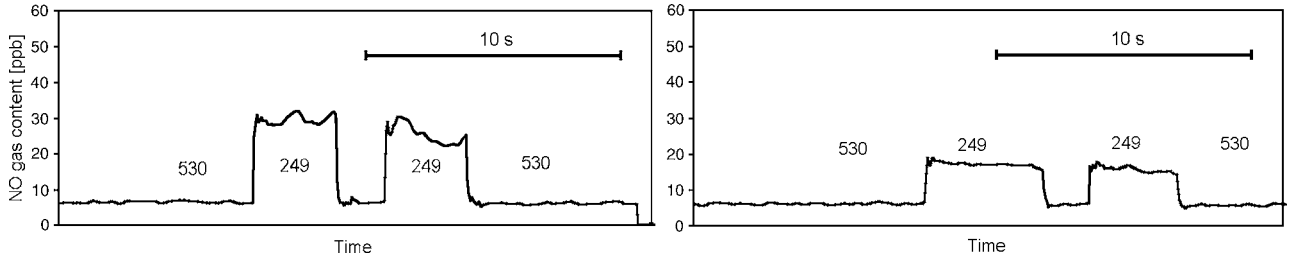


FIG. 7. NO-gas content in ppb observed in the tube model when it was excited with sinewaves at 530, 249, 249, and 530 Hz corresponding to the two lowest resonances in the tube in the absence of the Helmholtz resonator. The Helmholtz resonator was placed at the F2 node, near $L/2$. The resonator volumes were 20 and 10 ml, resulting in Helmholtz frequences at 255 Hz and to 305 Hz (left and right graphs, respectively).

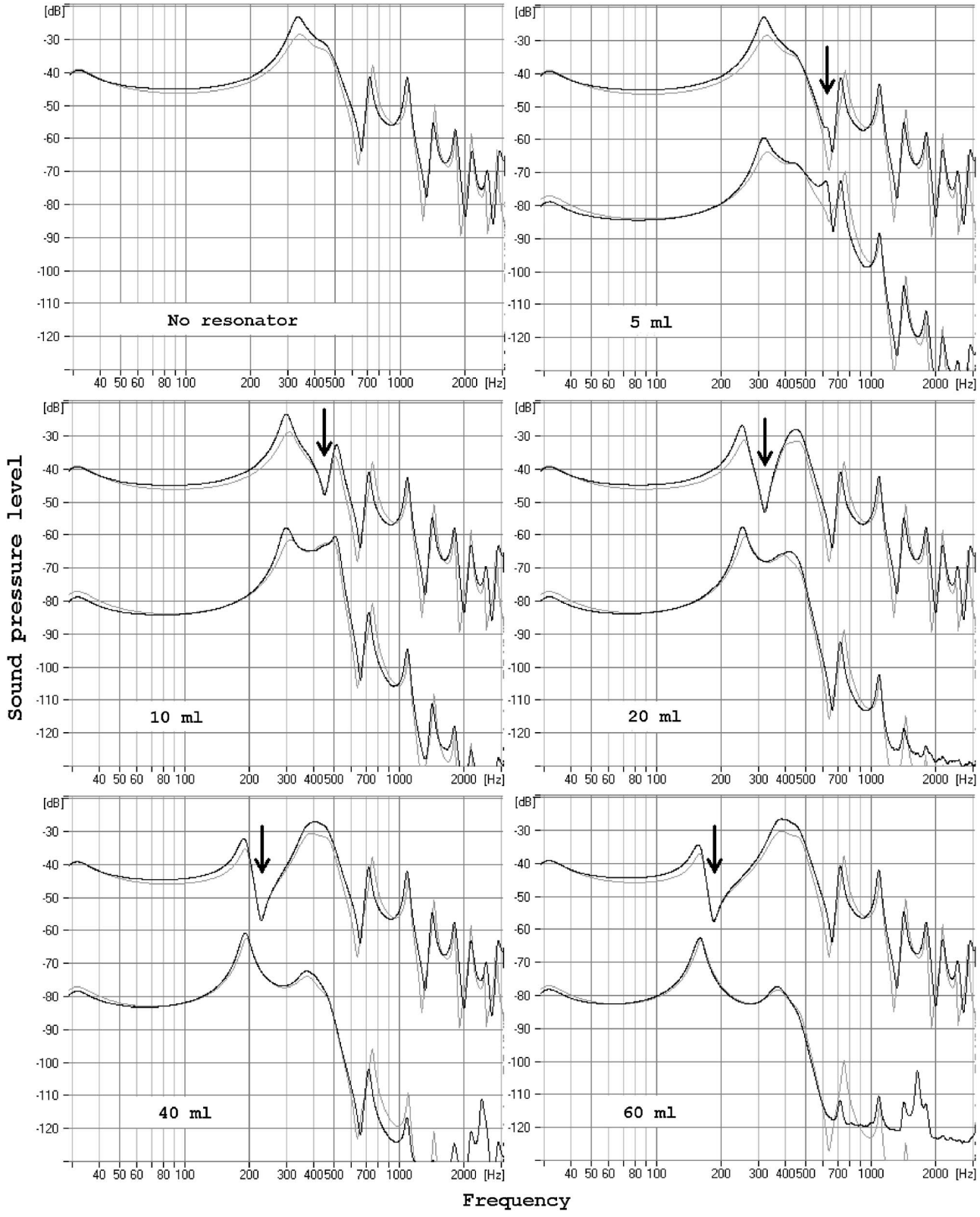


FIG. 8. Simulated and measured data. The top left panel represents the p_t pressure, when the Helmholtz resonator was removed and the hole was plugged. The other five panels represent five different resonator volumes (5, 10, 20, 40, and 60 ml, respectively). In each graph the upper curves represent the driving pressure p_t, and the lower the pressure inside the Helmholtz resonator cavity p_h. The black curves were measured on the physical model and the gray curves were derived from the computer model. The arrows indicate the Helmholtz frequency. The reference for the sound pressure level was arbitrary.

After the iterative manual adjustments, the following parameter values were used: $Q_{\text{tubes}}=30$, $Q_{\text{tubep}}=6$, $Q_h=13$, number of sections: 26. The volumes of the two cavities were: $V_1=2.2$ l, $V_2=0.19$ l. These volumes were also manually adjusted, since the actual volumes were hard to measure exactly due to the space occupied by the loudspeaker driver, and also the isothermal effects of the damping material in compartment V_1.

TABLE II. Volumes and calculated Helmholtz frequencies of the syringe used in Experiment C.

Volume (ml)	Frequency (Hz)
60	185
40	224
20	317
10	448
5	634

The loudspeaker parameters, taken from the data sheet and the impedance curve, were: $V_{as}=10.69$ l, Bl=6.11 N/A, $f_s=54.94$ Hz, $R_e=6.28\ \Omega$, $Q_{ts}=0.35$, lossy Le parameters (Leach, 2002) $K=3.18\times10^{-7}$, $n=0.78$, $S_d=84$ cm^2. The tube length was $L=465$ mm, tube diameter $d=20$ mm.

M_{ah} was determined for one of the syringe volumes by adjusting it such that the dip at F_h in the model's pressure curve coincided with the measured data. This M_{ah} value was then kept when using Eq. (10) for calculating the pressure curves for other syringe volumes.

The displacement of air in the resonator neck is proportional to the pressure inside the cavity. Furthermore, the maximum displacement amplitude in the neck does *not* occur at the Helmholtz resonance frequency f_h. The reason can be seen in the curve representing the driving pressure, which shows a minimum at f_h. Such a driving pressure minimum obviously will counteract an efficient excitation of the Helmholtz resonator. This indicates that the frequency for maximum exchange of NO gas in the paranasal sinuses probably cannot be predicted from dimensions of the sinuses and the channel to the nasal cavities alone. Also the rest of the nasal cavity would have to be taken into account.

IV. DISCUSSION

Our results appear to explain why the NO gas is evacuated from the paranasal sinuses during humming (Weitzberg and Lundberg, 2002); the oscillatory movement of the alternating pressure in the nasal tract at the ostium forces the airplug in the ostium to oscillate, thus increasing the NO gas evacuation. If this explanation is correct, the magnitude of the gas evacuation should be frequency dependent, and such a frequency dependence was also observed in Experiment A. The phenomenon was further examined by means of the physical model used in Experiment B. A syringe modeling a maxillary sinus was filled with NO gas and a dc flow was inserted near the loudspeaker end of the tube, at "A" in Fig. 3. It was shown that the rate of evacuation of NO gas varied depending on the frequency induced by the loudspeaker.

In Experiment C, we focused on the displacement of air in the resonator neck. It can be assumed that this air displacement is closely related to the ventilation of NO gas. As mentioned before, the displacement and flow in the physical model can be indirectly measured in terms of the pressure inside the resonator cavity. Since this pressure appears as a result of the flow through the neck into the cavity, the displacement is proportional to the pressure and the flow to its derivative.

It could be assumed that the flow U_h could be affected by several factors. One would be the relation between the Helmholtz frequency and the exciting frequency. A maximum flow could be expected if these frequencies are identical and if the exciting pressure is constant. On the other hand, the exciting pressure varies due to resonances in the tube. If the Helmholtz resonator neck is placed in a pressure antinode of such a resonance, a maximum could be expected to occur since the exciting pressure would then be maximized. However, the acoustic load from the resonator affects the resonances in the tube. So, if the Helmholtz resonator is placed at position where a pressure node would be expected in the unloaded tube, the node pressure drops. The simulations and measurements in this study indicate that the net effect of this is that two U_h flow maxima occur, one at each side of the original resonance/node frequency.

The manual matching of the computer model to the measured data was highly enlightening. It turned out to be fairly easy to achieve a good agreement in the low-frequency range ($<$500 Hz). However, in order to achieve a good match at higher frequencies, up to 2–3 kHz, the accuracy of the microphone placement and of the dimensions of the box, tube, and resonator turned out to be important. Another experience with the tube model was that it was fairly easy to overload the microphones, since the sound pressure inside the tube and resonator can easily reach very high levels.

Our results support the assumption that the evacuation of the NO gas from the paranasal sinuses is caused by sound pressure produced in the nasal tract during humming. The basic mechanism seems to be that the sound pressure causes the airplug in the ostium to move in and out of the sinus, thereby driving exhalation air into the cavity which then reacts by expelling a portion of the gas it contains.

The magnitude of the effect depends on the frequency. No gas evacuation occurs if the ostium is located at a pressure node. The risk for this case to occur in reality is small, however, since voice F0 is typically in the range 100–200 Hz, i.e., a quarter wavelength of 0.85–0.43 m. The frequency that would fail to ventilate the sinus would be much higher than F0, i.e., an overtone, but the lower spectrum partials would contribute to the ventilation.

Another factor that could be assumed to affect the ventilation is the size of the sinus. On the other hand, a doubling of the Helmholtz volume would cause a resonance frequency shift of no more than 1.4, other factors kept constant. Thus, the size of the sinus would have a modest influence on the ventilation effect. The dimensions of the ostium, on the other hand, would be influential. For example, an obstructed ostium, such as during sinusitis, will obviously prevent all NO-gas ventilation. Indeed, in patients with chronic sinusitis and completely blocked sinus ostia, the humming-induced nasal NO increase is completely absent (Lundberg *et al.*, 2003). The physiological and pathophysiological aspects of humming, nasal NO, and sinus ventilation have been recently discussed in more detail (Weitzberg and Lundberg, 2002; Maniscalco *et al.*, 2003; Lundberg *et al.*, 1995).

Singers often regard the nasal tract as an important part of their resonator system. Evidence of a more or less open velopharyngeal port was recently found for the vowel [a] as

sung by professional opera singers (Birch *et al.*, 2002). Thus in many singers a constant nasal air stream could be measured when the singers sang this vowel. It would be interesting and easy to use the current experimental setup for checking the relevance of the paranasal sinuses to such singers. NO gas in the nasal airflow should be a reliable indicator of excitation of the sinuses.

V. SUMMARY AND CONCLUSIONS

In this study we analyzed the acoustic background of the NO gas evacuation from the paranasal sinuses that has previously been observed to occur during humming. Experiment A revealed that this evacuation varied with frequency and Experiment B elucidated the acoustics underlying this frequency dependence by means of gas evacuation measurements in a tube provided with a radially mounted Helmholtz resonator containing this gas. Experiment C showed that the phenomenon can be explained in terms of a computer model and accurately replicated the complex transfer functions measured in a system similar to that used in experiment B. The computer model could easily be adjusted so as to reflect geometrical data obtained from human subjects. However, it is clear already from the present rather coarse modeling of the nasal resonator system that the increased NO ventilation during nasal murmur can be explained by the acoustic flow into and out of the paranasal sinuses. These sinuses act as Helmholtz resonators with the necks corresponding to the ostia. The sinusoidally alternating acoustic pressure forces the airplug in the Helmholtz resonator neck to vibrate, thus expelling NO from the cavity. This gas is transported to free air by the exhalatory air stream.

ACKNOWLEDGMENTS

This study was supported by The Swedish Heart and Lung Foundation. S.G.'s participation in this investigation was supported by the Swedish Research Council. The investigation was presented at the ASA meeting in New York, May 2004 and at the BNAM meeting of the Baltic-Nordic Acoustical Societies in Helsinki June 2004.

Aust, R., and Drettner, B. (**1974**). "The functional size of the human maxillary ostium in vivo," Acta Oto-Laryngol. **78**, 432–435.

Beranek, L. (**1954**). *Acoustics* (McGraw Hill, New York).

Birch, P., Gümoes, B., Stavad, H., Prytz, S., Björkner, E., and Sundberg, J. (**2002**). "Velum behavior in professional classic operatic singing," J. Voice **16**, 61–71.

Knops, J. L., McCaffrey, T. V., and Kern, E. B. (**1993**). "Inflammatory diseases of the sinuses," Otolaryngol. Clin. North Am. **26**, 509–534.

Leach, M. (**2002**). "Loudspeaker voice-coil inductance losses: Circuit models, and effect on frequency response," J. Audio Eng. Soc. **50**, 442–449.

Lundberg, J., Farkas-Szallasi, T., Weitzberg, E., Rinder, J., Lidholm, J., Anggaard, A., Hokfelt, T., Lundberg, J. M., and Alving, K. (**1995**). "High nitric oxide production in human paranasal sinuses," Nat. Med. **1**, 370–373.

Lundberg, J. O., Maniscalco, M., Sofia, M., Lundblad, L., and Weitzberg, E. (**2003**). "Humming, nitric oxide, and paranasal sinus obstruction," JAMA, J. Am. Med. Assoc. **289**, 302–303.

Maniscalco, M., Weitzberg, E., Sundberg, J., Sofia, M., and Lundberg, J. O. (**2003**). "Assessment of nasal and sinus nitric oxide output using single-breath humming exhalations," Eur. Respir. J. **22**, 323–329.

Small, R. (**1972**). "Closed-box loudspeaker systems. I. Analysis," J. Audio Eng. Soc. **20**, 798–808.

Weitzberg, E., and Lundberg, J. (**2002**) "Humming greatly increases nasal nitric oxide," Am. J. Respir. Crit. Care Med. **166**, 144–145.

Active control of buzz-saw tones: Experimental results from a laboratory-scale, no-flow rig

M. J. Wilkinson and P. F. Joseph[a)]
Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, United Kingdom

(Received 15 March 2005; revised 20 February 2006; accepted 28 February 2006)

This paper details the results of an experiment on a small scale no-flow rig to control a spinning mode with similar characteristics to the modes generated by buzz-saw noise. The spinning mode ($m=-3$, $n=1$) is generated by a mode synthesizer ring which has similar characteristics to that of a low frequency buzz-saw mode. A control system using seven microphones and seven control actuators is used to minimize the sum of the squared pressures at the duct wall using a filtered reference least mean squares (FXLMS) algorithm. It is shown that at certain frequencies the sound power transmitted to the duct termination is reduced by 14.5 dB. The control mechanisms at different frequency ranges are identified and discussed with reference to computer simulations. It is shown that cut-off modes excited by the secondary actuator array have a significant effect on control performance at frequencies close to the cut-on frequency of the ($m=-4$, $n=1$) mode. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188817]

PACS number(s): 43.20.Mv, 43.50.Ki [KAC] Pages: 2618–2627

I. INTRODUCTION

Buzz-saw noise is the noise radiated forward of aircraft engines on take-off when the tip speed of the fan becomes supersonic. Currently, passive liner sections are used to attenuate this subjectively annoying tonal sound field. However, the liners do not have sufficiently good performance to efficiently attenuate the low frequency components of the buzz-saw noise field. It is therefore proposed that active noise control (ANC) techniques be employed either after, or embedded into, the linear to attenuate these low frequency tones. This paper describes a laboratory experiment aimed at assessing the performance of an active control system in controlling a low-order spinning mode in a duct, typical of that found in a low frequency buzz-saw tone.

Characteristics of buzz-saw noise. Buzz-saw tones occur at multiples (engine orders, EO) of the shaft rotational frequency of the fan. The sound field at each EO in the duct comprises a single propagating spinning modes of order $m=m_B=\mathrm{EO}$. The pressure due to buzz-saw modes can therefore be expressed as the sum of the modal contributions summed over all multiples of the shaft rotational frequencies $\omega_l=l\Omega$,

$$p(x,r,\theta,t)=\sum_{l=1}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{n=1}^{\infty}A_{m,n,l}\psi_{m,n}(r,\theta)\times\exp\left(ik_l\left(\frac{\alpha_{mnl}-M_x}{(1-M_x^2)}\right)x\right)e^{-i\omega_l t} \quad (1)$$

where

$$\alpha_{mnl}=\sqrt{1-\frac{k_{r_{m,n}}^2}{k_l^2}(1-M_x^2)}. \quad (2)$$

In the above equation, k_l is the acoustic wave number ω_l/c.

Each of the spinning modes comprises a number of radial modes (of order n). At typical take-off fan tip speeds only the first order radial mode $n=1$ is usually cut-on and hence can radiate efficiently to the far-field. This simple modal structure, and the low frequencies at which it occurs, suggests that active control may be potentially very effective in reducing the buzz-saw sound field.

Previous active control studies on aircraft engines (Fruteau *et al.* 2000; Joseph *et al.* 1999a, 1999b; Kousen and Verdon, 1994; Kraft and Kontos, 1993; Maier *et al.*, 2001; Risi *et al.*, 1996; Smith and Burdisso, 1999; Thomas *et al.*, 1993; Walker, 2001) have focused on the control of rotor-stator interaction tones, which are radiated at subsonic tip-speeds and at the blade passing frequencies of the fan, i.e., EO=B, where B is the number of blades. At these higher frequencies, the sound field is more spatially complex since there is a potentially much larger number of modes which can propagate, whose spinning mode order m is given by the well-known Tyler and Sofrin condition (Tyler and Sofrin, 1962). Each of these spinning modes comprises multiple radial modes. Moreover, at these higher frequencies, there is also a greater possibility for the control actuators to excite large numbers of additional modes. This effect is known as modal spillover. The active control of buzz-saw tones should therefore be much more effective, and simpler to implement, than rotor-stator interaction tones.

The investigation by Walker (Walker, 2001) on the control of rotor-stator interaction tones discusses the degradation of control performance due to cut-off modes excited by the actuator array. This effect is also relevant in the control of buzz-saw modes. However, most of the spillover modes excited by the actuator array on our experimental rig are evanescent. Their effect on control performance is discussed in the final section of this paper.

[a)]Author to whom correspondence should be addressed. Electronic mail: pfj@isvr.soton.ac.uk

 0001-4966/2006/119(5)/2618/10/$22.50

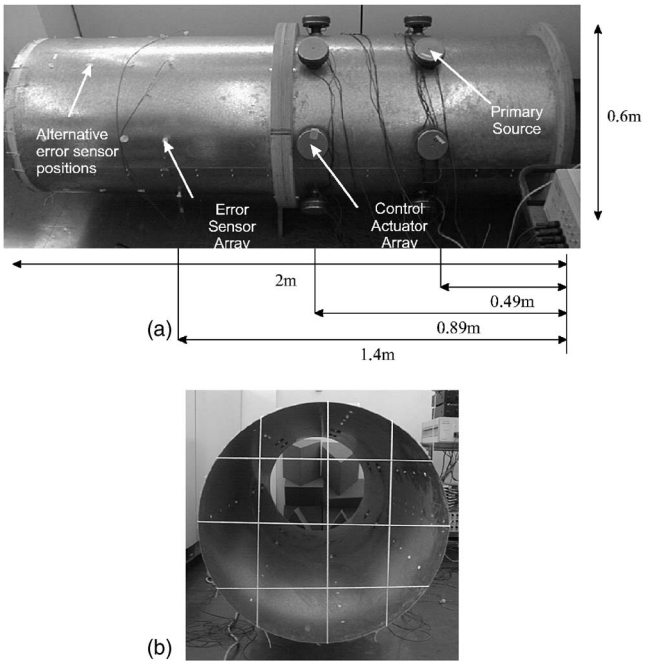


FIG. 1. (a) The duct section and active control system, (b) the measurement grid at the duct "inlet."

This paper describes the results from a laboratory experiment aimed at investigating the performance of a simple active control system in attenuating a single low-order spinning mode at frequencies above and below its cut-off frequency. This spinning mode was excited by a mode synthesizer ring in order to simulate the modal characteristics of a buzz-saw mode at EO=3. The control system consisted of a single ring of seven error microphones and a single ring of seven control actuators, situated in a circular duct made from steel plate of length 2 m and radius 0.3 m. Note that this number of sources and sensors is greater than the minimum criteria suggested by Walker (2001). The duct and the positions of the sources and sensors are shown in Fig. 1(a).

II. THE EXPERIMENTAL RIG

A. Mode synthesizer ring

Each of seven loudspeakers used in the primary source actuator ring shown to the far right of the duct in Fig. 1(a) was driven by a single frequency signal such that adjacent loudspeakers were delayed in time corresponding to a phase shift of $6\pi/7$. This phase variation is designed to excite the $m=-3$ spinning mode, independent of the excitation frequency. The amplitude and phase response of the loudspeakers were measured to be within 3 dB and 20° of each other and hence some degree of spillover into extraneous modes will occur, as occurs in a real engine duct.

In order to assess the performance of the mode synthesizer array in exciting this mode, a spinning mode decomposition was performed on the measured complex pressure amplitudes $p(\theta_l,\omega)$ made at $l=1-7$ error microphones. The procedure for determining these complex amplitudes from the measured time series is described in Wilkinson (2004). This was performed using the Fourier series decomposition,

$$\hat{a}_m(\omega)=\frac{1}{L}\sum_{l=1}^{L}p(\theta_l,\omega)e^{-im\theta_l}\left(\frac{-(L-1)}{2}\leqslant m\leqslant\frac{(L-1)}{2}\right). \tag{3}$$

Thus it is assumed that contributions to the pressure from cut-off modes $|m|\geqslant 4$ can be neglected. We will show in Sec. V that this is not the case after control and that the inability to discriminate the $m=\pm 4$ cut-off modes from propagating modes of lower order degrades control performance.

Control was implemented separately at 14 discrete frequencies between 740 and 950 Hz. This frequency range corresponds to cut-off ratios of $\zeta=0.97$ to 1.24 for the $(-3,1)$ mode. Note that by assuming $M_x=0$ the modal cut-off ratio ζ may be related to that produced by a rotor with blade tip Mach number M_t by

$$\zeta=\frac{mM_t}{k_{r_{m,n}}a}, \tag{4}$$

where $k_{r_{m,n}}$ is the radial component of the acoustic wave number and a is the duct radius. Equation (4) shows that the above-given range of ζ corresponds to a range of M_t of 1.36–1.74, which is significantly higher than the take off fan speed of full-scale rotors. This mode is therefore usually cut-off in a real engine duct. Nevertheless, many of the control principles established here also apply at higher engine orders.

Figure 2 shows the result of the spinning mode decomposition at each frequency.

The spinning mode amplitude of the $(-3,1)$ mode is observed to be approximately 10–20 dB greater than those of the extraneous modes, excited due to loudspeaker mismatch. Note that in this measurement, it is assumed that the contribution from the cut-off $m=\pm 4$ modes can be neglected due to the large separation distance between primary ring and error sensor ring.

B. The FXLMS controller

The filtered-x least mean squares (LMS) algorithm adopted here is a modification of the LMS algorithm and has been successfully used in earlier active control system applications in aircraft engines (Maier *et al.*, 2001; Thomas *et al.*, 1993). The approach followed in the experiment is a modified version of the multichannel control approach used by Elliott *et al.* (1987). In the experiment the excitation frequency is known. A reference signal is generated by the controller and not from a reference microphone in the duct. At each sample time, with the modified FXLMS algorithm the following operations are performed.

(1) The reference signal for the nth sample is generated using

$$x(n)=\cos(\omega nT).$$

(2) The matrix of filtered reference signals $\mathbf{R}(n)$ is constructed using the amplitude and phase of the single frequency transfer function matrix $\mathbf{Z}$, with elements of the form

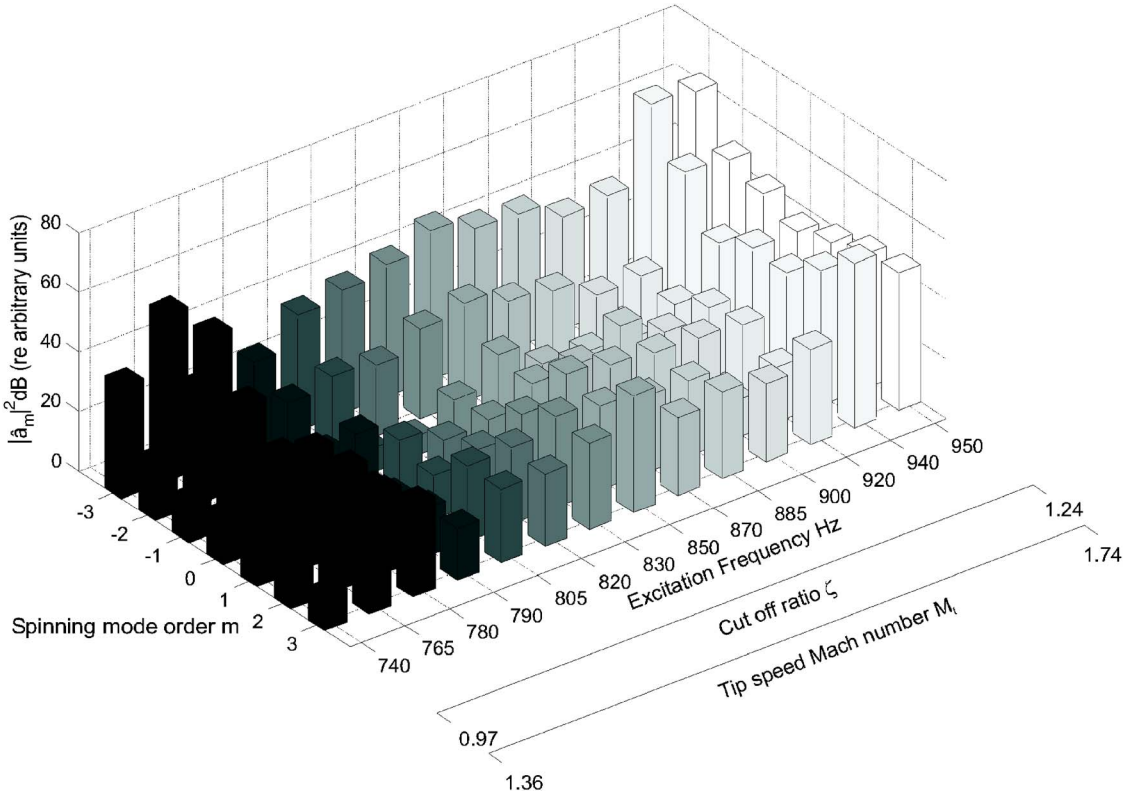


FIG. 2. Measured spinning mode decomposition at the wall at each excitation frequency.

$Z_{ml}=A_{ml}e^{i\phi_{ml}}$. The transfer functions are measured offline, prior to control. The internally generated reference signals are of the form,

$$\hat{r}_{ml}(n) = A_{ml}\cos(\omega nT + \phi_{ml}).$$

(3) The actuator u_m signals are generated by filtering the reference signal $\mathbf{x}=[x(n)x(n-1)]^T$ using the current values of the filter coefficients $\mathbf{w}(n)$,

$$u_m(n) = \mathbf{w}^T\mathbf{x}.$$

(4) Finally, the error signal $e(n)$ is obtained from the voltage signal measured at the microphone array. This is acquired for the nth sample and used together with the filtered reference signals to update the filter coefficients, according to

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \alpha\mathbf{R}^T e(n)$$

where $\mathbf{R} = [\mathbf{R}(n)|\mathbf{R}(n-1)]$.

III. COMPUTER MODEL

The measured control performance from this experiment will be compared against the performance predicted by the computer model detailed in Wilkinson (2004). The model includes the following assumptions:

(1) The engine inlet is a finite-length, axisymmetric circular duct with rigid walls containing an axial, uniform mean flow.
(2) The modal amplitudes of the primary field in the absence of reflections in the duct (the sound field generated by the buzz-saw noise) are known at the fan plane x_f.
(3) The secondary sources are mounted flush with the duct wall and behave as point monopoles.
(4) At tip speeds close to $M_t=1$, the buzz-saw tone associated with each engine order EO generates a single propagating primary mode $m=m_B=$EO (McAlpine and Fisher, 2001), which propagates linearly along the duct.
(5) Blockage of the sound field by the rotor and stator can be neglected.

IV. EXPERIMENTAL RESULTS

The active control system was used to minimize the sum of the squared pressures at the error sensors at 14 excitation frequencies in the frequency range $f_{-3,1}<f<f_{-4,1}$ where $f_{-3,1}$ and $f_{-4,1}$ are the cut-off frequencies of the $(-3,1)$ and $(-4,1)$ modes, respectively. Control was implemented at one frequency at a time. The reduction in the sum of the squared pressures determined using the signals recorded at the microphone array, before and after control, is plotted in Fig. 3. In this first experiment the error sensors are positioned closest to the secondary source array at a separation distance of 0.51 m. The effect on control performance of increasing the separation distance will be investigated in later sections.

The pressure reduction at the error sensors is shown to vary from 14 dB in the worst case at 900 Hz to 39 dB at

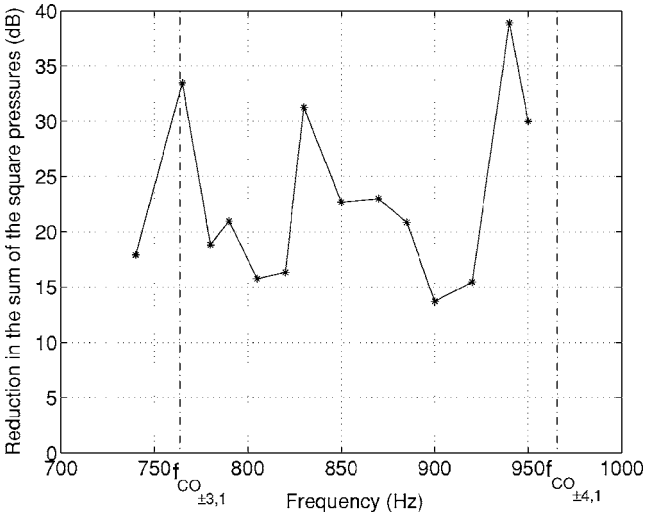


FIG. 3. Sound pressure level reduction with varying excitation frequency.

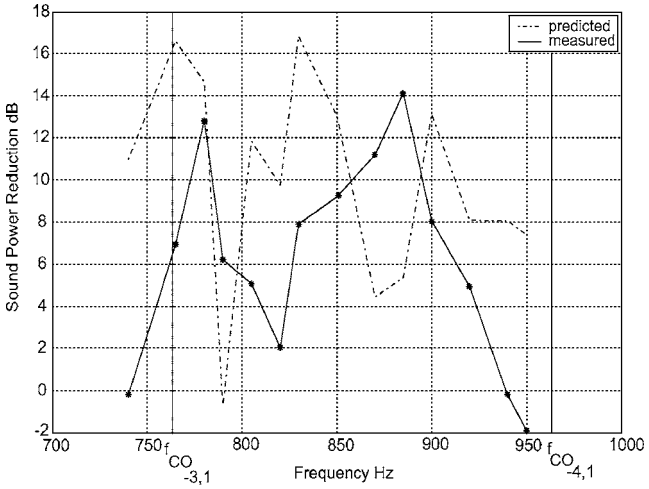


FIG. 4. Sound power level reduction with varying excitation frequency.

940 Hz. This variation in sound pressure reduction is likely to be due to inaccuracies, at each excitation frequency, in the transfer function measurements.

The corresponding reduction in the sound power level transmitted downstream of the error sensors after control is plotted in Fig. 4 (solid curve).

The sound power reduction is determined by integrating the axial intensity component over the duct termination using an intensity probe before and after control. This procedure was achieved by dividing the duct face into 16 sections, as shown in Fig. 1(b). The area-averaged intensity $\langle \bar{I}_n \rangle_i$ flowing normally across each section i was then measured and summed to give the total sound power.

The sound power reduction is observed to be highly frequency dependent, but unrelated to the variation in sound pressure reductions plotted in Fig. 3. The levels of sound power reduction are shown to be typically 10–20 dB less than the sound pressure reduction, suggesting that the sound power reduction cannot be directly inferred by the sound pressure reduction at the error sensors. The control mechanism, and hence the relationship between sound pressure and sound power, will be discussed in Sec. V.

Also plotted in Fig. 4 (dashed curve) is the sound power reduction predicted by the computer model. The level of agreement is generally poor. At most excitation frequencies the model predictions a sound power reduction within 5–10 dB of the measured value. The discrepancy between measured and predicted control performance is mainly due to the assumption in the model that the pressure is driven perfectly to zero. In the experiment, however, the pressure level is only reduced by between 20 and 40 dB. The precise phase relationship between modes after control, predicted by the computer model, is therefore not realized in practice, leading to the large discrepancies observed in Fig. 4. Nevertheless, we shall see in Fig. 6 that predictions obtained from the model of the detailed variation in induct pressure are, in general, in excellent qualitative agreement with measurement. Despite the poor agreement between predicted and measured sound power reduction, therefore, the theoretical model will be shown to provide useful insight into the con-

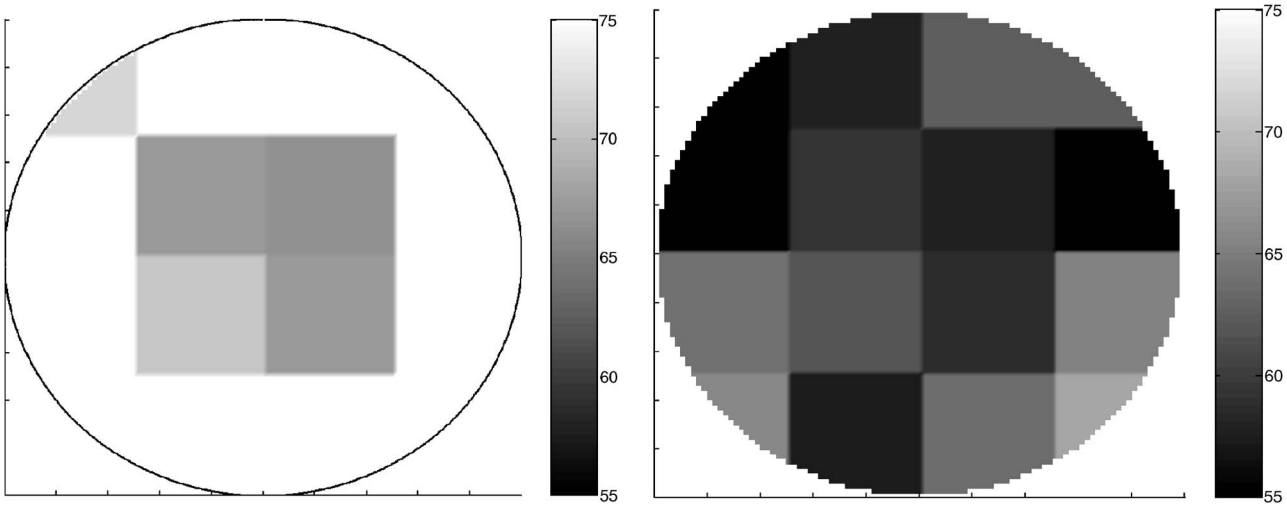


FIG. 5. Measured sound intensity distribution before control (a) and after control (b) at 885 Hz.

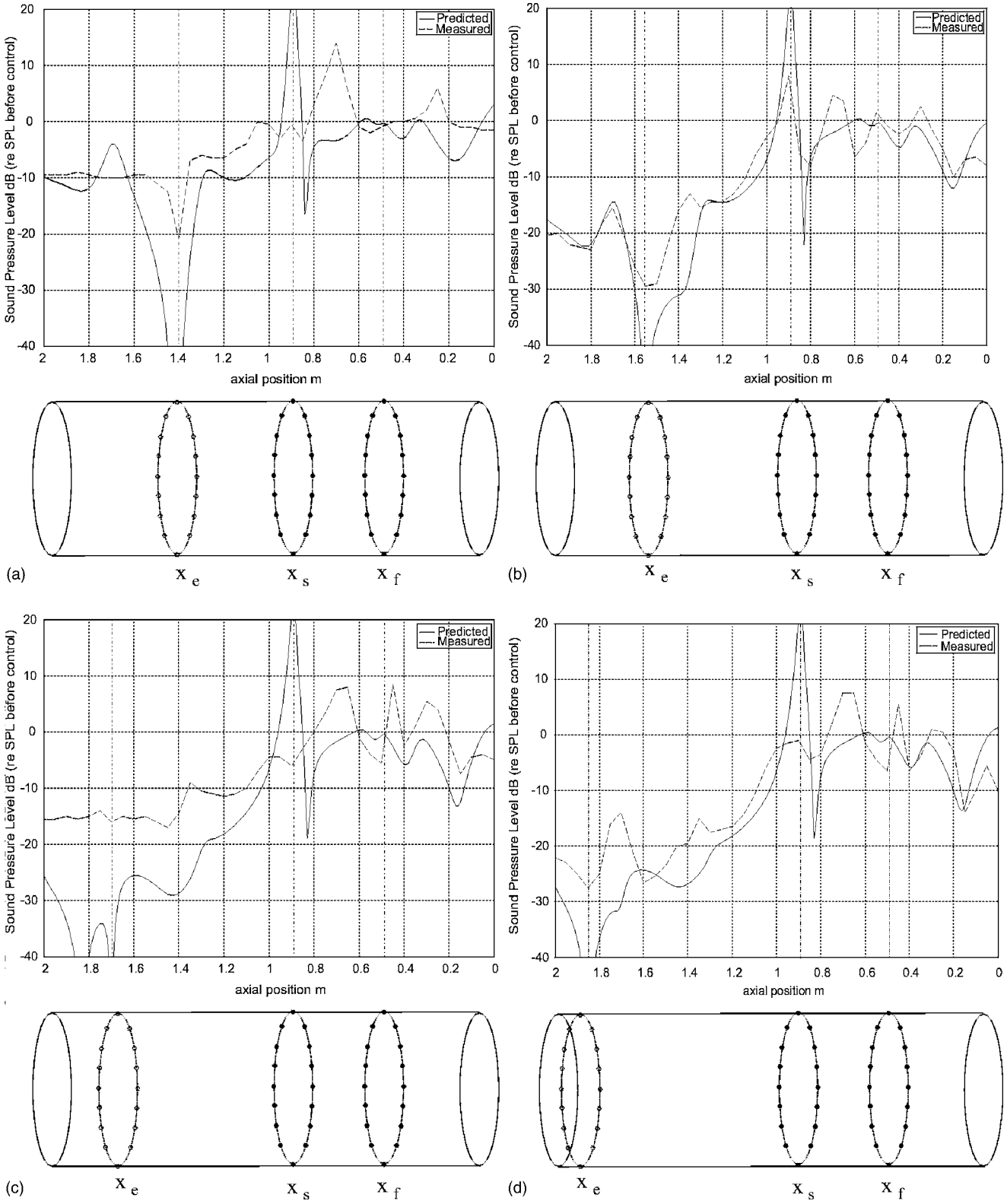


FIG. 6. Comparison of measured and predicted sound pressure reduction with varying axial position 885 Hz: $x_e = 1.4m$ (a), $1.55m$ (b), $1.7m$ (c), and 1.85 (d).

trol mechanism. A further reason for the discrepancy observed in Fig. 4 is that the hard-walled assumption made in the model may not be valid in the case of a steel plate duct. Vibration of the duct wall causes some sound to be reradiated into the duct or sound to "breakout" through the duct walls.

The acoustic intensity distribution across the duct "inlet" at a representative excitation frequency of 885 Hz, before and after control, is plotted in Figs. 5(a) and 5(b). Before control, the acoustic energy is concentrated close to the duct walls. This provides some evidence for the dominance in the sound field of the first radial mode order $n=1$. Figure 5(b) confirms that after control the intensity at the duct wall has reduced by approximately 15 dB, commensurate with the overall power reduction of 14 dB at this frequency seen in Fig. 4. Furthermore, the radial variation of sound intensity is

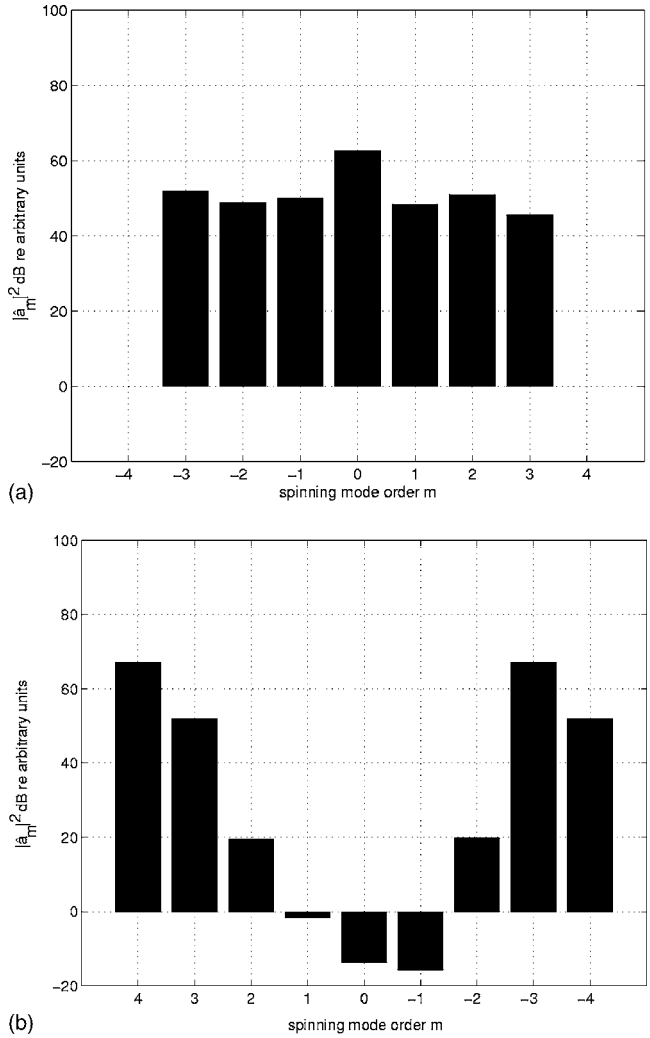


FIG. 7. (a) Measured and (b) predicted spinning mode amplitudes at the error sensors after control at 885 Hz.

now significantly smaller compared with that before control, suggesting that the single radial mode no longer dominates the sound field in the duct after control.

Further insight into the control process is obtained from measurements of the sound pressure level reduction versus axial distance along the duct at the duct wall. Figures 6(a)–6(d) show the sound pressure reduction at 885 Hz measured in 5 cm increments along the duct for the error sensor positions of $x_e = 1.4m$, $1.55m$, $1.7m$ and $1.85m$, respectively (solid curves). Also plotted are the predicted sound pressure reductions calculated using the theoretical model (dashed curves).

Good qualitative agreement is generally observed between the measured and predicted axial variation in sound pressure level reduction. An interesting feature of these figures is that the pressure drops to a minimum at the error sensor array and then rises to a finite level at the duct termination, before being radiated to the far-field. The control mechanism, from a modal point of view, therefore, does not simply involve driving the mode amplitudes to zero, which would give constant sound pressure level reductions downstream of the error sensor plane. This behavior indicates the existence of destructive interference between different modes. This control mechanism is explored in greater detail in the following section.

V. GENERAL CONTROL MECHANISM

The control system seeks to drive the pressure at L error sensors to zero. The pressure at the lth error sensor can be expressed in terms of the spinning mode amplitudes $\hat{a}_m$ at the duct wall in the form,

$$p(\theta_l, \omega) = \sum_{m=-\infty}^{\infty} \hat{a}_m(\omega) e^{im\theta_l}, \tag{5}$$

where $\hat{a}_m$ comprises a number of radial modes,

$$\hat{a}_m(\omega) = \sum_{n=1}^{\infty} a_{m,n} \Psi_{m,n}(a), \tag{6}$$

where $a_{m,n}$ is the amplitude of the mode (m,n), and $\Psi_{m,n}(a)$ is the radial mode shape function evaluated at the duct wall.

Provided there is a sufficient number of sensors, the spinning mode amplitudes $\hat{a}_m$ can be determined by performing a discrete spatial Fourier transform of the pressure signals at the error sensors using Eq. (3). Thus if the pressure is driven to zero at all seven error sensors, i.e.,

$$p(\theta_l, \omega) = 0 \quad (l = 1, \ldots, 7) \tag{7}$$

Eq. (3) suggests that the spinning mode amplitudes are independently driven to zero at the error sensor array, i.e.,

$$\hat{a}_m(\omega) = 0 \quad (-3 \leqslant m \leqslant 3). \tag{8}$$

Equation (8) assumes that the contribution from cut-off $m = \pm 4$ modes can be neglected. Subject to this assumption being valid, Eqs. (7) and (8) indicate that driving the pressure at the error sensors to zero, and driving the spinning mode amplitudes to zero, are identical control strategies.

The computer model was also used to predict the reduction in mode amplitudes following control, and hence the sound power reduction. The input to the model was the mode amplitude spectrum for $-3 \leqslant m \leqslant 3$ measured at the error sensors before control, as plotted in Fig. 2. For simplicity the amplitude of the two axi-symmetric $m = 0$ modes are assumed equal. A comparison of the measured and predicted spinning mode amplitudes after control is shown in Figs. 7(a) and 7(b), respectively.

Agreement between measurement and prediction is poor. Contrary to Eq. (8), Figs. 7(a) and 7(b) show that neither the predicted nor measured mode amplitudes are driven to zero. The main discrepancy between them arises for the $m = \pm 4$ modes. In the measurement it is assumed that these can be neglected in order for the modal decomposition of Eq. (3) to be valid. However, the amplitude of the $m = +4$ mode is predicted to be exactly equal to the $m = -3$ mode, and the amplitude of the $m = -4$ mode equal to that of $m = +3$ mode. The source of the discrepancy for the $m = 0$ and $m = \pm 1$ modes in Fig. 7 arises from imperfect cancellation between the two radial mode amplitudes $a_{m,1}$ and $a_{m,2}$. We shall show in the next section that the modal amplitudes $\hat{a}_0$ and $\hat{a}_{\pm 1}$ are driven to zero by exciting their component radial mode amplitudes

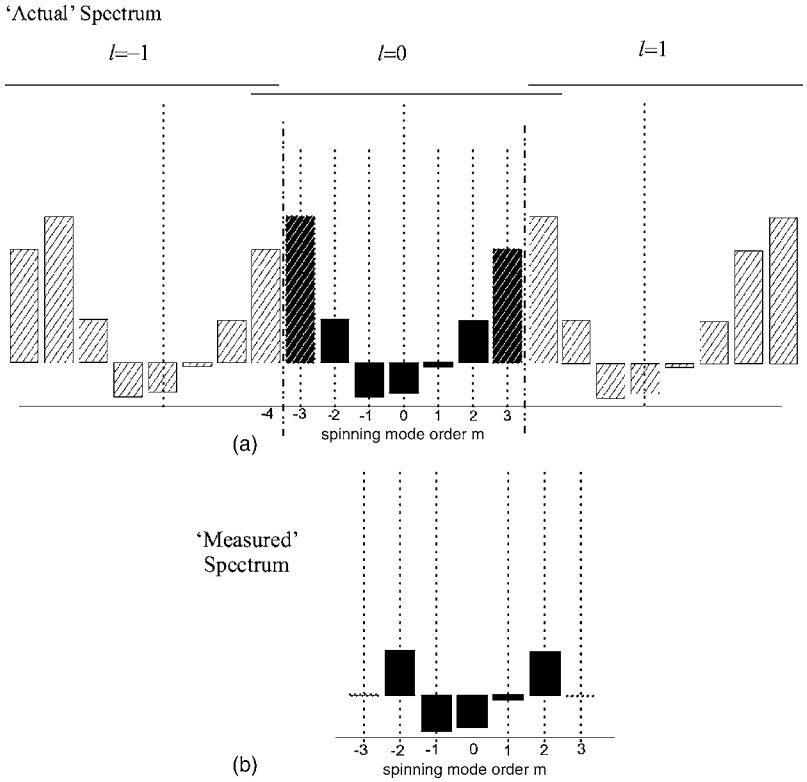


FIG. 8. The "measured," aliased spectrum of spinning mode amplitudes after control at 885 Hz.

to perfectly cancel. Since the single source and sensor ring cannot control and detect individual radial modes, they are excited to high amplitude and become the dominant modes (see Figs. 10 and 12). Imperfect cancellation of these modes due to imperfect pressure cancellation at the error sensors leads to large spinning mode amplitudes, as observed in the measurements of Fig. 7(a).

Aliasing of higher order evanescent modes. Due to the close proximity of the secondary sources to the error sensor in the duct rig, evanescent modes, which are close to cutting on at the frequencies of interest here, will make a significant contribution to the pressure at the error sensor array. The seven sources and sensors used in the control system were chosen to satisfy the Nyquist criterion for the maximum propagating mode order $m=\pm 3$. Pressure contributions from the $m=\pm 4$ cut-off modes will therefore be aliased and hence will contribute to the modal amplitudes observed by the control system in the range $-3 \leqslant m \leqslant 3$.

The use of seven error microphones in the control system for mode detection assumes periodicity in the spinning mode spectrum such that

$$\hat{a}_m = \hat{a}_{m+7l} \quad (l = -\infty, \ldots, 0, \ldots, \infty). \tag{9}$$

If the actual mode spectrum after control is of the form shown in Fig. 7(b) comprising modes in the range $-4 \leqslant m \leqslant 4$, then the mode spectrum that appears to the control system over the range $-3 \leqslant m \leqslant 3$ will be of the form shown in Fig. 8(a).

The periodicity property indicated in Fig. 8(a) shows that the spinning mode amplitudes $\hat{a}_{\pm 3}$ appearing at the error sensor array actually include an $m=\pm 4$ cut-off mode. From Eq. (9) the measured $m=+3$ and $m=-3$ modes comprise the following:

$$\begin{aligned} \underbrace{\hat{a}_3}_{measured} &= \underbrace{\hat{a}_3 + \hat{a}_{-4}}_{actual} \quad (l=-1), \\ \underbrace{\hat{a}_{-3}}_{measured} &= \underbrace{\hat{a}_{-3} + \hat{a}_4}_{actual} \quad (l=-1), \end{aligned} \tag{10}$$

Equation (10) suggests that the "measured" spinning mode amplitudes, of order $m=\pm 3$, have been driven to zero by arranging for them to cancel with a corresponding $m = \mp 4$ cut-off mode i.e.,

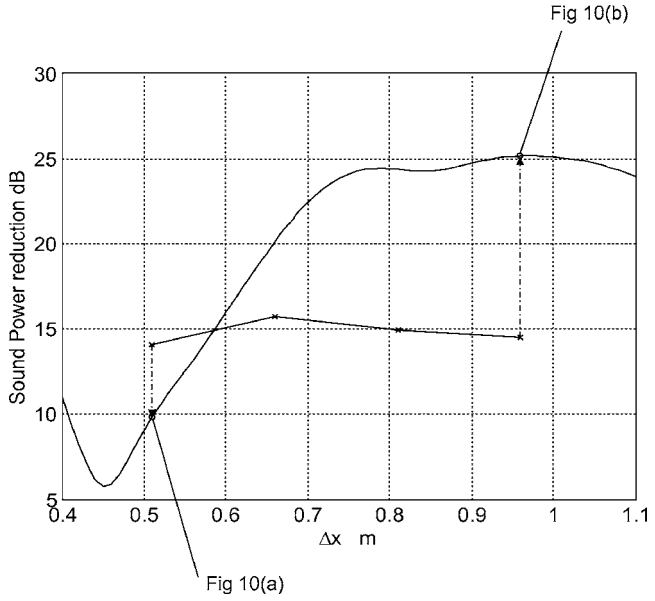


FIG. 9. Comparison between measured and predicted variation of sound power reduction with source/sensor separation distance at 885 Hz. Predicted mode amplitudes at $\Delta x=0.51$ and $0.96m$ are plotted in Fig. 10.

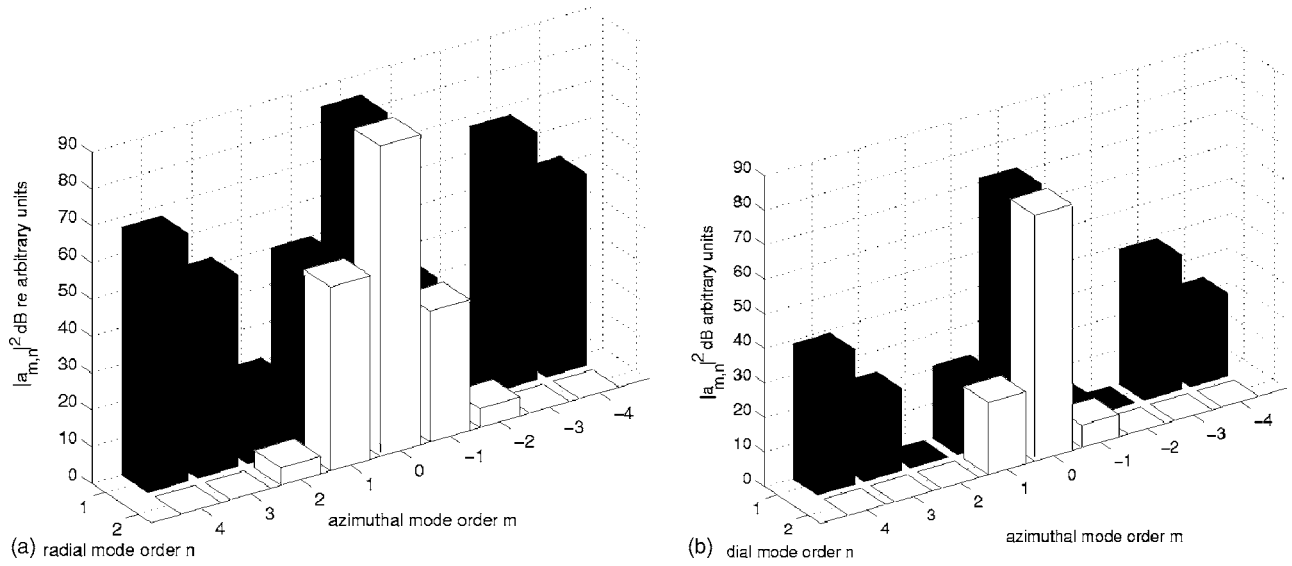


FIG. 10. Mode amplitudes after control at the error sensor array at 885 Hz, (a) $\Delta x = 0.56m$ and (b) $0.91m$.

$$\hat{a}_3 + \hat{a}_{-4} = 0, \qquad \hat{a}_{-3} + \hat{a}_4 = 0. \tag{11}$$

Precisely this behavior is observed in Figs. 7(a) and 7(b), where the amplitudes of $\hat{a}_3$ and $\hat{a}_{-4}$ are exactly equal.

The modal interpretation of the control mechanism involved in the cancellation of the pressure to zero at the error sensors in the experimental duct has been shown to be surprisingly subtle. It may be summarized as follows:

(1) The pressure is driven to zero by arranging for $\hat{a}_{\pm 4} = -\hat{a}_{\mp 3}$ at the error sensor plane. This is because the $m = \pm 4$ modes cannot be controlled independently by the seven sources and sensors of the control system.
(2) The amplitudes of the axi-symmetric $m = 0$ radial modes and the $m = \pm 1$ and $m = \pm 2$ are arranged to cancel, $\hat{a}_{0,1} = -\hat{a}_{0,2}$, since the two radial modes associated with $m = 0$ cannot be independently controlled by the single ring of sources and sensors, even though $\hat{a}_{\pm 1,2}$ are cut-off.

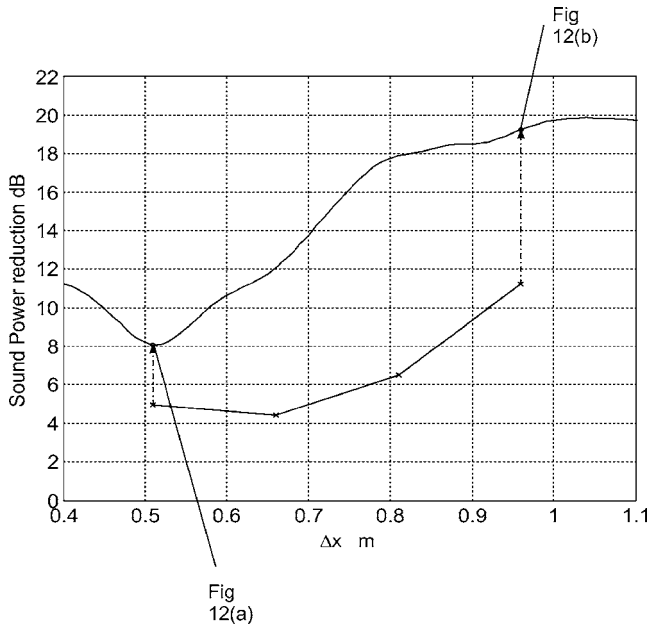


FIG. 11. Variation of sound power reduction with source/sensor separation distance at 920 Hz.

We now consider in detail the control mechanism versus source-sensor separation distance at 885 Hz and at frequencies greater than 900 Hz, since the control mechanism is substantially different in these two cases.

VI. CONTROL PERFORMANCE VERSUS SOURCE-SENSOR SEPARATION DISTANCE

One of the few parameters that may be varied in the experimental rig is the axial separation distance Δx between the sources and sensors. We now investigate control performance for different Δx at two different frequencies. At the lower frequency of 885 Hz, the excitation frequency is sufficiently low for the control system not to be affected by the presence of higher order cut-off modes. At 920 Hz, the cut-off $m = \pm 4$ modes make a significant pressure contribution at the error sensors and therefore begin to degrade control performance. The decay in SPL between sources and sensors of the (4, 1) evanescent mode can be determined from

$$\Delta \mathrm{SPL} = -8.7 \frac{2\pi}{c} \Delta x \sqrt{f_{4,1}^2 - f^2} \quad (f < f_{4,1}). \tag{12}$$

At 920 Hz, with a separation distance of $\Delta x = 0.51m$ the $m = \pm 4$ mode will decay by 24 dB before reaching the error sensor array, increasing to 32 dB at the lower frequency of 885 Hz.

A. Results at 885 Hz

The measured sound power reduction for the four separation distances Δx is plotted in Fig. 9 (asterisks). Also shown are the sound power reductions predicted using the theoretical model (continuous line).

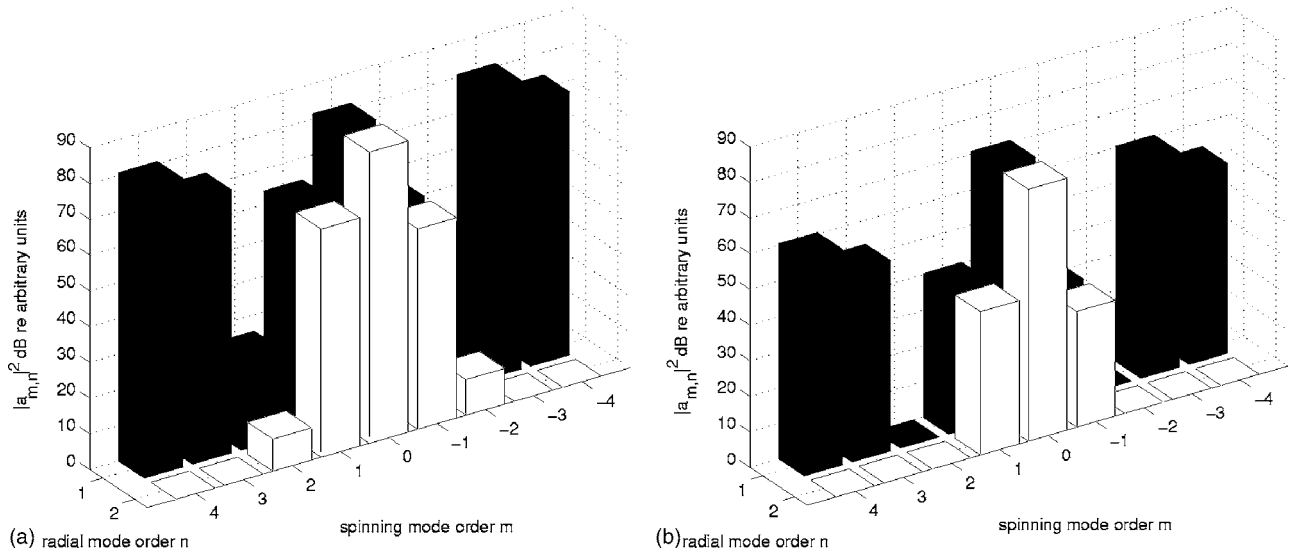


FIG. 12. Predicted mode amplitudes after control at the error sensor array at 920 Hz: $\Delta x=0.56m$ (a) and $0.91m$ (b).

Once again, agreement between the measured and predicted sound power reductions is generally poor. This can be explained by Figs. 7(a) and 7(b) showing the measured and predicted spinning mode amplitudes after control. Predicted individual mode amplitudes after control for $\Delta x=0.51m$ and $0.96m$ are plotted in Fig. 10. The theoretical prediction in Fig. 10(a) shows that the axi-symmetric $m=0$ radial modes are excited in anti-phase, to give $\hat{a}_m=0$. However, the measured pressure reduction at the error sensors is only 30 dB. Cancellation between the radial modes is therefore imperfect.

Figures 10(a) and 10(b) show that with the error sensors positioned furthest from the source array, $\Delta x=0.96m$, the axi-symmetric modes $m=0$ modes are excited to a lower level than at the smaller separation distance and hence the sound power reduction is predicted to be greater. A simple analytic model predicting this behavior is presented by the first author in Wilkinson (2004).

B. Results at $f \geqslant 900$ Hz

In this section we investigate the mechanism of control at frequencies greater than 900 Hz. The important difference here compared with the results presented in Sec. VI A is that the $m=\pm 4$ modes are significantly less cut-off, and hence now make a significant contribution to the pressure at the error sensors.

Figure 11 shows the measured and predicted variation in sound power reduction with varying Δx at 920 Hz.

Once again the measured sound power reduction is overestimated by about 10 dB. The overall trend in both measurement and prediction, however, is similar. Both indicate a sound power reduction that increases with increasing separation distance. This behavior confirms that the $m=\pm 4$ cut-off modes dominate the control mechanism at this higher frequency. This is verified in a prediction of the mode amplitudes at the error sensors at separation distances Δx of $0.51m$ and $0.96m$, as shown in Figs. 12(a) and 12(b).

With the source-sensor separation distance equal to $\Delta x=0.51m$, Fig. 12(a) shows that the amplitudes of the $m=\pm 4$ modes are slightly greater than that of the two $m=0$ radial modes. As explained in Sec. V, the $m=\pm 3$ modes will therefore be excited in order to cancel with this cut-off mode, and will therefore dominate the sound power after control. Global control performance is therefore diminished by the presence of the $m=\pm 4$ modes. As Δx is increased, the level of the cut-off modes reduces. With the error sensors positioned furthest from the source array, $\Delta x=0.91m$, the cut-off mode has decayed to a level that is lower than that of the axi-symmetric modes. The situation in this case is therefore similar to that described in Sec. VI A.

VII. CONCLUSIONS

The design and implementation of an active control system to control a low order spinning mode in a no-flow duct has been described. An $m=-3$ mode was generated using a mode synthesizer ring and this spinning mode was controlled using seven sources and seven error sensors. Sound power reductions of up to 14.5 dB were measured at the duct face when minimizing the sum of the squared pressures at the error sensors using a single frequency FXLMS algorithm. The control mechanism was examined using a theoretical model developed in previous work by the authors. It was seen that the level of the axi-symmetric $m=0$ modes dominate the sound power after control, and hence sound power reduction was found to be significantly less than the pressure reduction at the error sensors. At higher frequencies it was shown that aliasing of the $m=\pm 4$ cut-off modes also degrades control performance at the frequencies at which these cut-off modes contribute significantly to the pressure at the error sensors.

ACKNOWLEDGMENT

The research in this paper was funded by the EU SILENCE(R) project.

Elliott, S. J., Stothers, I. M., and Nelson, P. A. (**1987**). "A multiple error LMS algorithm and its application to the active control of sound and

vibration," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-35**, 1423–1434.
Fruteau, E., Joseph, P. F., and Nelson, P. A. (2000). "Final report on the reduction of aircraft noise by nacelle treatment and active control (RANNTAC)," Technical Report 00/36, ISVR, December 2000.
Joseph, P. F., Nelson, P. A., and Fisher, M. J. (**1999a**). "Active control of fan tones radiated from turbofan engines I External error sensors," J. Acoust. Soc. Am. **106**, 766–778.
Joseph, P. F., Nelson, P. A., and Fisher, M. J. (**1999b**). "Active control of fan tones radiated from turbofan engines II In-duct error sensors," AIAA J. **106**, 779–786.
Kousen, K. A., and Verdon, J. M. (**1994**). "Active control of wake blade-row interaction noise," AIAA J. **32**, 1953–1960.
Kraft, R. E. and Kontos, K. B. (**1993**). "Theoretical implications of active noise control for turbofan engines," in Proceedings of the 15th AIAA Aeroacoustics Conference, Long Beach.
Maier, R., Zillman, J., Roure, A., Winninger, M., Enghardt, L., Tapken, U., Antoine, H., and Bouty, E. (**2001**). "Active control of fan tone noise from aircraft engines," in Proceedings AIAA/CEAS Aeroacoustics Conference, Maastricht.
McAlpine, A., and Fisher, M. J. (**2001**). "On the prediction of 'buzz-saw' noise in aero-engine inlet ducts," J. Acoust. Soc. Am. **248**, 123–149.
Risi, J. D., Burdisso, R. A., and Fuller, C. R. (**1996**). "Analytical investigation of active control of radiated inlet fan noise," J. Acoust. Soc. Am. **99**, 408–416.
Smith, J. P., and Burdisso, R. A. (**1999**). "Active control of inlet noise from a turbofan engine using inlet wavenumber sensors," in Proceedings of the Fifth AIAA/CEAS Aeroacoustics Conference, Seattle.
Thomas, R. H., Burdisso, R. A., Fuller, C. R., and O'Brien, W. F. (**1993**). "Preliminary experiments on active control of fan noise from a turbofan engine," J. Sound Vib. **161**, 532–537.
Tyler, J. M., and Sofrin, T. G. (**1962**). "Axial flow compressor noise studies," SAE Trans. **70**, 309–332.
Walker, B. (**2001**). "Sensitivity issues in active control of circular duct modes using axially-spaced actuator arrays," Noise Control Eng. J. **49**, 6–14.
Wilkinson, M. J. (2004). "Active control of low frequency buzz-saw tones," Ph.D. thesis, University of Southampton.
Wilkinson, M. J., and Joseph, P. F. (**2004**). "Comparison of active control strategies for the active control of buzz-saw tones," in Proceedings of the Tenth AIAA/CEAS Aeroacoustics Conference, Manchester.

Broadband sound reflection by plates covering side-branch cavities in a duct

Lixi Huang[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong

(Received 9 December 2005; revised 3 February 2006; accepted 18 February 2006)

When a segment of a rigid duct is replaced by a plate backed by a hard-walled cavity, grazing incident sound waves induce plate bending, hence sound reflection. The mechanism is similar to the drumlike silencer with tensioned membranes [L. Huang, J. Acoust. Soc. Am. **112**, 2014–2025 (2002)]. However, the logarithmic bandwidth over which the reflection occurs is much wider than that of a drumlike silencer of the same cavity geometry, the typical difference being nearly one octave band. The difference in the silencing performance is explained in terms of the intermodal acoustic interference between the odd and even *in vacuo* vibration modes. For a given cavity volume, the widest stopband for noise in air is obtained by using long plates with two free lateral edges parallel with the duct axis. The optimal material should be stiff and light, and the critical parameter is found to be the ratio of the Young's modulus over the cube of density. Typically, this ratio is 250 times higher than those of common metallic materials like aluminum alloys, but it is within the reach of existing ultralight foam materials or composite beams with a light core. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2186431]

PACS number(s): 43.20.Tb, 43.20.Ks [KA] Pages: 2628–2638

I. INTRODUCTION

Duct noise control finds many applications such as ventilation systems in built-up environments. Medium and high-frequency noise is easily tackled by duct lining using porous materials, for which the technology is rather mature. Low-frequency noise remains a technical challenge. Active control could, in principle, deal with such problem with ease, but the actual applications of the technology have been few due to cost, reliability, and other practical concerns. The present author has been developing an alternative for the low-frequency duct noise control through the use of flexible walls (Huang, 1999) leading to a prototype device of drumlike silencer (Huang, 2002) in which a segment of tensioned membrane lines part of the duct wall, and each membrane is backed by a hard-walled cavity. Incident sound waves from upstream induce the membrane to vibrate and the vibration radiates sound in both directions of the duct. The upstream radiation forms the wave reflection while the downstream radiation combines with the incident wave and destructive acoustic interference ensues. In fact, the acoustics of such device is very similar to a successful scheme of active duct noise control using a single side-branch loudspeaker except that the secondary source in this case is the tensioned membrane, and there is absolutely no chance that the destructive acoustic interference in the downstream region could turn out to be constructive. This is so because the incident wave is the only energy source and the acoustic energy conservation guarantees that there is transmission loss of noise across such flexible segments. When compared with duct lining and splitter-type silencers, the drumlike silencer carries the extra merit of creating zero back pressure and using no porous material. The prototype device has been tested successfully for the conditions of without (Choy and Huang, 2002) and with flow (Choy and Huang, 2005).

It has been shown that the drumlike silencer achieves its best silencing performance under an optimal tensile force (Huang, 2004). When the membrane is too loose, reflection waves from different parts of the membrane simply cancel themselves out. When the membrane is too tight, it approaches the condition of a hard wall. The mechanism of the superior sound reflection capability achieved for the optimal configuration lies in the delicate acoustic interference between different parts of the tensioned membranes. If the membrane vibration is decomposed into *in vacuo* modes of velocity $\sin(j\pi x/L)$, where the membrane occupies $x \propto [0,L]$, the broadband performance is found to derive mainly from the first two modes, $j=1,2$. The first mode is very effective in reflecting low-frequency sound but is hard to excite due to the cavity air stiffness. The second mode is relatively ineffective in reflecting low-frequency sound but is rather easy to excite. The two modes are found to supplement each other in achieving a broadband silencing performance.

In this study, the tensioned membrane is replaced by a plate with its natural bending moment as the sole structural restoring force. The parametric relationship between the first two *in vacuo* modes is changed, and a corresponding change in the intermodal acoustic interference is also expected to take place. For example, the second membrane (string) mode has its eigenfrequency twice as much as that of the first mode. But, this is changed to four times when it comes to the plate. It is shown in later sections that the change in the intermodal relationship is beneficial for sound reflection provided that the parameters of the plate are optimized.

The use of a plate and, for this matter, the tensioned membrane in noise abatement may not be new, but its use as

a)Electronic mail: mmlhuang@polyu.edu.hk

 0001-4966/2006/119(5)/2628/11/$22.50

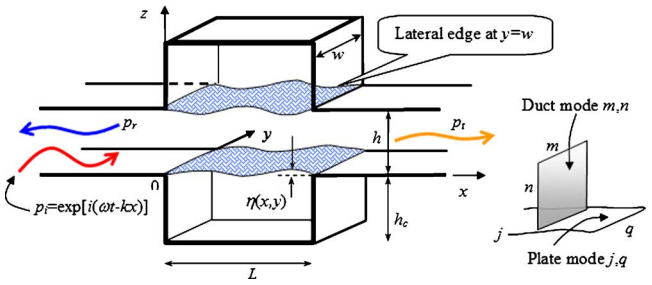


FIG. 1. (Color online) Theoretical model of sound wave reflection by two cavity-backed plates forming part of the otherwise rigid duct walls.

a side-branch wave reflector is. For example, a plate has been used as a means to adjust certain acoustic resonance. Panel absorbers are used in broadcasting studios and other architectural practices (Brown, 1964; Ford and McCormick, 1969; Horoshenkov and Sakagami, 2001), membrane absorbers are used as splitter silencers in the form of arrays of Helmholtz resonators (Frommhold *et al.*, 1994). However, in all these applications the panel, or the membrane, is a component of a resonator which works for a narrow frequency band. The structural mass is often a desirable property to achieve a low resonance frequency. Here, it is shown that the most desirable properties of a plate as a side-branch wave reflector are the high stiffness and low density.

In what follows, theoretical outlines are given in Sec. II, and the relationship between plate/cavity resonance and wave reflection is examined for the three-dimensional acoustic model and two-dimensional plate model. In Sec. III, the model is changed to two-dimensional duct and one-dimensional beam, leading to the optimal wave reflection performance. The effects of plate and cavity damping are also studied. Section IV examines the effect of plate mass and the issue of material selection.

II. THEORETICAL MODEL AND PLATE RESONANCE

A. Three-dimensional theoretical model

As shown in Fig. 1, a segment of a rigid duct of height h and width w is lined by two plates of length L and width w. Each plate is supported by a cavity of depth h_c. The four edges of the plate may have different structural boundaries. The leading and trailing edges located at $x=0,L$ are simply supported. This condition is preferred over the clamped condition for a uniform plate since it allows the plate to respond to incident waves with a higher vibration magnitude. The two lateral edges, defined as those located at $y=0,w$, can be either simply supported, as is the case in the analysis of this section, or free, which is the case for the later sections. A plane incident wave comes from the left to the right with a unit amplitude,

$$p_i = e^{i(\omega t - kx)}, \tag{1}$$

and it induces the plate to vibrate with a transverse displacement of complex amplitude $\eta(x,y)$ and velocity $V(x,y)$ with the same time dependence, $\exp(i\omega t)$, which is henceforth omitted. The plate motion radiates sound in both directions of the duct axis, $\pm x$, and it induces a fluid loading of

$$\Delta p = (\,p|_{z=0+} - \,p|_{z=0-})^{(\eta)}. \tag{2}$$

Note that Δp is defined as that induced by the plate vibration (hence the superscript η) excluding the loading by the incident wave. The plate vibration is governed by the following equation:

$$m_s \frac{\partial^2 \eta}{\partial t^2} + B\nabla^4 \eta + p_i + \Delta p = 0, \tag{3}$$

where $B=Es^3/[12(1-\nu^2)]$ is the bending stiffness, m_s, E, ν, s are, respectively, the plate mass per unit surface area, the Young's modulus, Poisson's ratio, and the plate thickness. The effect of damping is investigated in the next section where relevant formulation is given. In order to solve Eq. (3) both the plate vibration and the fluid loading on the plate are expanded into the *in vacuo* modes of the simply supported plate denoted by $\varphi_{jq}(x,y)$,

$$\varphi_{jq}(x,y) = \sin(j\pi x/L)\sin(q\pi y/w),$$

$$V(x,y) = \sum_{j=1}^{\infty}\sum_{q=1}^{\infty} V_{jq}\varphi_{jq}(x,y),$$

$$V_{jq} = \frac{4}{Lw}\int_0^w dy \int_0^L V(x,y)\varphi_{jq}dx,$$

$$\Delta p^{(\eta)} = \sum_{j,q} V_{jq} \sum_{j',q'} Z_{jqj'q'}\varphi_{j'q'}, \tag{4}$$

$$Z_{jqj'q'} = \frac{4}{Lw}\int_0^w dy \int_0^L \Delta p^{(\varphi_{jq})}\varphi_{j'q'}dx,$$

where $\Delta p^{(\varphi_{jq})}$ denotes the fluid loading induced by the modal vibration of indices (j,q) with a unit amplitude, and $Z_{jqj'q'}$ is the modal impedance defined as the fluid loading of mode (j',q') arising from the plate vibration of mode (j,q) with a unit velocity amplitude. The formulations for $Z_{jqj'q'}$ and the convergence of the solution for the truncated set of equations are documented in Huang (2002) and Huang and Choy (2005) for tensioned membranes. Since $Z_{jqj'q'}$ is the acoustic property of the duct and the cavity, it is independent of the structural mechanics. In other words, $Z_{jqj'q'}$ found for the membrane problem can also be used here. Briefly, the modal impedance problem becomes a standard radiation problem and the solution can be expanded into the duct acoustics modes (Doak, 1973) of indices m in the y direction and n in the z direction, illustrated in the right-hand part of Fig. 1. Using the standard Galerkin procedure, the following set of linear equations are obtained:

$$\mathcal{L}_{jq}V_{jq} + \sum_{j',q'} Z_{j'q'jq}V_{j'q'} + I_{jq} = 0, \quad j,q = 1,2,3,\cdots, \tag{5}$$

where $\mathcal{L}_{jq}$ is the structural mechanics operator,

$$\mathcal{L}_{jq} = m_s i\omega + \frac{B}{i\omega}\left[\left(\frac{j\pi}{L}\right)^4 + 2\left(\frac{j\pi}{L}\right)^2\left(\frac{q\pi}{w}\right)^2 + \left(\frac{q\pi}{w}\right)^4\right], \tag{6}$$

and I_{jq} is the modal excitation of the given incident wave p_i,

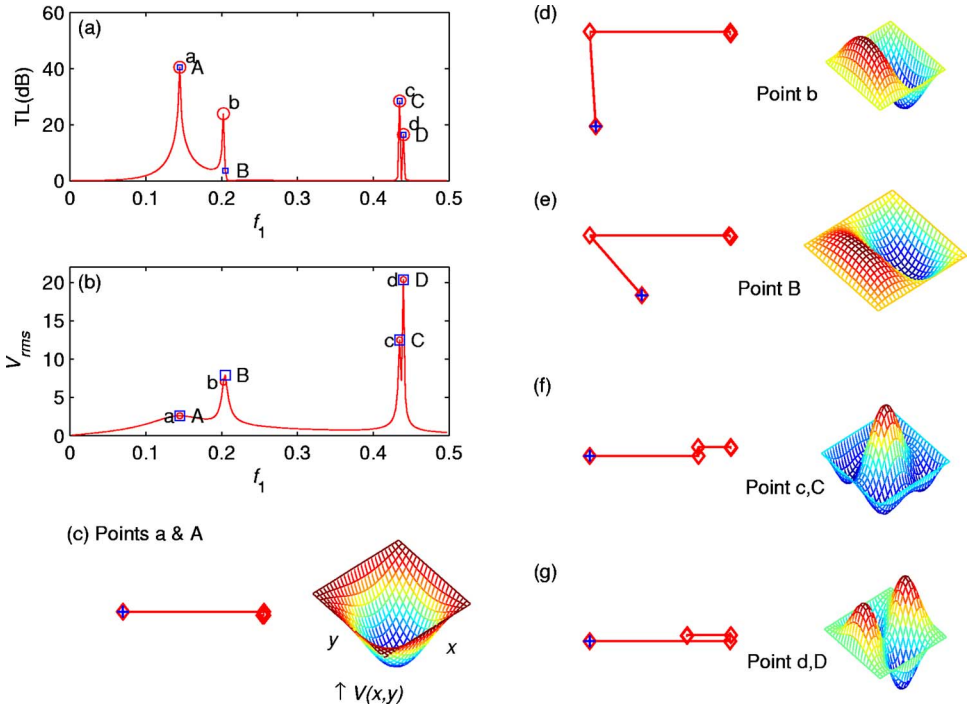


FIG. 2. (Color online) Resonance of the cavity-backed, side-branch plate with $L=h$, $h_c=h$, $B_1=0.001$. (a) System TL. (b) Plate vibration velocity (V_{rms}, relative scale). (c) Modal decomposition of the reflected sound and the instantaneous plate vibration velocity distribution for the first resonance frequency, point A, where TL also peaks. (d) Second TL peak point b. (e) Second resonance frequency point B. (f) Third and (g) fourth TL peak and resonance frequency points.

$$I_{jq} = \frac{4}{Lw}\int_0^w \left(\int_0^L p_i(x)\varphi_{jq}(x,y)dx\right)dy$$
$$= 4\frac{1-\cos(q\pi)}{q\pi} \times j\pi e^{ikL/2}\left[\frac{1-e^{i(-kL+j\pi)}}{(j\pi)^2-(kL)^2}\right]. \quad (7)$$

The set of linear equations (5) is truncated to a finite number of modes in all three dimensions: axial plate mode j, lateral plate mode q, and duct acoustics modes, m,n. Numerical tests show that the modal convergence properties for the plate problem are similar to the membrane problem, and details are not given here. In the numerical examples given below, the plate modes are truncated to 20 in both length and width directions while 50 modes are used for the duct acoustics. Once the complex modal velocity V_{jq} is obtained, the far-field sound radiation to the upstream represents the reflection wave calculated by

$$p_r = \sum_{j,q} p_{rjq} = \sum_{j=1}^{\infty} p_{rj}, \quad p_{rj} = \sum_{q=1}^{\infty} p_{rjq},$$

$$p_{rjq} = \frac{\rho_0 c_0}{2hw^2} V_{jq} \int_0^w \left(\int_0^L \varphi_{jq}(x,y)e^{-ikx}dx\right)dy, \quad (8)$$

where ρ_0, c_0 are the fluid density and speed of sound, respectively, and the contribution from each axial mode, j, is calculated by a summation over the lateral mode of index q. The radiation to the downstream combines with the incident wave to become the transmitted wave, which is written below together with the definition of the transmission loss TL,

$$p_t = p_i + \frac{\rho_0 c_0}{2hw^2}\sum_{j,q} V_{jq}\int_0^w \left(\int_0^L \varphi_{jq}(x,y)e^{+ikx}dx\right)dy, \quad (9)$$

$$\text{TL} = 20\log_{10}\frac{|p_i|}{|p_t|}.$$

When the two lateral edges at $y=0,w$ are set free, the problem is degenerated into two dimensions for the duct acoustics and one dimension for the plate (or beam). The above equations can be easily adapted by eliminating the modes in the y direction with indices q and m. The formulations for the 2D problem are similar to Huang (2002) and are not listed separately here for the sake of brevity.

In all numerical examples, parameters are normalized. Three basic normalization quantities are chosen, fluid density ρ_0, speed of sound c_0, and duct height h. Time and pressure quantities are normalized by c_0/h and $\rho_0 c_0^2$, respectively. By doing so, three most important parameters emerge,

$$f_1 = \frac{\omega h}{2\pi c_0}, \quad m_1 = \frac{m_s}{\rho_0 h}, \quad B_1 = \frac{B}{\rho_0 c_0^2 h^3}, \quad (10)$$

which are, respectively, the dimensionless frequency, the plate-to-fluid mass ratio, and the dimensionless bending stiffness. Here, a dimensionless variable is identified by a subscript of "1" added to its dimensional variable of the same symbol, while dimensional variables are used as much as possible. For the sake of comparative study, tensioned membranes are also investigated, and the relevant dimensionless tension is defined as $T_1=T/(\rho_0 c_0^2 h^2)$, where T is the dimensional tensile force applied.

B. Compact plate resonance

In order to best appreciate the fluid-plate interaction, an example of compact plate resonance is given with the following parameters:

$$m_1 = 1, \quad L = w = h_c = h, \quad B_1 = 0.001.$$

The results are shown in Fig. 2 with seven subfigures. Figure 2(a) shows the transmission loss (TL) spectrum and Fig. 2(b) the root-mean-square value of the plate vibration velocity integrated over the whole plate denoted by V_{rms}. Peaks in the TL spectrum are indicated by open circles and marked by lower-case letters, while the V_{rms} peaks are marked by open squares and upper-case letters. The two sets of peaks are shown in both subfigures. Figures 2(d)–2(g) are details of the vibration velocity distribution and the sound reflection for individual frequencies marked in Figs. 2(a) and 2(b).

As shown in Figs. 2(a) and 2(b), the first TL peak, "*a*," is over 40 dB and the same frequency position happens to be a rather flat peak of V_{rms}, denoted by "*A*." This common peak features mainly the fundamental mode vibration, $j=q=1$, shown in Fig. 2(c). In the left-hand side of Fig. 2(c), the sound reflection is decomposed into contributions from various axial modes of index j; cf. Eq. (8). The contribution of each axial mode is represented by a vector bounded by open diamonds based on the complex amplitude of the sound reflection. The vectorial summation indicates how the modal contributions interfere with each other. In this particular frequency of $f_1=0.145$ for points *a* and *A*, there is hardly any contribution from modes of $j>1$. The reason why V_{rms} is so low is as follows. At this frequency, most sounds are reflected by the plate. The plate becomes a pressure release surface, and the actual sound pressure is an antinode ($p \approx 0$) there. In other words, part of the excitation force is canceled out by the plate response. As shown in Figs 2(a) and 2(b), TL decreases when $f_1>0.145$ before it increases again, while V_{rms} decreases only slightly before it approaches its next peak. The TL peak deviates from the V_{rms} peak although they are not too far apart. At the point where the plate vibration is the highest, point "*B*," TL is actually very low. This is the best illustration of the fact that the plate resonance is not necessarily good for sound reflection. Modal decomposition of the sound reflection and the instantaneous plate vibration velocity are shown in Fig. 2(d) for point "*b*" and in Fig. 2(e) for point *B*. Figure 2(d) shows that good silencing performance is achieved when the reflections by the first two axial modes of $j=1,2$ are essentially out of phase (the vectors being perpendicular to each other). As shown in Fig. 2(e), a slight adjustment of frequency to point *B* shifts such phase relation, and the two modes interfere destructively. When the dimensionless frequency is further increased towards 0.44, two more sets of peaks are identified in Figs. 2(a) and 2(b). This time, the TL peaks also coincide with the V_{rms} peaks like the first point at $f_1=0.145$, but they can actually be separated by using a finer frequency resolution. Details of the vibration are shown in Fig. 2(f) for points "*c*" and "*C*," and Fig. 2(g) for points "*d*" and "*D*." Two observations are made here. First, the main contributions of sound reflection derive from the first and third modes, $j=1,3$. Second, the plate vibrations are very different for the two sets of peaks, although the frequencies are not too far apart.

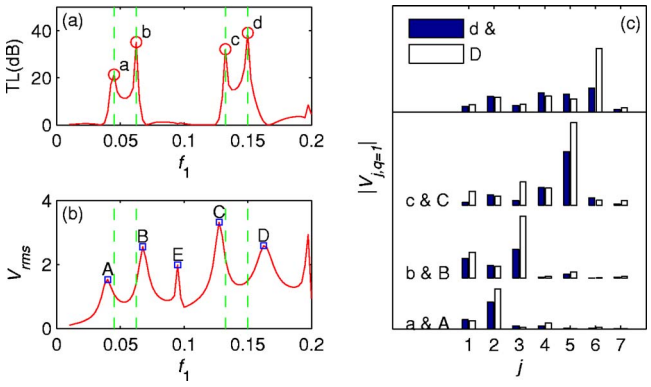


FIG. 3. (Color online) Response of a long plate with $L=5h$, $h_c=h$, $B_1=0.003$. (a) Transmission loss. (b) Root-mean-square plate vibration velocity V_{rms}. (c) Comparison of modal responses at the TL peaks (filled bars) and a nearby resonance (open bars) with frequency locations labeled in the other two subfigures.

C. Noncompact plate

The results of the short plate studies shown in Fig. 2 indicate that good silencing performance occurs near the plate resonance. This implies that, in order to obtain good low-frequency performance, the plate has to be long. The results for the following configuration:

$$m_1 = 1, \quad L = 5h, \quad w = h_c = h, \quad B_1 = 0.003,$$

are shown in Fig. 3 with frequencies zoomed in to $f_1=0$ to 0.2. For such a long plate, the first mode has very high radiation resistance and is very difficult to excite at low frequencies. Comparing the results in Figs. 3(a) and 3(b) with those of Figs. 2(a) and 2(b), there are much larger and clearer separations between the TL peaks, which are marked a,b,c,d in Fig. 3(a), and the V_{rms} peaks, which are marked by A,B,C,D in Fig. 3(b), for the long plate than for the shorter ones studied above. The resonance frequency labeled as "*E*" does not have a corresponding TL peak and is left out of the discussion. The modal amplitudes for the first peaks, *a* and *A*, are shown in the bottom row of Fig. 3(c). It is found that there is a strong second mode vibration, $j=2$, instead of the first, $j=1$, since the latter is too difficult to excite at low frequencies due to the cavity fluid stiffness. Nevertheless, the first mode still makes the dominant contribution to the sound reflection (not shown) at the first TL peak point *a*. The modal amplitudes for the second group of peak points, *b* and *B*, are shown in the second row from the bottom of Fig. 3(c). Here, the first mode response is enhanced but the third mode dominates. Point *c* and *C* feature the fifth mode, while *d* and *D* the sixth mode. It may be concluded that, for low structural bending stiffness, resonances featuring dominant axial modes occur at increasing frequencies, but not all resonance can produce a TL peak at its nearby frequencies. There is a clear frequency deviation between TL peaks and the resonance frequencies; the former generally consists of a mixture of different modes interacting constructively.

III. CAVITY AND PLATE OPTIMIZATIONS

In the previous studies for the drumlike silencer (Huang and Choy, 2005), it is found that high tension is necessary to obtain better silencing performance since a strong structure promotes low-order modes which are more effective in reflecting sound than high-order modes. However, the lateral structural constraints prohibit the overall response, and are generally unhelpful for responses at nonresonance frequencies. This is also found to be the case for plate silencer. Details are not given for such similar conclusions. In this section, a comparative study is conducted for the drumlike silencer and the plate silencer, both limited to the two-dimensional configuration in which the two lateral edges, $y=0,w$, are free to move. Since the frequency of interest is limited to that below the cut-on frequency of the duct, $f_1=0.5$, plane incident wave does not excite the lateral modes. The plate therefore behaves as a one-dimensional beam, and the duct acoustics is limited to two dimensions. The proof of such 1D structural response is given in Huang and Choy (2005) for the drumlike silencer, and is not repeated here. For simplicity, the configuration with two free lateral edges is referred to as the 2D model in the following discussions. To materialize such a 2D configuration, a small gap has to be allowed between the lateral plate edges and the duct walls. Too big a gap would cause serious leaking of sound, but a previous study for the 2D membrane configuration (Choy and Huang, 2002) showed that common design techniques used in minimizing the leakage flow between stationary and moving parts in a rotary machine would work. In the said experimental study using a duct of 5 by 5 cm cross section, a slit of 2 mm in depth was created in each lateral duct wall allowing the membrane to be inserted into the recess with an all-round clearance of no more than 0.5 mm. The experimental results demonstrated that the acoustic leakage was prevented satisfactorily. It is believed that, when the dimension of the duct cross section is increased, such gap would prevent acoustic leaking even more effectively.

In this section, the cavity shape is optimized for a given volume. The optimal shape is then fixed for the parametric study of other variables such as the plate and cavity damping. Comparative study is conducted between the plate silencer and the drumlike silencer to understand the commonality and the differences in terms of the intermodal interference.

A. Cavity shape optimization

The procedure described in Huang (2004) is followed here. For a given cavity volume, Lh_c, the aspect ratio, L/h_c, can be varied by changing the length and cavity depth simultaneously. For a given structure-to-air mass ratio, m_1, the dimensionless bending stiffness, B_1, can be varied by changing the material elasticity. The objective function is the frequency ratio, f_b/f_a, where the transmission loss, TL, is above a criterion value, $\mathrm{TL}_{\mathrm{cr}}$, over the whole frequency band $f\propto[f_a,f_b]$, or $f_1\propto[f_{1a},f_{1b}]$ in dimensionless form, where $f_{1a}=f_ah/c_0, f_{1b}=f_bh/c_0$. The choice of $f_b/f_a=f_{1b}/f_{1a}$ as the objective value instead of f_{1b}-f_{1a} is based on the consideration that the low-frequency noise control is a true technical challenge. The criterion value $\mathrm{TL}_{\mathrm{cr}}$ is somewhat arbitrary but a procedure described in Huang (2004) recommends that the value be chosen as the peak transmission loss for an expansion chamber whose cavity volume is three times the actual cavity volume in the silencer. For the example of volume $Lh_c=5h^2$, the criterion value thus derived is rounded up to 10 dB. This configuration and the $\mathrm{TL}_{\mathrm{cr}}$ value are adopted here. The results of the optimization are shown in Fig. 4 for the following given parameters: $m_1=1, Lh_c=5h^2$. The procedure is described as follows. With some parameters fixed as above, two variables remain: dimensionless cavity length $L_1=L/h$ and plate bending stiffness B_1. The optimization begins by fixing the value of L_1 and searching for the optimal B_1 so that f_b/f_a is the highest. This set of optimal results may be called a local optimum. The highest value of f_b/f_a for all lengths is identified as the global optimum. The local optimal parameters, f_{1a}, f_{1b}, and $B_{1\mathrm{opt}}$, are functions of the cavity length L_1, and are presented in Figs. 4(a)–4(c). Figure 4(a) shows the peak logarithmic bandwidth as a function of L_1. It is found that the global optimum is achieved right at $L_1=5$, $h_c/h=1$, since the result for $L_1=5.1$ is worse than that of $L_1=5$. Figure 4(b) shows that the dimensionless lower frequency limit, $f_{1a}=f_ah/c_0$, as a function of L_1 and the lowest f_{1a} is achieved for $L_1=5$. Figure 4(c) shows that the optimal bending stiffness, $B_{1\mathrm{opt}}$, increases with L_1 and there is no peaky appearance around $L_1=5$. The following set of optimal results is found:

$$L_{\mathrm{opt}}=5h,\quad B_{1\mathrm{opt}}=0.129,\quad f_{1a}=0.0353,$$

$$f_{1b}=4.25f_{1a}=0.149.$$

The right-hand column of Fig. 4 shows some spectral comparisons with frequency points where $\mathrm{TL}=\mathrm{TL}_{\mathrm{cr}}$ marked by

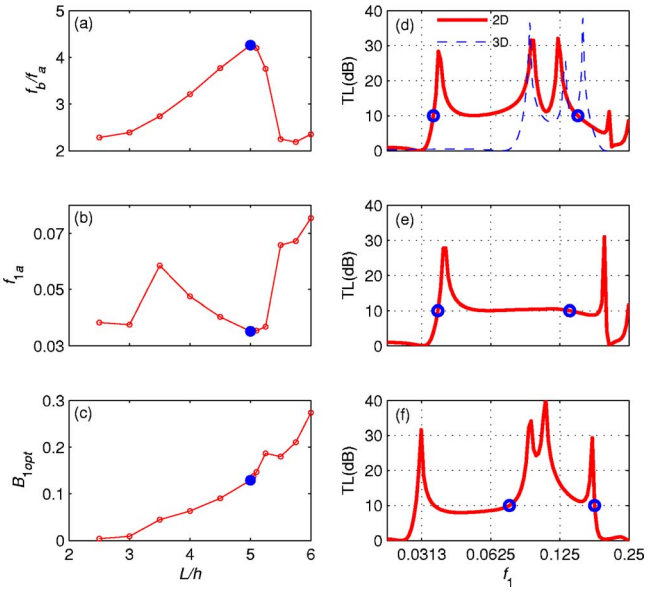


FIG. 4. (Color online) Logarithmic bandwidth optimization in terms of the B_1 and L/h while the cavity volume is fixed at $5h^2$. (a) Logarithmic bandwidth f_b/f_a. (b) Dimensionless lower bandlimit f_{1a}. (c) Dimensionless optimal bending stiffness $B_{1\mathrm{opt}}$. (d) Comparison of the optimal performance of the 2D configuration (solid line) with $L=5h$ and the optimal 3D configuration with $L=4h$ (dashed line). (e) Optimal performance for $L=5.25h$ showing a nearly flat stopband. (f) Optimal design for $L=6h$.

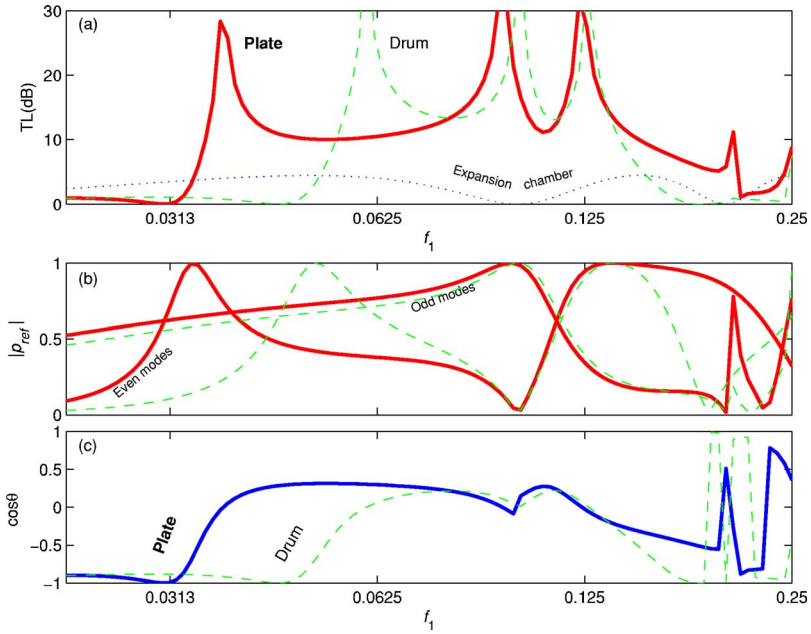


FIG. 5. (Color online) Comparison of the plate silencer (solid curves, with $B_{1\text{opt}}=0.129$) with drumlike silencer (dashed curves, with $T_{1\text{opt}}=0.475$) with the same cavity geometry of $L=5h$, $h_c=h$, and the same mass ratio of $m_1=1$. (a) TL. (b) Reflection sound decomposition into odd and even modes. (c) Odd-even interference index.

open circles for easy identification of stopbands. Figure 4(d) compares the overall optimum for $L=5h$ (solid curve) with the typical results for a three-dimensional model in which the plate is simply supported along all four edges (dashed curve) with $L=4h$, $h_c=1.25h$. The latter is not optimized under the strict criterion of $\text{TL}\geqslant\text{TL}_{\text{cr}}$ since TL=10 dB is hard to achieve for such 3D configuration described in the last section. The comparison shows that the major improvement of the 2D model from the 3D model is the low-frequency performance. The reason why the 2D configuration performs better is common with that of the drumlike silencer (Huang and Choy, 2005) and detailed analysis is not given here.

Figure 4(e) shows a very interesting 2D configuration of $L=5.25h$ for which the TL is almost flat over much of the stopband. Similar spectrum also appears for $L=5h$ when B_1 exceeds the optimal value of $B_{1\text{opt}}=0.129$, and the second and third spectral peaks merge. For the case of $L=5.25h$ shown in Fig. 4(e), the logarithmic bandwidth f_b/f_a is deemed to be lower than that of $L=5h$ due to the slight deterioration of TL near f_{1b}. Note also that there are only two spectral peaks within the band. Figure 4(f) shows that, when L_1 is further increased to 6, the bandwidth is severely narrowed due to the rather uniform drop of TL between the first two peaks in the spectrum. If the strict rule of $\text{TL}\geqslant\text{TL}_{\text{cr}}$ is somehow relaxed, it may be said that the silencer works for an ultralow frequency for a given geometry. If the dimensional frequency of interest is given, it means that the plate dimension may be reduced. For this reason, it might well be that this configuration is practically more preferable.

B. Comparison with the drumlike silencer

Coincidentally, the optimal cavity shape for the drumlike silencer for the specific volume of $Lh_c=5h^2$ is $L_{\text{opt}}=5.7h$ (Huang, 2004), not too far from $5.0h$ found for the plate silencer here. For simplicity, the performance of the drumlike silencer with $L=5h$, $h_c=h$ and the dimensionless optimal tension $T_1=0.475$ is chosen for comparative study in Fig. 5. The results for the plate silencer are shown in solid lines and those of the drumlike silencer in dashed lines. The spectral comparison is shown in Fig. 5(a) together with the performance of the expansion chamber of the same cavity geometry (dotted line). Evidently, the superiority of the plate silencer lies in the low-frequency region. Figure 5(b) compares the decomposition of the sound reflection, cf. Eq. (8) into the summation over odd *in vacuo* modes, $j=1,3,5,\ldots$ and that over the even *in vacuo* modes, $j=2,4,6,\ldots$. Focusing on the stopband below the dimensionless frequency of around 0.15, it is found that the odd-mode contribution for the two types of silencers is almost identical. For the even modes, however, the plate-type silencer outperforms the drumlike silencer with its first peak at a much lower frequency. In fact, this peak of the even-mode reflection is related to the second-mode resonance determined by the structural restoring forces (Huang, 2002), tension or bending moment, and the total mass of the coupled system. The virtual mass imposed by the cavity air is found to be much more than the structural mass for this example, and the acoustic impedance analysis for the drumlike silencer remains unchanged. The difference between the two silencers is this. For the plate silencer, a bending stiffness of $B_{1\text{opt}}=0.129$ is chosen in such a way that the second mode resonance occurs at a frequency lower than that of the drumlike silencer under $T_{1\text{opt}}=0.475$. Of course, it is also possible to choose a lower value of tension so that the second mode resonance for the drumlike silencer also occurs at a lower frequency. But, then the TL in the low-frequency regime cannot be maintained at a level above TL_{cr} due to the adverse cross-modal interference, which is shown in Fig. 5(c) as the part of the dashed curve near the label "Drum" with negative values. If the phase angle between the reflected wave from the odd modes and those of the even modes is denoted by θ, the interference index is given by $\cos\theta$, which varies from the lowest extreme of complete cancellation to complete addition. For frequency $f_1=0.03$ to 0.05, the plate silencer

works since its odd-even interference is neutral, $\cos\theta \approx 0$, while the drumlike silencer fails as its odd-even interference is rather destructive. Note that the structural responses at the three spectral peaks are rather similar to those of Huang, (2002), namely, the first peak is from the combined effect of the first and second *in vacuo* modes, the second peak is from the first mode, and the third peak is from the second mode.

C. Effects of plate and cavity damping

The effect of damping is investigated by two approaches. One is to consider various levels of structural damping on the plate, and the other is to add sound absorption materials in the cavity. For plate damping, the model adopted is written as

$$m_s(1 - i\sigma_m)\frac{\partial^2 \eta}{\partial t^2} + B\nabla^4 \eta(1 + i\sigma_s) + p_i + \Delta p = 0 \tag{11}$$

for harmonic oscillations, $\eta \propto e^{i\omega t}$, where σ_m and σ_s are, respectively, the mass and stiffness damping coefficients in the Rayleigh damping model. The mass damping coefficients acts mainly on the high-frequency oscillation while the stiffness damping on low frequencies. The exact damping mechanisms are hard to model. For the purpose of qualitative illustration, it may suffice to assume a balanced damping mechanism as

$$\sigma = \sigma_m = \sigma_s, \tag{12}$$

and focus on the overall effect of σ on the silencing performance. For the second method of adding sound absorption material into the cavity, the fluid inside the cavity is modeled as one in which the density and the speed of sound are modeled as complex quantities. Following Mechel and Vér (1992), the mass and momentum conservation laws are combined in a familiar manner to form the wave equation in the porous media,

$$\frac{1}{\rho_c c_c^2}\frac{\partial p}{\partial t} + \nabla \cdot \mathbf{v} = 0, \quad \rho_c \frac{\partial \mathbf{v}}{\partial t} + \nabla p = 0 \rightarrow \nabla^2 p - \frac{1}{c_c^2}\frac{\partial^2 p}{\partial t^2} = 0, \tag{13}$$

where $\rho_c, \mathbf{v}, c_c$ are, respectively, the complex density, particle velocity and wave speed. Note that the particle velocity is defined in such a way that $\rho_0 \mathbf{v}$ represents the true oscillating mass flow of the fluid through the porous material per unit cross section instead of $\rho_c \mathbf{v}$. Similar to sound wave propagation in lossless fluid, a velocity potential can be introduced and is related to sound pressure and velocity perturbations as follows:

$$\mathbf{v} = \nabla\phi, \quad p = -\rho_c \frac{\partial \phi}{\partial t}, \quad \left(\nabla^2 - \frac{1}{c_c^2}\frac{\partial^2}{\partial t^2}\right)\phi = 0. \tag{14}$$

The boundary condition on a solid wall with its unit normal vector $\mathbf{n}$ and the plate surface at $z=0$ are given as follows:

$$\left.\frac{\partial \phi}{\partial n}\right|_{\text{hard wall}} = 0, \qquad \left.\frac{\partial \phi}{\partial z}\right|_{\text{plate}} = \frac{\partial \eta}{\partial t}. \tag{15}$$

The sound pressure on the lower surface of the plate is obtained as $p|_{z=0-} = -\rho_c \partial\phi/\partial t$ and is used in calculating the fluid loading, $\Delta p = p|_{z=0+} - p|_{z=0-}$, used in Eq. (3). The complex wave number and impedance are given by empirical formulas (Mechel and Vér, 1992) using the data of glass fiber in the low frequency range,

$$Z_c = \rho_c c_c, \quad Z_0 = \rho_0 c_0, \quad k_0 = \omega/c_0, \quad R = -U^{-1} dp/dx,$$

$$k_c/k_0 = c_0/c_c = 0.396E^{-0.458} + i(1 + 0.135E^{-0.646}),$$

$$Z_{1c} = Z_c/Z_0 = (1 + 0.0668E^{-0.707}) - i \times 0.196E^{-0.549},$$

$$E = \rho_0 f/R \leqslant 0.025. \tag{16}$$

Note that R is the dimensional flow resistance, and a dimensionless version is defined below in the current study,

$$R_1 = \frac{Rh}{\rho_0 c_0}. \tag{17}$$

The parameters of σ and R_1 are collectively called damping factors for simplicity.

The effects of plate damping and sound absorption material filled in the cavity are investigated separately. For each type of damping mechanism and the level of damping factor chosen, the plate bending stiffness B_1 is varied to seek the widest logarithmic stopband f_b/f_a, while the geometry of $L = 5h$, $h_c = h$, and the dimensionless plate mass of $m_1 = 1$ are fixed in this optimization exercise. Again, the global optimum is identified for the highest bandwidth. The results are shown in Fig. 6. Figure 6(a) shows the optimal stopband as a function of the damping factor for the two damping mechanisms, with the global optimal point marked by a filled circle (with cavity damping) and a diamond (plate damping). For plate damping (solid line) the overall optimum is found at $\sigma_{\text{opt}} = 0.1$, $f_b/f_a = 4.54$, which is only slightly better than the bandwidth of $f_b/f_a = 4.27$ without damping. When σ exceeds around 0.12, the bandwidth suffers a sudden drop. For cavity damping (dashed line), a maximum bandwidth of $f_b/f_a = 4.84$ is achieved at $R_{1\text{opt}} = 0.07$. There is no sudden drop in bandwidth as R_1 increases further but excessive damping reduces the bandwidth. Figure 6(b) shows the optimal bending stiffness, $B_{1\text{opt}}$, as a function of the damping factors with the global optimal points shown in the same filled symbols as in Fig. 6(a). Spectral comparisons between the plate damping factors of $\sigma = 0$, σ_{opt}, 0.2 are given in Fig. 6(c). It is found that sharp peaks predicted for $\sigma = 0$ are smoothed out when there is plate damping, and excessive damping causes the TL to drop between the second and third peaks; hence, the sudden drop of f_b/f_a shown as the solid line in Fig. 6(a). The results for the cavity damping of $R_1 = 0$, $R_{1\text{opt}}$, 0.2 are shown in Fig. 6(d). It is found that the cavity damping brings a rather uniform TL across the whole frequency band, and excessive damping suppresses the responses near the lower limit of the stopband.

The contributions of the damping mechanisms towards the overall silencing effect are shown in Fig. 7 when the plate bending stiffness is fixed at its optimal value found in Fig. 6. For a plate with damping, the optimal value of $B_{1\text{opt}} = 0.132$ is given and the results are shown in Fig. 7(a). With the optimal damping factor of $\sigma_{\text{opt}} = 0.1$ (thick solid line), the average contribution of sound absorption by the damped

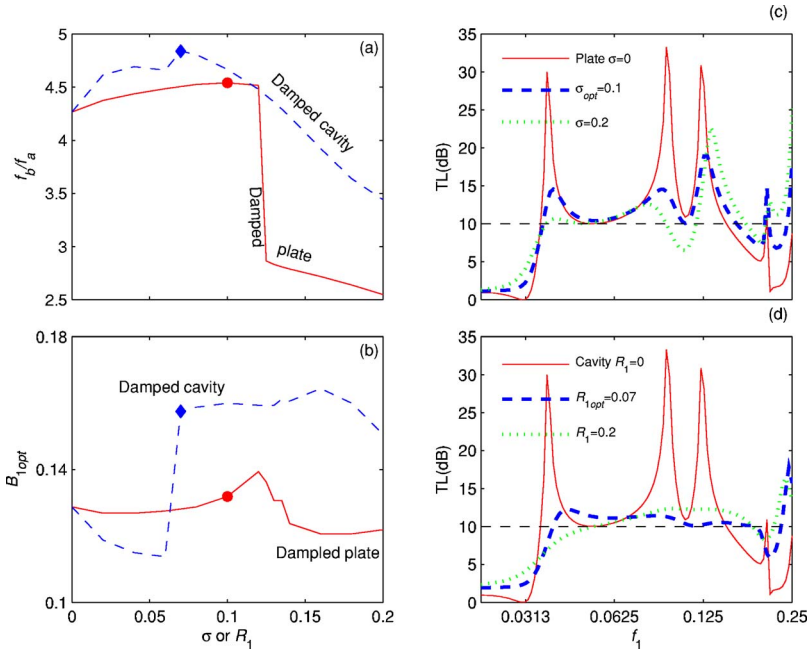


FIG. 6. (Color online) Effect of plate damping and porous material filled cavity for $L=5h$, $h_c=h$. (a) Logarithmic bandwidth. (b) Optimal bending stiffness. (c) Spectral comparison for plates with various damping factors. (d) Spectral comparison for cavity filled with sound absorption materials of various flow resistance.

plate is around 9% in the stopband marked by the two thin vertical lines. When the damping factor is increased to 0.2, the sound absorption is indeed increased to an average of 16%, but the overall TL is reduced, as shown in Fig. 6(c), due to the reduced sound reflection. For a plate without damping but the cavity filled with sound absorption material, the optimal bending stiffness is $B_{1\text{opt}}=0.158$ and the results are shown in Fig. 7(b). The contribution of sound absorption in this case is much more prominent than that of plate damping. Within the stopband marked by the two thin vertical lines, the average contribution by the cavity damping is 46% when $R_{1\text{opt}}=0.07$ is used, and is 63% when a higher value of $R_1=0.2$ is used. Again, excessive use of sound absorption mechanism stifles the plate response and reduces the overall transmission loss.

IV. EFFECT OF PLATE MASS AND CHOICE OF MATERIAL

A. Effect of plate mass

The effect of plate mass is investigated by varying the plate-to-air mass ratio m_1 while the geometry is fixed as $L=5h$, $h_c=h$. For each mass ratio given, bending stiffness is varied to find the optimal bandwidth f_b/f_a. The results are shown in Fig. 8. Figure 8(a) shows the optimal bandwidth as a function of m_1. All calculated points are marked by crosses. The overall optimum is shown by a filled triangle, while the result for $m_1=1$ is shown by a filled square. The horizontal scale is logarithmic except the special point of $m_1=0$, which is shown as an open circle. For the range of $m_1=0.05$ to 0.5, the bandwidth hovers around $f_b/f_a\approx 6.0$ with the overall optimum located at $m_1=0.25$. Figure 8(b) gives $B_{1\text{opt}}$ as a function of m_1 which shows a monotonous trend of increase. Figures 8(c) and 8(d) compare the optimal spectra for different values of m_1. Figure 8(c) explains why the bandwidth of $m_1=0$ suffers from a sudden drop of f_b/f_a from $m_1=0.05$ to 0.01 in Fig. 8(a). The reason is found as the mild decrease of TL between the first and second spectral peaks. For practical purposes, such drop can be ignored and a general conclusion can be drawn that a lighter plate provides wider stopband. Figure 8(c) also shows that, when the plate mass is increased, the stopband shrinks from both low and moderately high frequencies. Figure 8(d) shows the overall optimum (thick solid line) of $m_1=0.25$, and a rather heavy plate of $m_1=10$ (dashed line). The latter has nearly one octave stopband ranging from $f_1=0.029$ to 0.055. The bandwidth may be low, but the location in the very low frequency region could prove to be a practical advantage for the noise control at very low frequencies. In addition, a higher mass ratio also makes it easier for such a plate to be realized with a low-cost material and structure.

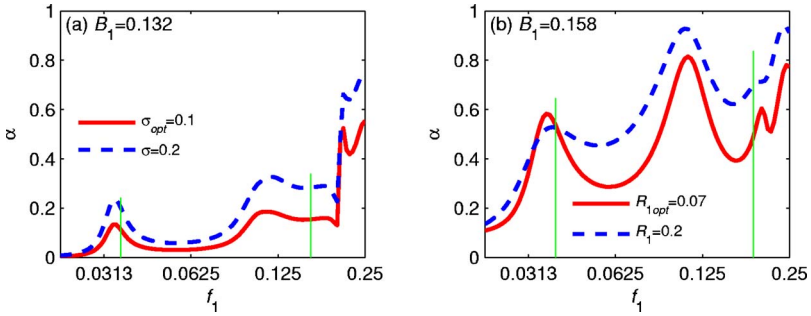


FIG. 7. (Color online) Absorption coefficient of sound energy, α. (a) Damped plate with $B_1=0.132$; (b) damped cavity with $B_1=0.158$.

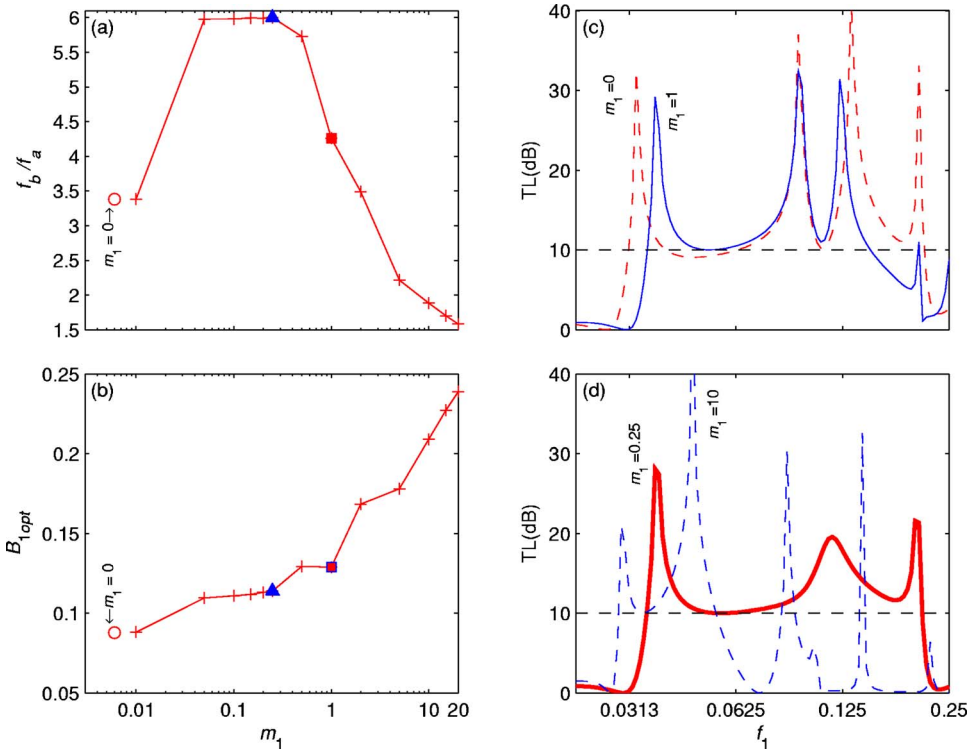


FIG. 8. (Color online) Effect of mass on (a) logarithmic bandwidth f_b/f_a, and (b) dimensionless optimal bending stiffness $B_{1\text{opt}}$, with spectral comparisons for (c) light, and (d) heavy plates. The cavity geometry is $L=5h$, $h_c=h$.

B. Choice of plate material and structure

As is shown below, it is generally difficult to satisfy the condition of a high bending stiffness with a low mass ratio. Stiffness-enhancing structures may have to be used. Denote the overall thickness of the plate as s and the volume density of the plate as ρ_s, which may be much lower than the density of the core material. The dimensionless bending stiffness, mass ratio, and the dimensional moment of inertia I are expressed below,

$$B_1 = \frac{EI}{\rho_0 c_0^2 w h^3}, \quad m_1 = \frac{\rho_s s}{\rho_0 h}, \quad I = \int_A y^2 dA = Cwh^3, \quad (18)$$

where the constant C is 1/12 for a homogeneous plate of rectangular cross section and the ratio of the actual value of C to the normal value of 1/12 can be called stiffness efficiency for a sandwich beam construction. Substituting the thickness s by $m_1 h\rho_0/\rho_s$, the dimensionless bending stiffness can be rewritten as follows:

$$B_1 = Cm_1^3 \frac{E}{\rho_0 c_0^2}\left(\frac{\rho_0}{\rho_s}\right)^3. \quad (19)$$

For a given material and sandwich beam construction, the most influential parameter is the dimensionless mass ratio m_1. When the mass ratio is increased by increasing the plate thickness, the required optimal bending stiffness also increases, cf. Fig. 8(b), and the bandwidth reduces, cf. Fig. 8(a). For $m_1=0.5,1,5$, the required values of the optimal bending stiffness are $B_1=0.114$, 0.1289, and 0.1780. The increase of $B_{1\text{opt}}$ with respect to m_1 is mild, roughly proportional to $m_1^{0.15}$ in this region of mass ratio. In other words, the requirement for $B_{1\text{opt}}$ is much easier to satisfy if the allowed mass ratio is increased with the price of having a much smaller stopband. For $m_1=5$, the logarithmic bandwidth is $f_b/f_a=2.2$. If this performance is accepted as the lower limit of a design, the mass ratio as high as 5 is considered to be relevant, and is included in the following analysis alongside the moderate mass ratio of unity.

Equation (19) also shows that, for a given set of m_1, B_1 to be satisfied, the most crucial material property is E/ρ_s^3, which is typical in the selection of weight-saving materials in aerospace applications. According to Chart 1 given in Ashby (1999), and Ashby and Johnson (2002), the best material for this purpose would be balsa, a type of wood, polyvinychloride (PVC), a type of polymer, or carbon fiber-reinforced polymer. Two materials are examined here using the normal rectangular cross section for the plate, for which $C=1/12$. The first material is the aluminum alloy with $\rho_s=2700$ kg/m^3, and $E=75$ GPa. Taking the air properties as $\rho_0=1.225$ kg/m^3, $c_0=340$ m/s, the dimensionless bending stiffness calculated by Eq. (19) is $B_1=4.12\times10^{-6}$, 5.15×10^{-4} for the mass ratios of 1 and 5, respectively, far below the required values of 0.1289 and 0.1780. It is concluded that a simple aluminum plate is about 250 times too elastic for a plate silencer even with a mass ratio of 5. The second material chosen is the PVC foam with $\rho_s=40$ kg/m^3, and $E=200$ MPa, a set of moderate values taken from Chart 1 of Ashby and Johnson (2002). The relative bending stiffness is found as $B_1=0.0034, 0.4226$ for the mass ratios of 1 and 5, respectively. In fact, this material would satisfy the optimal bending stiffness condition shown in Fig. 8(b) when the mass ratio is around 3.5.

A composite beam with a light core material like PVC foam bonded together with aluminum facings can be employed to achieve the effect of having an overall stiffness determined mainly by the elastic modulus of aluminum while the total mass is reduced drastically by the light core. Sandwich beams with honeycomb, I-beam, or corrugated

cores may also be considered to achieve the same purpose. Among the above structures, the I-beam made of a homogeneous material is perhaps most suitable for the purpose of illustrating how such structures can help satisfy the parametric requirement predicted by the current study. An I-beam cross section is mainly characterized by its depth, here denoted as s, and its flange width, w. Take the standard I-beam designated as S380×74 as an example. It has a depth of $s=381$ mm and width $w=143$ mm, the latter being roughly twice the second designation number, cf. Gere (2004). The discussion here is limited to the shape instead of the actual dimension of w and s. As listed in the tables of Gere (2004), the moment of inertia is $I=201\times10^6$ mm^4 and the cross section is 9500 mm^2. To achieve the same moment of inertia I, a plate of rectangular cross section must have a different depth s calculated by $I=\frac{1}{12}ws^3$, $s=\sqrt[3]{12I/w}\approx256$ mm. The cross section of such rectangular plate is $256\times143=36\,608$ mm^2, which is $36\,608/9500=3.85$ times that of the I-beam of the same moment of inertia. In other words, the weight-saving efficiency for such a typical I-beam is 3.85. Since the PVC foam described above can satisfy the parametric requirement of the plate silencer when the structure-to-air mass ratio is about 3.5, the weight-saving efficiency of 3.85 means that an I-beam made purely by PVC foam can satisfy the parametric requirement with a mass ratio of less than 1.0. The use of a thin rigid facing material would further enhance the structural rigidity for any given thickness s. As a result, the requirement of very light core material can be further relaxed. It is expected that a more common material like balsa may be adopted in a sandwich structure.

When using such smart structures, however, care must be taken to check whether the composite structure vibrates as one piece as assumed by the theoretical model, and whether the structural damping coefficient is low enough to allow vigorous plate response to achieve a high degree of sound reflection. More practical work needs to be done in this aspect, but it is argued that the optimal parameters predicted in this study are entirely within the reach of existing known light materials possibly with the help of composite beam construction.

V. CONCLUSIONS

This theoretical study reveals that, in a duct, a plate covering a side-branch rigid cavity can function effectively as a low-frequency wave reflector over a very broad frequency band, about one octave band broader than a drumlike silencer of the same cavity geometry. Effects of various boundary conditions and parameters are investigated. The following specific conclusions are made:

(1) When all edges of the plate are restrained to the duct walls, the response of the plate under grazing incident waves shows various axial and lateral vibration modes which consist of a combination of *in vacuo* modes in both directions depending on the excitation frequency and plate stiffness. Although the transmission loss depends very much on the magnitudes of the induced plate vibration, the spectral peaks in the plate vibration velocity, e.g., resonances, do not necessarily correspond to the peaks in the spectrum of the transmission loss. The latter depends more crucially on the acoustic interference between the wave reflections from various vibration modes. When the plate is simply supported at the leading and trailing edges and is free at the two lateral edges, the performance of the plate silencer is much better than when all edges are fixed.

(2) The first *in vacuo* plate vibration mode is effective in reflecting sound but is difficult to excite due to cavity stiffness. The second *in vacuo* mode is easy to excite but its capability to cause sound reflection is significant only when there is significant phase difference of sound over the length of the plate. The acoustic interference between the odd and even *in vacuo* modes is crucial to achieving a broad stopband. The mechanics of the plate bending differs from that of a tensioned membrane, leading to a different pattern of acoustic interference in favor of the plate silencer towards the ultralow frequencies like $f_1=0.03$.

(3) The effect of plate mass is generally counterproductive, but zero mass is not quite the optimal configuration, which differs somewhat from the drumlike silencer. The effect of the plate damping is marginally beneficial for the bandwidth, so is the use of sound absorption material in the covered cavity. However, the contribution of sound absorption in the overall noise abatement is higher in the latter case. This result can be easily understood since a highly damped plate generally responds less to the incoming sound. In other words, the effect of damping on the reduction of sound reflection is more serious when the damping is embedded in the plate motion than if it is distributed in the cavity.

(4) The optimal parameters predicted for sound reflection in air require a plate material which is very light and yet extremely stiff. When aluminum is considered, it is found that the Young's modulus E is at least 250 times too low. The critical parameter is E/ρ_s^3, for which nonmetallic materials like PVC foam can easily satisfy the requirement. In addition, composite beam structure can be used to reduce the crucial structural mass while achieving a fairly high structural stiffness.

ACKNOWLEDGMENT

The research reported here is supported by grants from the Research Grants Council of the Hong Kong SAR (PolyU 5169/02E, 5298/03E).

Ashby, M. F. (**1999**). *Material Selection in Mechanical Design*, 2nd ed. (Butterworth Heinemann, Oxford).

Ashby, M. F., and Johnson, K. (**2002**). *Material and Design: the Art and Science of Material Selection in Product Design* (Butterworth Heinemann, Oxford).

Brown, S. (**1964**). "Acoustic design of broadcasting studios," J. Sound Vib. **1**, 239–257.

Choy, Y. S., and Huang, L. (**2002**). "Experimental studies of a drumlike silencer," J. Acoust. Soc. Am. **112**, 2026–2035.

Choy, Y. S., and Huang, L. (**2005**). "Effect of flow on the drumlike silencer," J. Acoust. Soc. Am. **118**(5), 3077–3085.

Doak, P. E. (**1973**). "Excitation, transmission, and radiation of sound from source distributions in hard-walled ducts of finite length. I. The effects of duct cross-section geometry and source distribution space-time pattern," J. Sound Vib. **31**, 1–72.

Ford, R. D., and McCormick, M. A. (**1969**). "Panel sound absorbers," J. Sound Vib. **10**, 411–423.
Frommhold, W., Fuchs, H. V., and Sheng, S. (**1994**). "Acoustic performance of membrane absorbers," J. Sound Vib. **170**, 621–636.
Gere, J. M. (**2004**). *Mechanics of Materials*, 6th ed. (Brooks/Cole, London).
Horoshenkov, K. V., and Sakagami, K. (**2001**). "A method to calculate the acoustic response of a thin, baffled, simply supported poroelastic plate," J. Acoust. Soc. Am. **110**, 904–917.
Huang, L. (**1999**). "A theory of passive duct noise control by flexible panels," J. Acoust. Soc. Am. **106**(4), 1801–1809.
Huang, L. (**2002**). "Modal analysis of a drumlike silencer," J. Acoust. Soc. Am. **112**, 2014–2025.
Huang, L. (**2004**). "Parametric studies of a drumlike silencer," J. Sound Vib. **269**, 467–488.
Huang, L., and Choy, Y. S. (**2005**). "Vibroacoustics of three-dimensional drum silencer," J. Acoust. Soc. Am. **118**(4), 2313–2320.
Mechel, F. P., and Vér, I. L. (**1992**) "Sound-absorbing material and sound absorbers," *Noise and Vibration Control Engineering: Principles and Applications*, edited by L. L. Beranek and I. L. Vér (Wiley, New York), Chap. 8.

Determination of the nonlinear parameter by propagating and modeling finite amplitude plane waves

F. Chavrier, C. Lafon, and A. Birer
Inserm, U556, Lyon, F-69003 France, Université Claude Bernard Lyon 1, Lyon, F-69000 France

C. Barrière and X. Jacob
Laboratoire Ondes et Acoustique, Université Paris 7-CNRS-ESPCI, Paris, F-75231 Cedex 05, France

D. Cathignol
Inserm, U556, Lyon, F-69003 France, Université Claude Bernard Lyon 1, Lyon, F-69000 France

(Received 11 October 2005; revised 3 March 2006; accepted 3 March 2006)

The acoustic nonlinear parameter, B/A, is an important piece of data whenever high intensity pressure fields are under consideration. In this work, an alternative method is proposed to measure this parameter. First, the method involves measuring the sound velocity and nonlinear waveform distortion of a finite amplitude plane wave propagating through a medium, Butanediol, whose density and attenuation law have been preliminarily determined. Measurements were performed in the nearfield of a piston where plane wave propagation regime exists. Impulse response of the hydrophone was determined and pressure waveforms were obtained by a convolution process. Then, the method involves modeling, in time domain and under experimental conditions, the theoretical nonlinear waveform distortion and fitting it to the experimental results by adjusting the B/A parameter. Comparative measurements were performed using the technique of parametric interaction. The respective results for the two methods ($B/A=11.0\pm10\%$ and $10.9\pm5\%$) are in a good agreement despite a smaller degree of accuracy for the proposed method. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2190161]

PACS number(s): 43.25.Ba, 43.80.Ev, 43.25.Zx [CCC] Pages: 2639–2644

I. INTRODUCTION

The acoustic nonlinear parameter, called B/A, is a commonly used term in the field of nonlinear acoustics. It is derived from the Taylor series expansion of the variations of the pressure in a medium in terms of variations of the density or alternatively of the variations of the sound velocity in terms of variations of the pressure. So, whenever high amplitude regime and/or highly nonlinear medium are involved, it is of major interest to know this parameter. This is especially true for medical applications. For example, when thermally ablating tissues using high intensity focused ultrasound, the nonlinear waveform distortion produces harmonics frequencies in tissues near the focus of the transducer. Consequently, the total attenuation of the wave increases and leads to enhanced local heating.[1–4] Similarly, in harmonic imaging, Verma *et al.* (Ref. 5) recently showed that a high degree of nonlinear distortion may occur during propagation through biological fluids and produce significant temperature rises in the application of three-dimensional (3D) ultrasound imaging in obstetrics. In the field of lithotripsy, knowledge of the nonlinear parameter of a medium or media through which the wave passes is indispensable for predicting the pressure waveform at the focus.[6] These nonexhaustive examples illustrate that nonlinear parameter must be known for accurate modeling and understanding of modern US medical applications. Obviously, the interest for the acoustic nonlinear parameter is not limited to biological tissues. For example, in other fields and applications such as nondestructive evaluation in industrial applications[7] or search for hydrocarbon reservoirs,[8] the nonlinear characteristics of materials were investigated.

Beyer[9] listed different methods described in the literature to determine B/A. He insisted on the fact that method accuracy was rarely better than 5%. The most common method is the thermodynamic method where the sound velocity is measured as a function of the pressure and the density. The finite amplitude method consists in measuring the amplitude of the second harmonic of a distorted plane monochromatic wave.[10] In the phase comparison method,[11] B/A is determined from the time shift of tone bursts due to adiabatic variation of the hydrostatic pressure. Sarvazyan[12] developed a specific method for aqueous solutions and Everbach[13] another one for immiscible mixtures. More recently, Barrière and Royer[14] presented a new method based on the parametric interaction of a high frequency acoustic wave with a low-frequency acoustic pulse. High-frequency acoustic wave gets modulated in phase when interacting with a lower frequency acoustic pulse and B/A is determined by detecting the phase modulation.

The goal of this study was to develop an alternative method for determining the nonlinear parameter B/A of a medium. First, the method involves measuring the sound velocity and the nonlinear waveform distortion of the pressure for a finite amplitude acoustic wave propagating through a medium whose density and attenuation law have been preliminarily determined. Then, the method involves modeling, in time domain and under experimental conditions, the theoretical nonlinear waveform distortion and fitting it to the

experimental result by adjusting the B/A parameter. Attenuation, nonlinearity, and diffraction must be taken into account for simulating the waveform distortion of a finite amplitude wave. As taking into account diffraction effects increases dramatically time of calculation, it was proposed to limit the model to finite amplitude plane waves. So, measurements were performed in the nearfield of a circular piston where plane wave propagation regime exists.

II. PRINCIPLE

Let us consider an acoustic plane wave propagating in the z direction and traveling in thermoviscous medium. Nonlinear distortion of the pressure waveform can be described by the Burgers equation,

$$\frac{\partial P(z,\tau)}{\partial z} = \frac{\beta}{\rho_0 c_0^3} P(z,\tau)\frac{\partial P(z,\tau)}{\partial \tau} + \frac{\delta}{2c_0^3}\frac{\partial^2 P(z,\tau)}{\partial \tau^2}, \tag{1}$$

where $P(z,\tau)$ is the sound pressure, $\beta=1+B/2A$ is the coefficient of nonlinearity, ρ_0 is the ambient density, c_0 is the sound speed, $\tau=t-z/c_0$ is the retarded time and δ is the dissipative coefficient of the medium. In many media such as biological tissues, this equation can be written in another form:[15]

$$\frac{\partial P(z,\tau)}{\partial z} = \frac{\beta}{\rho_0 c_0^3} P(z,\tau)\frac{\partial P(z,\tau)}{\partial \tau} + \int_{-\infty}^{\tau} g(z,\tau-\tau') \,.\, P(z,\tau')d\tau', \tag{2}$$

where the kernel, $g(z,\tau)$, can be evaluated from the frequency dependence of the attenuation $\alpha(\omega)$.[16]

Thus, the nonlinear propagation of a finite amplitude plane wave through a medium is determined by four intrinsic parameters: The density, ρ_0, the sound velocity, c_0, the frequency-dependent attenuation coefficient, $\alpha(\omega)$, and the acoustic nonlinear parameter, B/A. If ρ_0, c_0, and $\alpha(\omega)$ are known or measured elsewhere, B/A parameter can be obtained through matching experimental data to theoretical results.

Based on this observation, the principle of measuring the nonlinear parameter involves:

(1) Determining the parameters ρ_0 and $\alpha(\omega)$ of the medium being studied;
(2) measuring the distortion of a finite amplitude plane wave through a sample of medium with a calibrated hydrophone. Waveforms will be collected at the input and output of the sample;
(3) converting these electrical signals into pressure signals;
(4) determining, from these measurements, the parameter c_0 of the medium and of the water in experimental conditions;
(5) developing a numerical model to solve Eq. (2) and injecting the input pressure waveform and the parameters ρ_0, c_0, and $\alpha(\omega)$ into this model;
(6) varying the parameter B/A numerically until theoretical and experimental output pressure waveforms coincide as closely as possible.

III. NUMERICAL MODEL

The model implemented is a time-domain algorithm based on the principle of the fractional steps method, widely developed by Lee,[17] Tavakkoli,[18] and Khokhlova.[19] In summary, Eq. (2) is separated into two equations, Eqs. (3) and (4), and these equations are solved independently over each incremental step Δz.

$$\left.\frac{\partial P(z,\tau)}{\partial z}\right|_{NL} = \frac{\beta}{\rho_0 c_0^3} P(z,\tau)\frac{\partial P(z,\tau)}{\partial \tau}, \tag{3}$$

$$\left.\frac{\partial P(z,\tau)}{\partial z}\right|_{Att} = \int_{-\infty}^{\tau} g(z,\tau-\tau') \,.\, P(z,\tau')d\tau'. \tag{4}$$

Because Eq. (4) is a convolution sum, $P(z,\tau)$ at step z (index k) is sampled by a uniform time interval $\Delta\tau$ and labeled P_i^k at time $\tau_i=i\Delta\tau (i=0,\ldots,N-1)$.

Nonlinearity: The evolution equation for a plane wave in a nonlinear medium, Eq. (3), has an analytical solution, called the Poisson's solution, Eq. (5). To avoid multivalued waveform solutions, Δz must satisfy the condition given by Eq. (6)

$$P(z+\Delta z,\tau) = P(z,\tau+\Delta\tau_{NL}) \text{ with } \Delta\tau_{NL} = \frac{\beta\Delta z}{\rho_0 c_0^3} P(z,\tau), \tag{5}$$

$$\Delta z < \frac{\rho_0 c_0^3}{\beta \max(\partial P/\partial\tau)}. \tag{6}$$

Numerically, using the waveform samples P_i^k, an adapted Δz is first calculated according to Eq. (6). Then, for each time index, i, $\Delta\tau_{NL}$ is calculated and P_i^{k+1} is determined using 2nd degree polynomial interpolations.

Attenuation: In biological tissues, the attenuation effect, Eq. (4), is frequency-dependent and can be written in the general form:

$$G(\Delta z,\omega) = \exp\{-\alpha(\omega) \,.\, \Delta z + i\omega[c(\omega)^{-1} - c_0^{-1}] \,.\, \Delta z\}, \tag{7}$$

where $c(\omega)$ and $\alpha(\omega)$ are, respectively, the frequency-dependent sound velocity and attenuation coefficient of the medium. If the $\alpha(\omega)$ law is known, the Kramers-Kronig relationship between attenuation and phase velocity (i.e. dispersion) makes it possible to calculate $G(\Delta z,\omega)$ as a minimum-phase digital filter.[18,20] In the time domain, Eq. (4) is numerically implemented as a convolution sum, Eq. (8), where the impulse response g is the inverse Fourier transform of $G(\Delta z,\omega)$.

$$P_i^{k+1} = \sum_{j=0}^{+\infty} P_j^k . g_{i-j}^k \,.\, \Delta\tau. \tag{8}$$

Propagation: For precise synchronization of theoretical results with measurements, the absolute time $t_i^k=t_0^k+i\Delta t$ with $\Delta t=\Delta\tau$ must be determined for each sample P_i^k. Let us consider that $\Delta z_{\text{water}}^k$ and $\Delta z_{\text{medium}}^k$ are, at step z, the total distances of propagation through water and medium, the absolute time t_0^k is written,

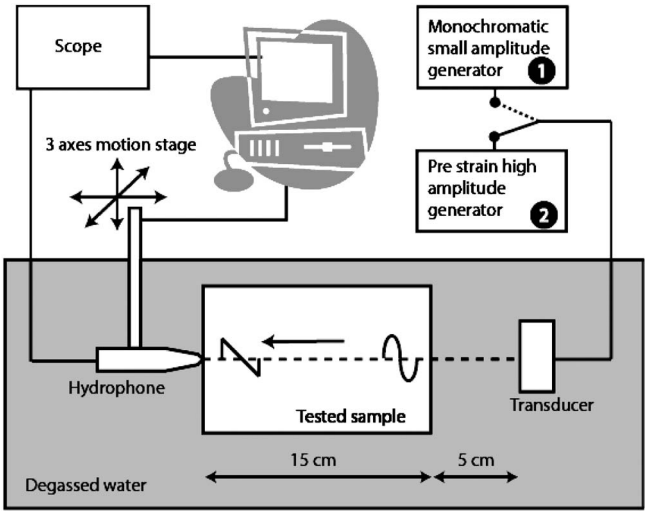


FIG. 1. Experimental setup. To find the acoustic axis, the transducer was excited monochromatically (position 1). To collect signals S1–S5, the transducer was excited with pulses (position 2).

$$t_0^k = t_0^0 + \frac{\Delta z^k_{\text{water}}}{c_{0\ \text{water}}} + \frac{\Delta z^k_{\text{medium}}}{c_{0\ \text{medium}}}. \tag{9}$$

Interface transmission: Finally, because the plane wave passes experimentally through several media in succession (sample of medium immersed in water), reflections at the interfaces should be taken into account in the model. If Z_w is the impedance of water and Z_m the impedance of the medium, the amplitudes of the pressure transmitted at the water-medium and medium-water interfaces can be written, respectively,

$$P_m = \frac{2Z_m}{Z_m + Z_w} P_w \quad \text{and} \quad P_w = \frac{2Z_w}{Z_m + Z_w} P_m. \tag{10}$$

IV. MATERIALS AND METHODS

A. Experimental setup

High-pressure pulses were generated at the surface of a piezo-composite transducer using electrical pre-strain. This method and the electronic driving system which enables more pressure to be generated from a piezo-composite transducer, have been described in detail elsewhere.[21]

The experiments were performed with a 50 mm diameter and 720 kHz central frequency piezo-composite planar transducer (Imasonic, Besançon, France). The transducer was air-backed and had a quarter wavelength matching layer. It was positioned in a tank filled with degassed water at ambient temperature (Fig. 1). The material being studied was conditioned in a sample-holder. A stand joined to the transducer enabled the sample-holder to be centered on the acoustic axis and its front to be positioned 50 mm away from the transducer. At this distance, the wave propagating through the front face of the sample is plane over a long temporal window. Furthermore, no electro-acoustical coupling prevents measuring pressure. The sample-holder was a 150 mm long cylinder, 90 mm in diameter. A polyethylene membrane was stretched over both ends of the cylinder. This extremely fine membrane (25 μm) was nevertheless sufficiently stiff to ensure that the faces of the sample were parallel to the transducer. A preliminary experiment was performed with a view to testing this membrane's transparency to ultrasound. The transducer was excited with a high-intensity pulse (nonlinear regime) and the hydrophone was placed at 280 mm from the transducer. Two signals were recorded when the sample-holder, filled with degassed water, was placed on its stand and then when it was withdrawn from it. The estimated frequency-dependent transmission loss increases from less then 0.04 dB at 1 MHz to 0.4 dB at 10 MHz and 1.5 dB at 20 MHz. These results show that transmission losses are low on this frequency band. A "Golden lipstick 0.2 mm" hydrophone (Onda, Sunnyvale, CA) was moved by a computer-controlled three-axis drive unit (Newport Corporation, Irvine, CA). The output signal from the hydrophones was collected using a Waverunner LT264 oscilloscope (Lecroy, Chesnut Ridge, NY).

Measurements were made on the acoustic axis, where the time window corresponding to the plane wave (preceding the arrival of the "edge wave") is maximal. According to the length and the position of the sample and according to the diameter of the transducer, the usable time window covered the first arch of the signal ($\approx$1 μs). It should be noted that the excitation system generates waveforms which present maximal nonlinear distortion in the first arch (where amplitude and slope were maximal). Moreover, the sensitivity of measurements increases with the propagation distance through the medium. So, this configuration is the good trade off to maximize the sensitivity of measurements (see Sec. IV).

B. Methodology of measurements

During setup, a special phase of the control program was used to accurately determine the acoustic axis by means of a series of measurements made with low amplitude monochromatic signals.

In order to determine the velocities of sound in experimental conditions, the transducer was switched to the pulse generator, whose discharge voltage was set to 117 Volts. The maximum pressure amplitude at the surface of the transducer was then 0.04 MPa (linear regime). With no sample, signals S1 and S2 were recorded when the hydrophone was placed at 50 and 201 mm from the transducer, respectively. With sample, signal S3 was recorded when the hydrophone was placed at 201 mm from the transducer. Measurements S1, S2, and S3 were the result of an average over 5 pulses.

To measure nonlinear distortion, the discharge voltage of the generator was then set to 3850 Volts. In these conditions the maximum pressure amplitude at the surface of the transducer was around 2 MPa (nonlinear regime). With the sample still in place, signal S4 was recorded when the hydrophone was placed at 201 mm from the transducer. Without sample, signal S5 was recorded when the hydrophone was placed at 50 mm from the transducer. At this level of excitation, the measurements were very reproducible. Averaging was, therefore, dispensed in order to protect the hydrophone. The choice of this sequence makes it possible to con-

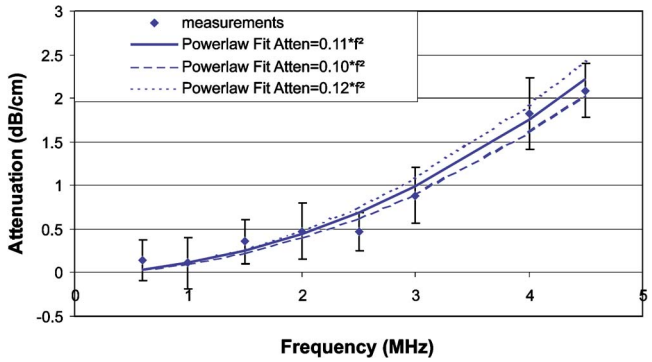


FIG. 2. Attenuation of 1.3 Butanediol over the band 0.5–4.5 MHz. Estimation of the attenuation law.

serve the alignment by keeping sample manipulations to a minimum and keeping the same excitation levels for the measurements in water and in the sample.

Finally, to be compatible with the numerical model, the electrical signals, S4 and S5, must be converted into pressure signals, P4 and P5. Given the spectrum of nonlinear signals, convolution sum was necessary, as $P(t)=S(t)*h(t)$ where $P(t)$ is the investigated pressure signal and $h(t)$ the pulse response from the hydrophone. To ensure that this conversion operation is carried out accurately, the pulse response from the hydrophone was measured using the "thick transducer" method.[22,23] The principle behind this method is to excite a transducer (50 mm in diameter and 20 mm thick) whose physical parameters are accurately known by a step of voltage (rise time=12 ns). As the transducer is sufficiently wide and thick, it is possible to separate the time taken by the response coming from the front of the transducer from those coming from its rear and its edges. In this way, by placing the sensor at 6 mm from the transducer, we can get its frequency response (amplitude and phase) and then, by inverse Fourier transform, its impulse response (4 μs). It should be noted that the characteristics of our calibration bench limit measurement to the frequency range situated between 0.25 and 20 MHz.

V. RESULTS

The material studied was 1.3—Butanediol (99%, Sigma-Aldrich Chimie, Saint-Quentin Fallavier, France) proposed as a tissue phantom for shockwave propagation.[24] Its density, 1005 kg/m^3, was found in manufacturer's data sheets. Water density, 1000 kg/m^3, and attenuation law, 0.0022 dB/cm/MHz^2, were available from the literature. Using a specific bench, the attenuation law of Butanediol was estimated from the series of attenuation measurements made in transmission mode on the frequency band 0.5–4.5 MHz (Fig. 2). Given the equipment used, the maximum error on these measurements was calculated as ±16%. The powerlaw $\alpha(f)=\alpha_0 f^2$ with $\alpha_0=0.11$ dB/cm/MHz^2 was found to be the best fit for experimental results. Considering the error bars on the measurements, α_0 ranges between 0.10 and 0.12 dB/cm/MHz^2 resulting in a ±10% accuracy. These results are in a good agreement with Andreev's data[25] measured in a 3–25 MHz frequency range.

Before the experiment, the Butanediol was degassed and the sample-holder immersed in the tank until the temperatures stabilized ($T=25$ °C). According to the described methodology, the electrical signals S1 to S5 were collected at the output of the hydrophone (Fig. 3) and recorded (time and amplitude).

From signals in linear regime, velocities, $c_{0\text{water}}=1496.93$ m.s^{-1} and $c_{0\text{but}}=1530.21$ m.s^{-1}, were, respectively, measured from the time shift (T_s) of the first arch of signals S1 and S2 ($T_s=100.873$ μs through 151 mm of water) and S1 and S3 ($T_s=98.694$ μs through 150 mm of Butanediol plus 1 mm of water).

Electrical signals S4 and S5 were converted into pressure signals P4 and P5. Then, signal P5 was injected into the model (time step=1 ns, $\alpha_{\text{but}}=0.11$ dB/cm/MHz^2, propagation distance=150 mm through Butanediol+1 mm through water). In order to approach the B/A value, simulations were made for B/A varying between 5.0 and 13.0 with increment $\Delta B/A=1.0$. The measured pressure, P4, and the simulated pressures $\{P4_S:\Delta B/A=1.0\}$ are shown in Fig. 4(a). The set of simulated pressures show that, for a given input signal, the

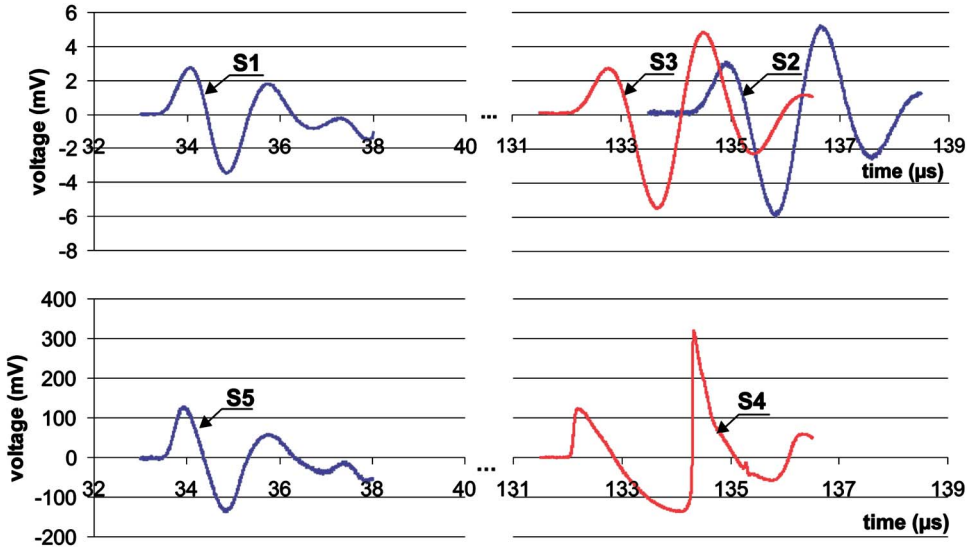


FIG. 3. Signals S1, S2, and S3 measured under linear regime and S4 and S5 measured under nonlinear regime. Time resolution = 1 ns.

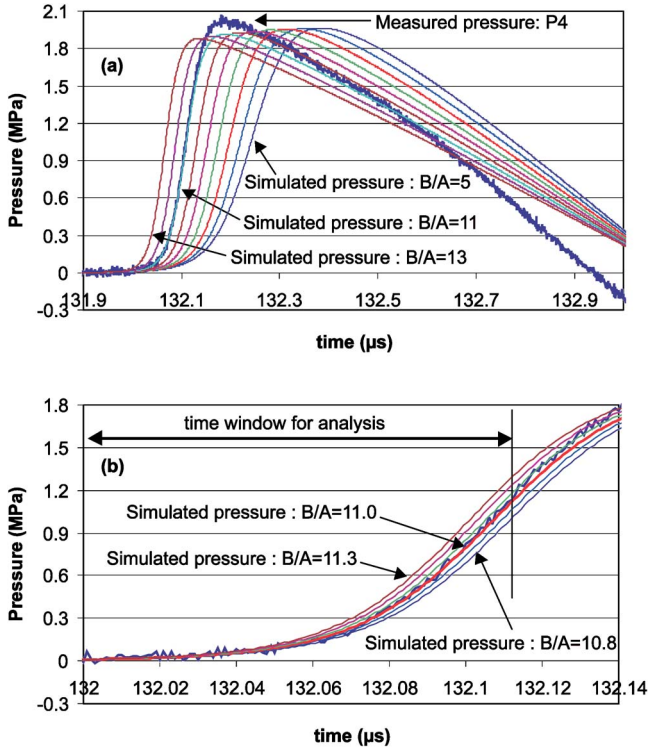


FIG. 4. Experimental response, P4, and theoretical responses: (a) For B/A varying between 5 and 13 with an increment of 1.0, (b) for B/A varying between 10.8 and 11.3 with an increment of 0.1. Time step = 1 ns, $\alpha(f) = 0.11\,f^2$, propagation distance = 150 mm through Butanediol +1 mm through water.

amplitude, the slope and the position of the first leading edge depend of the B/A parameter. The accuracy of the calibration of hydrophone, the uncertainty of the attenuation law and/or the nonuniformity of pressure distribution (amplitude and phase) on the edges of the transducer may be the cause of the discrepancy on amplitude between measured and simulated results. Furthermore, the slope of the first leading edge also varies with the α_0 value. Finally, for a given input signal, the position of the first leading edge appears as the most significant factor. Indeed, it principally depends of the velocity (accurately determined from linear signals) and of the B/A parameter. Moreover, the position of the first edge is much more sensitive to B/A than amplitude and slope. According to these observations, the time window for analysis was focused to a smaller area corresponding to the first part of the leading edge and more discriminating simulations were made for B/A varying between 10.8 and 11.3 with increment $\Delta B/A = 0.1$. The measured pressure, P4, and the simulated pressures $\{P4_S : \Delta B/A = 0.1\}$ are shown in Fig. 4(b). The best correlation coefficient (99.7958%) was obtained for the value $B/A = 11.0$.

VI. DISCUSSION

The nonlinearity parameter measured in the butanediol, $B/A = 11.0$, is different from the value given by Granz ($B/A = 7.3$, estimated according to Ballou's rule). In order to validate our measurement method, a sample of Butanediol was sent to the *Laboratoire Ondes et Acoustique (Paris, France)* where the nonlinearity parameter was measured using the nonlinear interaction of two waves with a high-frequency ratio and generated by planar transducers. The measurement principle and the experimental setup are described in detail by Barrière[14] and Jacob.[7] Briefly, the nonlinear collinear interaction of a 10 MHz tone burst with a 1 MHz central frequency pulse leads to a phase modulation of the highest frequency wave. The nonlinearity parameter is deduced from the phase modulation measurement and from the calibration of the low-frequency transducer with an optical interferometer. Several measurements were taken at 25 °C, for three interaction distances (10, 15, and 20 mm) and three different central frequencies of the lower frequency pulse (0.25, 0.5, and 1 MHz). The average value deduced from these experiments is $B/A = 10.9 \pm 0.6$. This means that the two methods concur.

Obviously, the theoretical model used in this work is valid for determining the nonlinear parameter B/A of liquids and biological tissues that satisfy both equations $B/A = 2(\beta - 1)$ and $\alpha(f) = \alpha_0 f^p (1 \leqslant p \leqslant 2)$. However, the proposed method remains suitable for determining β of many solids.

The sensitivity of the proposed method can be discussed. First, the results (Fig. 4) show that, for a finite amplitude plane wave propagating through a medium, the B/A parameter is directly linked to the position of the first leading edge. Consequently, the fit process can be focused on this part of the measured signal in order to increase the sensitivity. Additionally, simulations carried out with different propagation distances, show that the sensitivity increases with the length of the sample. For example, if the input pressure amplitude is 2 MPa and if the sample length is 80 mm, the simulations for B/A varying between 10.8 and 11.3 with an increment of 0.1 are nearly superposed. In these conditions, the determination of B/A with a decimal accuracy requires increasing the pressure at the surface of the transducer. Maximizing the sample length is preferable for protecting the hydrophone. Finally, the sensitivity of the method depends of several factors such as the dynamic range of the excitation system, the transducer diameter and/or the sample dimensions.

Concerning measurements, it should be noted that the accuracy of the proposed method is dependent upon the accuracy of the pressure measurements (amplitude and time) and the estimation of the attenuation law. First, the method turns out to be very sensitive as far as pressure injected into the model is concerned. A variation of ±2% in the pressure injected leads to a variation of ±3% in evaluating parameter B/A. In these conditions, a perfectly calibrated hydrophone must be used. Generally, hydrophone calibration procedures provide, in terms of frequency, a limited range of sensitivities (between 0.25 and 20 MHz for our calibration procedure). To ensure the accuracy of pressure measurements, the transducer was chosen with a low central frequency (720 kHz) and the excitation level was adapted so that the maximal frequency of the first leading edge remains below 20 MHz. Thus, the accuracy of pressure measurements can be estimated at ±5% (accuracy of our hydrophone calibration procedure). Concerning time measurements, it should be noted that the time shift of the synchronization signal (±5 ns) could lead to a variation of ±3% in evaluating parameter B/A.

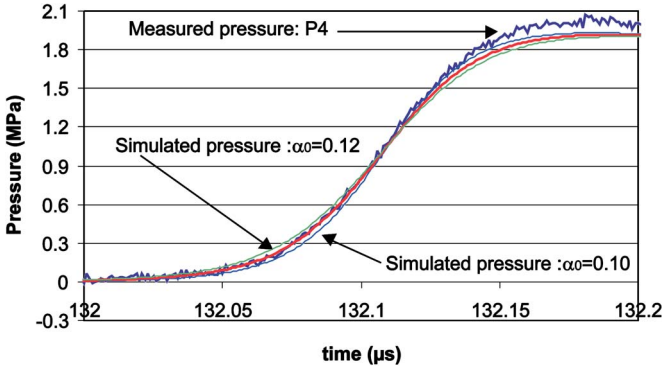


FIG. 5. Experimental response and theoretical responses for α_0 varying between 0.10 and 0.12 dB/cm/MHz^2 by increments of 0.01 dB/cm/MHz^2. Time step = 1 ns, $\alpha(f) = \alpha_0 f^2$, propagation distance = 150 mm through Butanediol + 1 mm through water, $B/A = 11.0$.

Another source of uncertainty is related to estimating the attenuation law for the material being studied. For Butanediol, a quadratic law was measured in agreement with data of Andreev.[15] However, with our bench, the dispersion of measurements in transmission on a restricted frequency range (Fig. 2) does not allow us to estimate α_0 with any great accuracy (±10%). Figure 5 shows simulations made for different attenuation laws, $\alpha(f) = \alpha_0 f^2$, when α_0 varies between 0.10 and 0.12 dB/cm/MHz^2 by increments of 0.01 dB/cm/MHz^2 (time step = 1 ns, propagation distance = 150 mm through Butanediol + 1 mm through water, $B/A = 11.0$). These results confirm that the amplitude and the slope of the first leading edge depend of α_0. So, by fitting the slopes, it seems possible to approach α_0 more accurately. Here, it clearly appears that simulation with $\alpha_0 = 0.11$ dB/cm/MHz^2 provides the best fit of measured pressure. Finally, in this work, the poor accuracy of the method used to measure the attenuation law does not really affect the accuracy of B/A measurement. Unfortunately, this conclusion is valid if the exponent of the powerlaw of the attenuation is known with accuracy. Indeed, if α_0 and p are unknown, different couples of values (α_0, p) may provide a good fit of measured pressure and consequently the accuracy of B/A measurement decreases. So, the necessity to accurately know the attenuation law $\alpha(f) = \alpha_0 f^p$ over a broad frequency range remains a constraint of the proposed method in comparison with other methods that require only few values of attenuation at specific frequencies.

In conclusion, the proposed method turns out to be relatively simple to use, from both the theoretical and experimental standpoints. It makes it possible to obtain reasonable results with a degree of accuracy (of ±10% in our experimental conditions) which is, to a great extent, correlated with the accuracy of the measurement of the pressure (amplitude and time) and of the medium's attenuation law.

[1]V. A. Khokhlova and O. A. Sapozhnikov, "Nonlinear effects in HIFU propagation and attenuation in biological tissue," J. Acoust. Soc. Am. **102**(5), 3155 (1997).
[2]R. L. Clarke and G. R. ter Haar, "Production of harmonics in vitro by high-intensity ultrasound," Ultrasound Med. Biol. **25**(9), 1417–1424 (1999).
[3]P. M. Meaney, M. D. Cahill, and G. R. ter Haar, "The intensity dependence of lesion position shift during focused ultrasound surgery," Ultrasound Med. Biol. **26**(3), 441–450 (2000).
[4]E. A. Filonenko and V. A. Khokhlova, "Effects of acoustic nonlinearity on heating of biological tissue by High-Intensity Focused Ultrasound," Acoust. Phys. **47**(4), 468–475 (2001).
[5]P. K. Verma, V. F. Humphrey, and F. A. Duck, "Broadband attenuation and nonlinear propagation in biological fluids: An experimental facility and measurements," Ultrasound Med. Biol. **31**(12), 1723–1733 (2005).
[6]M. A. Averkiou and R. O. Cleveland, "Modeling of an electrohydraulic lithotripter with the KZK equation," J. Acoust. Soc. Am. **106**(1), 102–112 (1999).
[7]X. Jacob, C. Barriere, and D. Royer, "Acoustic non-linearity parameter measurements in solids using the collinear mixing of elastic waves," Appl. Phys. Lett. **82**(6), 886–888 (2003).
[8]R. M. D'Angelo, K. W. Winkler, T. J. Plona, B. J. Landsberger, and D. L. Johnson, "Test of hyperelasticity in highly nonlinear solids: Sedimentary rocks," Phys. Rev. Lett. **93**(21), id. 214301 (2004).
[9]R. T. Beyer, in *The parameter B/A. In nonlinear acoustics.*, edited by M. F. Hamilton and D. T. Blackstock (Academic Press, New York, 1998), pp. 25–39.
[10]L. Adler and E. A. Hiedemann, "Determination of the nonlinearity parameter B/A for water and m-Xylene," J. Acoust. Soc. Am. **34**, 410–412 (1962).
[11]S. Zhu, M. S. Roos, W. N. Cobb, and K. Jensen, "Determination of the acoustic nonlinearity parameter B/A from phase measurements," J. Acoust. Soc. Am. **74**, 1518–1521 (1983).
[12]A. P. Sarvazyan, T. V. Chalikian, and F. Dunn, "Acoustic nonlinearity parameter B/A of aqueous solutions of some amino acids and proteins," J. Acoust. Soc. Am. **88**, 1555–1561 (1990).
[13]E. C. Everbach, Z. Zhu, P. Jiang, B. T. Chu, and R. E. Apfler, "A corrected mixture law for B/A," J. Acoust. Soc. Am. **89**, 446–447 (1990).
[14]C. Barriere and D. Royer, "Diffraction effects in the parametric interaction of acoustic waves: application to measurements of the nonlinearity parameter B/A in liquids," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(6), 1706–1715 (2001).
[15]V. G. Andreev, V. A. Khokhlova, O. V. Rudenko, and O. A. Sapozhnikov, "Suppression of nonlinear damping of a sound wave in a medium which contains a resonant absorber with a finite line width," Moscow Univ. Phys. Bull. **3**, 67–71 (1985).
[16]M. O'Donnell, E. T. Jaynes, and J. G. Miller, "Kramers-Konig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701 (1981).
[17]Y. S. Lee and M. F. Hamilton, "Time domain modeling of pulsed-amplitude sound beams," J. Acoust. Soc. Am. **97**(2), 906–917 (1995).
[18]J. Tavakkoli, D. Cathignol, and R. Souchon, "Modeling of pulsed finite-amplitude focused sound beams in time domain," J. Acoust. Soc. Am. **104**(4), 2061–2072 (1998).
[19]V. A. Khokhlova, R. Souchon, J. Tavakkoli, O. A. Sapozhnikov, and D. Cathignol, "Numerical modeling of finite-amplitude sound beams: Shock formation in the near field of a cw plane piston source," J. Acoust. Soc. Am. **110**(1), 95–108 (2001).
[20]R. Kuc, "Modeling attenuation of soft tissue with a minimum-phase filter," Ultrason. Imaging **6**(1), 24–36 (1984).
[21]A. Birer, M. Ghohestani, and D. Cathignol, "Generation of higher pressure pulses at the surface of piezo-composite materials using electrical prestrain," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**(7), 879–886 (2004).
[22]F. Lakestani, "Generation of a stepped ultrasonic plane wave: application to piezoelectric transducer calibration," J. Phys. D **9**, 547–555 (1976).
[23]J. C. Baboux, in *Pulsed ultrasonic field of a thick transducer*, edited by Guildford (IPC Sciences and Technology Press, England, 1979), pp. 462–467.
[24]B. Granz, "Measurement of shock wave properties after the passage through a tissue mimicking material," Ultrasonics Symposium, Cannes, 1847–1851 (1994).
[25]V. G. Andreev, A. A. Karabutof, Y. A. Pischalnikov, and N. B. Podymova, "Wideband acoustic spectroscopy of liquid phantoms of biological tissues," 16th ICA and 135th ASA meeting, Seattle, edited by P. K. Kuhl and L. A. Crum (1998), pp. 2509–2510.

Acoustical origin of rainbands in an ideal tropical hurricane

S. A. Elder[a)]
Physics Department, United States Naval Academy, Annapolis, Maryland 21402

(Received 29 November 2004; revised 29 August 2005; accepted 13 January 2006)

An acoustical model of hurricane rainband formation is proposed. Self-excited acoustic resonance inside the spinning eye produces strong azimuthal and radial standing waves. These modes can cause large temperature fluctuations about the dew point which are believed to be responsible for observed dry/wet bands in the outer storm disk. Distortion of standing wavefronts into spirals occurs in the transition layer region between approximately geostrophic flow inside the eyewall and irrotational flow outside. Tonal emissions from hurricane eyes should be detectable at great distances and are likely to be in a range below 0.01 Hz. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2181087]

PACS number(s): 43.25.Ts, 43.25.Cb, 43.28.Ra [RR] Pages: 2645–2650

I. INTRODUCTION

An ideal tropical cyclone is characterized by a clear/unclouded and elegantly rounded eye cavity, from which issue alternating spirals of rainy and dry bands. In this treatment we consider only counter-clockwise (CCW) rotating storms, the type found in the northern hemisphere. In order to seek a physical understanding of possible acoustical involvement of this natural phenomenon, it is helpful to visualize the underlying steady flow, upon which the acoustical field is superposed, in terms of a simple model. Emanuel[1,2] has derived an approximate formula for the maximum theoretical azimuthal velocity of a tropical hurricane, based on thermodynamic quantities such as the amount of enthalpy available from the warm moist air close to the ocean surface. At the eyewall, which occurs at a slightly smaller radius than that defined by this limiting speed, the steady flow is approximately geostrophic. Thus the eye is in essentially rigid-body rotation, important for understanding the acoustics of the problem. The reasoning for this assumption is that both viscous and inertial forces in the Navier-Stokes equation are likely to be small compared to the Coriolis force in the eye.

That is, in the eye, the steady-state Navier-Stokes equation reduces to

$$\frac{1}{\rho}\nabla p \cong -2\boldsymbol{\Omega}\times\mathbf{u} \tag{1}$$

where p is the pressure, ρ is the density of the air, $\boldsymbol{\Omega}$ is the angular speed of rotation of the earth, and $\mathbf{u}$ is the velocity of the fluid inside the eye cylinder.

Dropping the inertial term near the center is justified by the fact that this term varies as the square of the azimuthal velocity and must therefore be negligible near $r=0$, where r is measured from the center of the slowly translating storm. Furthermore, any friction between the nearly circular rotating layers of fluid within the eye should tend to make all layers rotate at the same angular speed, removing the differential effects of viscosity. Therefore, within the eyewall, the Coriolis and pressure gradient forces dominate, with low Rossby and Ekman numbers[3] almost out to the radius of maximum winds. Beyond the eye uniform angular rotation can no longer be maintained, and the wind begins to slow down. In the extreme far field, on the other hand, conservation of angular momentum requires that the velocity profile must approach irrotational flow, since nearly all the vorticity is confined to the region in and immediately around the eye. This also simplifies acoustical considerations. We may think of the region between these two extremes as a sort of transition layer over which the necessary adjustments are made by both the steady and acoustical motions. It is the presence of this layer that distinguishes real hurricanes from ideal free vortex motion.

II. PRESSURE PROFILE IN THE EYE OF AN IDEAL HURRICANE

Geostrophic flow leads to a barometric pressure in the eye cavity which rises parabolically from a minimum at the center out to the eyewall as shown in Fig. 1. This gives a pressure gradient proportional to the Coriolis acceleration, which is itself in proportion to the radius. Beyond the eyewall, pressure should rise more slowly, since it is assumed to approach some ambient atmospheric pressure in the far field. Outside the concentrated vorticity within the eye cavity, the velocity field beyond becomes increasingly irrotational, eventually allowing the azimuthal velocity to fall off approximately as the distance from the center in the far field. Figure 1 gives a plausible barometric pressure profile for the entire lowered pressure region based on the above understanding. It is assumed that the eyewall cloud begins at the radius for which the effect of the inertial and viscous terms begin to be felt, ending the conditions for geostrophic flow.

III. IDEAL FREE VORTEX VELOCITY, VORTICITY, AND CIRCULATION PROFILES

Ideal free vortex profiles of azimuthal velocity, vorticity, and circulation are shown in Fig. 2 for comparison.[4] Since, within the eye of the free vortex, the fluid is assumed to be in solid-body rotation, a linear azimuthal velocity profile prevails out to the beginning of the eyewall. The ideal free vortex has uniform vorticity within the eye and zero outside.

[a)]Retired.

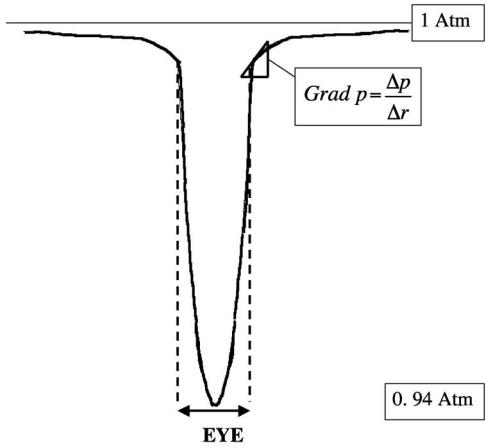


FIG. 1. Barometric pressure profile, $p(r)$, for ideal hurricane, showing radial gradient of barometric pressure at a particular radius. Dashed lines indicate diameter of the eyewall.

Circulation grows in proportion to the square of the radius, from zero at the center to a maximum at the eyewall. With all vorticity contained within the eye, the circulation must remain constant thereafter, providing pure irrotational motion.

IV. CUTAWAY VIEW OF HURRICANE OVER OCEAN

In Fig. 3(b) we see a closeup of the eye cavity region taken from the schematic hurricane cross section shown in Fig. 3(a). The CCW geostrophic motion inside the eye cavity

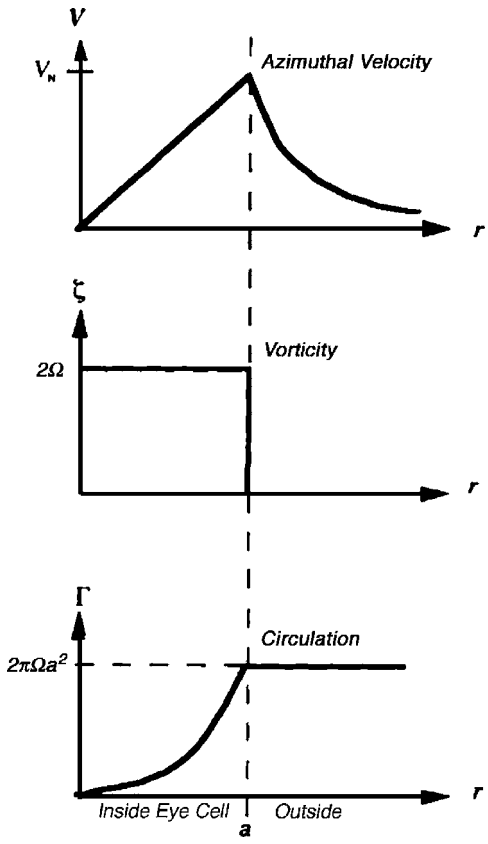


FIG. 2. Free vortex profiles are shown for comparison. Ideal velocity, vorticity, and circulation are plotted as a function of radius from the center. Adapted from Ref. 4.

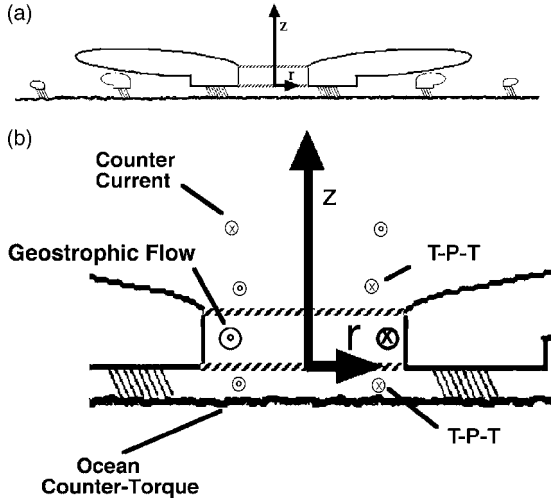


FIG. 3. Geostrophic flow in eye cavity. (a) Idealized hurricane cross section, showing a cylindrical acoustical resonator at the center, surrounded by the eyewall and outer rainbands. (b) Closeup of the eye cylinder, showing directions of steady currents. Approximate geostrophic flow extends above and below as well as in the cavity, illustrating the consequences of the Taylor-Proudman theorem (T-P-T).

is indicated by dot and "x" symbols. Since the Rossby number is small but not zero, some CCW flow occurs above and below the eye cylinder as well. Thus results the CCW twisting of warm moist air up from the ocean surface into the eye cylinder leading, in the steady state, to approximately geostrophic conditions within the eye cavity. That is, the regions of the cylindrical column immediately above and below the cloud layer also partake of this motion, consistent with the Taylor-Proudman theorem.[5] A diminishing azimuthal speed, accompanied by a counter torque, must occur in a boundary layer at the air-sea interface of the ocean, however, due to the inertia of the heavier fluid.

At the top of the eye cylinder, CCW rotation of the fluid weakens in the horn-shaped region defined by the outward sloping wall clouds above the hurricane. In order to balance lost angular momentum, a clockwise counter flow is generated above the top of the eye cavity by this effect. Along the vertical axis the active acoustical cavity is most effective from the point where the wall clouds begin to appear above the ocean and continues up to the point at which the slope of the pressure gradient changes abruptly.

V. CLOSE-UP OF PRESSURE PROFILE AT THE EYEWALL BOUNDARY

Figure 4 is a close-up that shows what a plausible pressure profile might look like at the radius where it begins to depart from the geostrophic conditions inside the eye. That is, the edge of the eyewall is the beginning of the transition regime between geostrophic flow and the outer region leading to irrotational flow. Beyond this radius the relative strength of inertial and drag forces are both greater than the Coriolis force. It is here that the rotating mass has to be able to settle down into a stable configuration consistent with a gentler pressure gradient, a process that tends to challenge the *de facto* fixed cylinder of the eye cavity. In the process of adjustment, spontaneous fluctuations—indicated by the

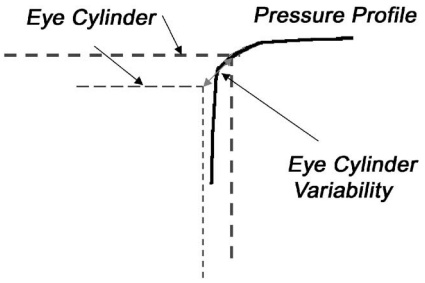


FIG. 4. Close-up of pressure profile $p(r)$ near the eyewall.

arrow—doubtless occur about the set point along the pressure profile on account of instability due to the high Reynolds number of the hurricane field. These cause the radius and height of the acoustically active eye cavity to vary slightly.

It is therefore reasonable to examine whether acoustic modes of the eye cylinder could be excited by instability of the set point at the eyewall. In the steady state, small deviations from the mean radius of the eyewall tend to be offset by the restoring forces associated with the balance between the effects of inertia and drag on the neutral radius.

Any fluctuation of the radius, however, would result in an acoustic pulse traveling from the eyewall both outward and inward, introducing a new phase lag into the oscillation. Reflections will cause the sound pulse to return to the eyewall, producing a second inward pulse. In a short time the circular boundaries will result in standing radial waves on account of the need to avoid a singularity at the center. If the wall has sufficient acoustic impedance mismatch, inward acoustic pulses will continue to be generated. Though the original pulse could either advance or retard the strength of the acoustic field inside the cavity, depending on the time elapsed in transit, there exist some special timings that reinforce sound energy density. Sooner or later the acoustic excitation will find these resonant modes and the oscillation frequency inside the cavity will stabilize, since timings that do not favor reinforcement will tend not to be repeated. Such fluctuations can be amplified by the extraction of energy from modulations of the steady wind corkscrewing up through the eye. Initially there is little to stop this positive feedback mechanism. On account of the very low linear damping at the expected infrasonic frequencies, the acoustic energy within the cavity will therefore tend to build up until it reaches an intensity limited by some nonlinear loss process. This behavior is typical of self-excited oscillators, which tend to lock in on a resonant frequency of the system when the gain exceeds some amount and the phase of the feedback is just right. Witness the screaming of sound amplification systems in a closed room when the gain is turned up too much!

The circular symmetry of the eye, together with the fact that rain clouds can have as much as 10% lower impedance[6] than dry air, makes it possible for some sound energy to be reflected at the wall and remain trapped in the cavity. In addition, known higher temperatures inside the eye[7] provide additional impedance mismatch at the boundaries, which should enhance the acoustic reflectivity of the cavity wall. Besides radial acoustic modes, azimuthal and vertical standing wave modes can be expected to couple in. Thus amplification of the acoustically resonant eye chamber by the pumping action of the disturbed winds in the eye will cause the internal sound level to rise until limited by nonlinear damping. Even though the Q of the cavity is probably small, very intense resonances can build up inside because nonlinear losses, being amplitude dependent, must maintain a strong internal sound level.

Since rain may occur at the altitude where clouds form, a likely way for the trapped sound energy to be dissipated could be through the formation of sound-induced rain clouds around the eye. At a height above the ocean where the ambient temperature is close to the dew point, sufficiently intense infrasonic field could induce evaporation and condensation by causing fluctuations of the local temperature. Such processes must take place outside the eye, since the higher temperature inside, previously noted, inhibits clouds from forming there.

Later we will show that the boundary with the eyewall cloud is where the typical rainband spirals that show up in the skirts of hurricanes begin to be formed, in conjunction with intense acoustic resonances in the eye. Such acoustical resonances may even be a source of stability for the hurricane, since a damped acoustic mode, associated with the set point, limits the variability of the operating point of the dynamical system.

VI. RESONANT MODES

Without going through a detailed calculation of the internal modal frequencies to be expected from a cylinder bordered by rain clouds, we can get an upper limit on expected frequencies associated with horizontal sound modes by analogy. For a hollow finite rigid cylinder the acoustic modal frequencies are given by[8]

$$f_{lmn} = \frac{c}{2\pi}\sqrt{k_l^2 + k_{mn}^2}, \tag{2}$$

where l, m, and n are the indices of the z, θ, and r modes, respectively, k_l is the wave number for modes along the vertical axis, and k_{mn} is the wave number for radial Bessel function and azimuthal modes in the horizontal plane. If $k_l=0$, the modes are determined by k_{mn} alone. Thus, for an eyewall of radius, a, one may take

$$f \approx f_{mn} = \frac{c(k_{mn}a)}{2\pi a} \tag{3}$$

for the modal frequencies of the $l=0$ modes.

The boundary conditions that cause downward reflection at the top of the cylinder are likely associated with the same conditions that cause cloud layers to have a top. The fact that there is a cloud ceiling shows that conditions that favor cloud formation at lower altitude must encounter an abrupt change at some height. Such a sudden change in medium must necessarily produce a corresponding change in acoustical impedance, so the problem is calculable for any location for which vertical properties of the atmosphere are known. One may reasonably conjecture that the cloud ceiling represents

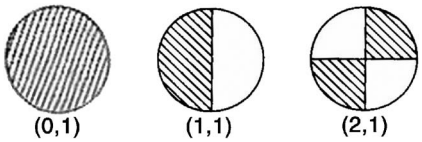


FIG. 5. The three lowest acoustic pressure modes and frequencies for a lightly mass-loaded cylindrical resonator, where c=speed of wave.

something more akin to a pressure-release condition than a rigid one. Likewise, the bottom of the cloud-layer probably also represents an approximate pressure release, though one cannot ignore the possibility of enough sound passing on to the ocean surface to produce interesting resonance conditions between sea and sky. For the purposes of this study we will ignore z-axis modes, except to say that they will have some affect on the resonant frequencies of the radiated sound for the case $l>0$. Regardless of the presence of vertical modes, horizontal modes can be compared with the case of a vibrating circular membrane. The three lowest resonant modes of membrane displacement with a rigid circular boundary[9] are shown in Fig. 5. By analogy, this may also be used for the acoustic case if we replace membrane displacement by acoustic pressure, since approximate pressure-release at the wall cloud is equivalent to approximate zero displacement at the nearly rigid boundary of a membrane.

Turning to the modal patterns again, let us assume that the shaded regions of a pattern are where the acoustic pressure, p, and therefore the local temperature variation, ΔT, are negative. Allan Pierce, in his excellent textbook[10] on acoustics, shows that sound pressure is a good surrogate for temperature fluctuation in the disturbed medium and supplies the following formula relating these variables:

$$\Delta \boldsymbol{T} = \left(\frac{\beta \boldsymbol{T}}{\rho \boldsymbol{c}_p} \right) p(\boldsymbol{r},\boldsymbol{t}), \tag{4}$$

where the constants β and c_p are, respectively, the thermal expansion coefficient and specific heat at constant pressure. This formula is accurate to first order in the variables $p(r,t)$ and ΔT.

A possible scenario is as follows: as a hurricane begins to form, a cloud begins to develop around the eyewall, providing the reflecting surface that allows the lowest frequency acoustic mode (0,1) to appear. This has motion in the r direction, but lacks motion in the θ direction. However, remembering that the eye is filled with vorticity, strange things can happen. If the radius of a vortex is increased, rotational motion is momentarily slowed down, which could excite the (1,1) mode. Furthermore, if axial modes are involved, there are more possibilities. If one stretches a vortex, its diameter shrinks and its rotation speeds up. If one compresses it, the diameter grows and the rotation slows down. This means that exciting even a pure axial sound wave produces r and θ fluctuations because they are all interrelated by the conservation of angular momentum. It may be that transverse inertia-gravity waves between temperature layers in the atmosphere, which some have associated with spiral rainbands,[11] provide just such a mechanism for pumping the axial mode. So once a rain cloud is formed, some internal modes will doubtless be excited. For the purpose of illustration, let us consider the (1,1) mode.

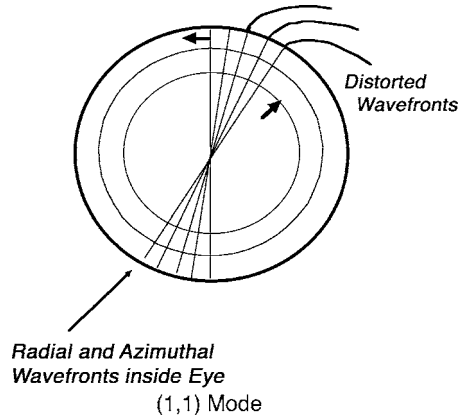


FIG. 6. Protrusion of the azimuthal standing sound waves into the transition layer region at the eyewall causes the standing wave-fronts to be distorted into spirals. Over one period, T, of the eye rotation, the starting point progresses a distance, $S=a\omega T=2\pi a$, along the circumference of the eyewall as the eye is rotated at angular speed ω. Multiple startup wavefronts are created as the modal structure, oscillating at a frequency $f=1/T_{\text{sound}}$, revolves about the center of the eye, with a separation of startups along the wall governed by the magnitudes of the relative periods of oscillation and revolution.

VII. OUTGOING WAVE-FRONTS FORMED IN THE EYEWALL

Consistent with the concept that the eyewall cloud is the beginning of a region where transition takes place between solid-body rotation inside and irrotational motion outside, it is likely that this is where the formation of rainbands begins to take place. Acoustically, the standing waves inside have azimuthal wave-fronts which are like spokes of a wheel that is slowly rotating, superposed on radial waves, as shown in Fig. 6. Outside the transition region wave-fronts are dominated by radial wave modes, as they approach the Hankel function shape of the outgoing far-field wave. The effect of decreasing angular speed at the eyewall causes distortion of the standing wave pattern into a spiral pattern within the transition layer, made visible by the formation of rain clouds.[12] These cloud bands formed in standing waves quickly become vestigial, being preserved in the irrotational flow as the invisible sound waves move on. Thus at the infrasonic wavefront formed just outside the active eye cylinder, there is momentarily produced a matching rain cloud structure which does not propagate with the outgoing sound. This must take place within the eyewall layer just outside the active eye cylinder.

Due to the continuing falloff of wind velocity with distance in the irrotational field outside the wall cloud, the outward sound rays will be refracted as they proceed away from the eyewall. The spiral formed near the eye will continue to be transported by the medium as it rotates. Any continuing rain in the spiral clouds away from the eye is probably of local origin from other causes, such as inertial gravity waves

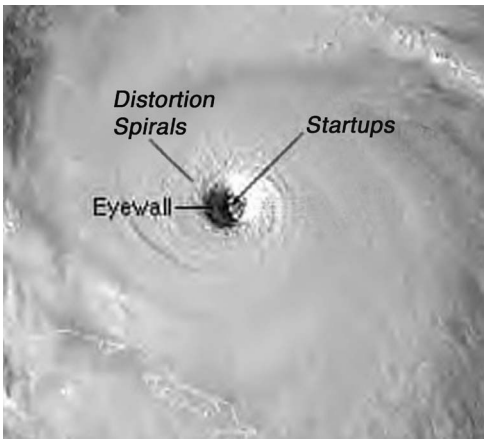


FIG. 7. Satellite view of startups along the eyewall of an actual hurricane, photographed and posted on the web by NOAA. Count of startups for a hurricane of given radius and azimuthal speed at the eyewall gives an estimate of the resonant mode frequency in the eye.

suggested by Abdullah[11] and others,[13] though some continuing sonic excitation might be possible from outgoing waves if the source is sufficiently intense.

VIII. PATTERN OF WAVE STARTUPS AROUND THE CYLINDER

The evidence of startup spirals may be observed in Fig. 7, which is an actual satellite photograph of an unknown hurricane in the open ocean, probably posted on the web by NOAA sometime in 2004. Notice the regular array of distortion spirals visible along the top of the storm cylinder, and the ribbonlike startup "wave-fronts" below them. Not all the hurricane pictures posted on the web are as suitable for observing these effects, since the eye is not always this clear. Recent hurricanes like Katrina, for example, were photographed crossing over islands in the Caribbean where the clarity of the eye was somewhat disturbed by local conditions. Nevertheless, in one of these pictures—Katrina NOAA-15—three or four startups may be seen evenly spaced along the eyewall on the right side.

Since typically the frequencies of the acoustic modes are an order of magnitude or more greater than the frequency of rotation of the eye, there will be a discrete sequence of many tilted wave-fronts formed in the transition layer for each complete rotation of the eye. It should be pointed out that, because of the solid-body rotation of the central region, the acoustic modal pattern itself must revolve at the angular speed of the eye. By measuring the spacing along the eyewall of these startups and knowing the diameter of the eye, one can estimate the number of wave startups, N_{mn}, in one revolution for a particular mode, f_{mn}. For the case pictured, incidentally, there are about 28 startups. The startup number can be equated to the ratio of acoustic-mode frequency to rotation frequency, giving the handy formula

$$N_{mn} = \frac{(\boldsymbol{k}_{mn}a/2\pi a)\boldsymbol{c}}{(V_H/2\pi a)} = \left(\frac{\boldsymbol{k}_{mn}a}{V_H}\right)c, \tag{5}$$

where $(k_{mn}a)$ determines the wave number for the (m,n) mode for the ideal pressure release boundary, a=radius of the eye, c=speed of sound, and V_H is the maximum hurricane speed at the eyewall. From this one may deduce that the frequency of the (m,n) mode in the cavity is given by

$$f_{mn} = \frac{(k_{mn}a)}{2\pi a}c = \frac{N_{mn}V_H}{2\pi a}. \tag{6}$$

So, from Eq. (5), one may find the value of $(k_{mn}a)$, guessing the values of m and n, and then use that to find the frequency of the acoustic resonance in the eye chamber with Eq. (6). The radius and speed of the hurricane pictured here are unknown. However, historical data on famous hurricanes are posted on the web by the National Hurricane Center (NHC). An example is Hurricane Luis in 1995. For that storm NHC reported a maximum speed of 138 mph and an eye radius of 23 miles. (For more precise comparison it would be helpful to ascertain the height at which the wind was measured.) If we assume a (1,1) mode for which the pressure release model gives $(k_{11}a)$=3.82, and taking c=777 mph, we find that

$$f_{11} = 0.95N_{11}\ \text{cycles/hour} = 0.000264\text{N}_{11}\ \text{Hz}. \tag{7}$$

N_{11} for this case has the value 21.5 [from Eq. (5)] corresponding to a frequency of about 0.006 Hz. The theoretical frequency, taking the end correction for radiation from a cylinder into account, would have been somewhat lower than this.

As one can see, the ideal model is still in a primitive stage, and more data from real hurricanes is needed to check out some of the assumptions. However, it suggests that very-low-frequency infrasound is probably present in hurricanes with well-formed eyes. A microphone with a response nearly down to dc would be required to detect such emissions. Actually a recording manometer would suffice at these frequencies. Since at low frequencies the absorption of sound is approximately proportional to the square of the frequency,[14] it is clear that infrasonic hurricane tones should be transmitted to great distances away from the source and therefore available to anyone, with the right equipment, who wishes to look for them.

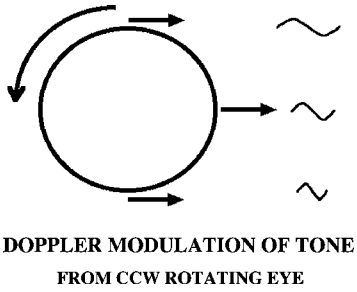


FIG. 8. Doppler effect on radiated sound produces a warble-tone about the modal frequency detected in the far field.

IX. DOPPLER COMPLICATIONS FROM THE ROTATING SOURCE

Figure 8 shows the effect of Doppler shifts on the moving sources. Where the sources are retreating from a distant observer, the wavelength is stretched, sending a lower-frequency tone, while for the approaching sources the fre-

quency will be raised. The sound coming from the nearest point of the cylinder should have a frequency midway between the two extremes and therefore represents the actual resonant value inside the cavity. Thus the sound received at a distance from the hurricane will be a sort of warble tone with pitches continually going up and down about the actual modal frequency inside.

Making use of high-resolution satellite photos such as those from the NOAA website it should be possible to compare the frequencies received via microphone with modal frequencies calculated from the spacing of the multiple sources around the eyewall.

ACKNOWLEDGMENT

The author is deeply indebted to Professor Wesley L. Nyborg, who first taught him the beauty of nonlinear acoustics and who has been very helpful in discussing the present paper.

[1]K. Emanuel, "Tropical Cyclones," Annu. Rev. Earth Planet Sci. **31**, 75–104 (2003). See p. 84, Eq. (8).

[2]Emanuel[1] on p. 81, states that some "air is slowly subsiding in the eyes, usually at velocities of $5-10\ \mathrm{cm\,s^{-1}}$." This important z axis feedback effect occurs at such small velocities compared to the azimuthal velocities in the hurricane that it may be neglected in considering the overall steady flow in the eye for acoustical purposes.

[3]For the benefit of nonspecialists, Rossby and Ekman numbers are nondimensional ratios of the inertial force and viscous force, respectively, to the coriolis force.

[4]Taken from S. A. Elder and J. Williams, *Fluid Physics for Oceanographers and Physicists*, 2nd ed. (Butterworth-Heinemann, Oxford, 1996), p. 219.

[5]For a physically intuitive and very helpful treatment of this subject, see D. J. Tritton, *Physical Fluid Dynamics* (Van Nostrand Reinhold Co., London, 1977), p. 163 ff.

[6]A report dealing with rain clouds, brought to the author's attention by Dr. Nyborg, is W. L. Nyborg and D. Mintzer, *Review of Sound Propagation in the Lower Atmosphere*, Brown University, May 1955, WADC TECHNICAL REPORT 54-602, Aero Medical Laboratory, Contract No. AF 33(616)-340, Project No. 7212, Wright Air Development Center. The report indicates that the velocity of sound with atmospheric water droplets can be lowered to as much as 90% of that for dry air.

[7]See Ref. 1, p. 81.

[8]See, for example, T. D. Rossing and N. H. Fletcher, *Principles of Vibration and Sound* (Springer-Verlag, New York, 1995), p. 155.

[9]Reference 8, pp. 69–70.

[10]A. D. Pierce, *Acoustics, An Introduction to its Physical Principles and Applications* (McGraw-Hill, New York, 1981), p. 17.

[11]A. J. Abdullah, "The spiral bands of a hurricane: a possible dynamic explanation," J. Atmos. Sci. **23**, 367–375 (1966).

[12]The profile in Fig. 3, left-cut, in Ref. 1 suggests that the eyewall begins at the point at which linear velocity profile begins to fall off so that maximum velocity occurs inside the eyewall. This is consistent with the fact that the eye radius is defined by the region of linear velocity profile, and that the distortion pattern begins immediately inside the eyewall.

[13]Emanuel[1] refers to several earlier papers on possible association of inertial-gravity waves with spiral bands, beginning with A. J. Abdullah, "Cyclogenesis by a purely mechanical process," J. Meteorol. **6**, 86–97 (1949).

[14]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 1982), p. 148, Eq. 7.36.

The radiation of atmospheric microbaroms by ocean waves

Roger Waxler[a] and Kenneth E. Gilbert
National Center for Physical Acoustics, University of Mississippi, University, Mississippi 38677

(Received 24 August 2005; revised 6 March 2006; accepted 7 March 2006)

A two-fluid model, air over seawater, is used to investigate the radiation of infrasound by ocean waves. The acoustic radiation which results from the motion of the air/water interface is known to be a nonlinear effect. The second-order nonlinear contribution to the acoustic radiation is computed and the statistical properties of the received microbarom signals are related to the statistical properties of the ocean wave system. The physical mechanisms and source strengths for radiation into the atmosphere and ocean are compared. The observed ratio of atmospheric to oceanic microbarom peak pressure levels (approximately 1 to 1000) is explained. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2191607]

PACS number(s): 43.28.Dm, 43.25.Lj, 43.28.Ra [MFH] Pages: 2651–2664

I. INTRODUCTION

Energetic systems of ocean waves, such as those produced by storms at sea, radiate detectable levels of infrasound into both the atmosphere and the ocean and, through the ocean, generate seismic waves at the sea floor which radiate into the earth.[1–6] The radiation is in a frequency band approximately 0.1 Hz wide centered at about 0.2 Hz. This radiation is referred to as microbaroms in the atmospheric case and as microseisms in the seismic case. Although the oceanic signals are sometimes referred to as microseisms, in this paper the terminology oceanic microbaroms will be used.

The microbarom and microseism peak is a permanent feature of the oceanic and seismic noise floors, respectively.[1,7,8] Spectra measured on the floor of the Pacific Ocean show peak levels up to 100 Pa/$\sqrt{\text{Hz}}$.[7,8] Atmospheric microbaroms are detected at large distances (up to thousands of kilometers) from the waves which produced them. Their propagation is highly dependent on the direction of the atmospheric winds.[9–12] Atmospheric microbarom spectra measured downwind from the wave system have typical peak levels of about 0.1 Pa/$\sqrt{\text{Hz}}$.[13] Note that oceanic microbarom signals are a thousand times greater than atmospheric signals.

In this paper a detailed calculation of the microbarom source strength is presented. It is shown that the physical mechanism for the radiation of microbaroms into the atmosphere is different from that for the radiation of microbaroms into the ocean. The predicted radiation into the atmosphere is three orders of magnitude less than the radiation into the ocean, in agreement with observation. Further, in order to relate sea states to observed microbarom levels, a direct connection is obtained between the stochastic models used to describe ocean waves[14,15] and the statistical properties of the received microbarom signal.

The problem of the generation of oceanic microbaroms was solved by Longuet-Higgens.[2] He showed that microbaroms and, through their interaction with the ocean floor, microseisms are produced through a second-order non-linear effect by the interaction of gravity waves of nearly equal frequency and nearly opposite propagation direction. He found that to obtain the source strength for the radiation of oceanic microbaroms it is sufficient to assume that the ocean is incompressible and that the atmosphere is a vacuum. In particular, once the ocean surface is in motion, the atmosphere's effect on the radiation of microbaroms into the ocean is negligible. Hasselmann[3] extended Longuet-Higgens' result to general sea states and made a direct connection between the stochastic models used to describe ocean waves[14,15] and the statistics of the observed microseism signals.

It was pointed out by Brekhovskikh *et al.* in Ref. 5 that the ocean's effect on the radiation of atmospheric microbaroms is significant. A large part of the atmospheric microbarom signal is due to sound radiated into the atmosphere from pressure fluctuations produced in the water by the motion of the ocean surface. It will be seen in this paper that the rest of the signal is due to the compression of the air by the motion of the ocean surface.

It follows that the study of atmospheric microbaroms requires a two-fluid model, consisting of a rare fluid, the atmosphere, over a dense fluid, the ocean. The interface between atmosphere and ocean must be allowed to deviate from its equilibrium position, assumed in this paper to be at $z=0$. For simplicity, in this paper the ocean will be assumed to be infinitely deep.

Of interest is the case in which there is a region of the ocean surface in which the sea state contains energetic counter-propagating waves. The acoustic signal radiated from this region to a distant sensor is considered. The mechanism through which the sea surface has been excited (presumably through strong winds produced by storms[14,15]) is not considered here. As in Refs. 3 and 5, the statistical properties of the microbarom signal are related to those of the sea state.

The paper is organized as follows. In Sec. II the notion of a source region, the region of the ocean surface from which the infrasound is radiated, is defined. The equations of fluid mechanics are then solved in the source region to sec-

[a]Electronic mail: rwax@olemiss.edu

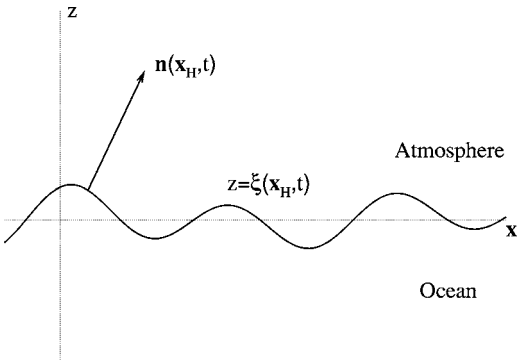


FIG. 1. A two-dimensional cross section of a portion of the source region. The air/water interface and an upwardly pointing normal vector to the interface are shown.

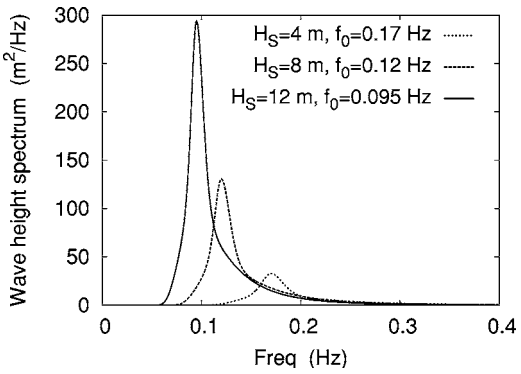


FIG. 2. The JONSWAP model for the frequency spectrum $\bar{F}(f) = \int_0^{2\pi} F(f,\theta)\, d\theta$ for significant wave heights H_S equal to 4, 8, and 12 m and peak frequencies f_0 equal to 0.17, 0.12, and 0.095 Hz, respectively.

ond order in the ratio of wave height to acoustic wavelength. In Sec. III the statistical properties of the received microbarom signal are related to those of the sea state. The source strength spectra for both atmospheric and oceanic microbaroms are obtained and compared. An explanation for the nearly three order of magnitude difference between the atmospheric and the oceanic signal strength is given. In Sec. IV an example is presented: a simple model is considered for the atmospheric microbarom signal received from storms over the deep ocean far from land masses. Section V contains our conclusions.

II. THE SOURCE REGION

Let S be a region of the ocean surface in which the sea is extremely active. Let z represent altitude relative to the undisturbed air/water interface at $z=0$. Assume that the surface of the water has been disturbed in such a way that the actual air/water interface is at $z=\xi(\mathbf{x}_H,t)$. Here, and throughout, $\mathbf{x}_H$ denotes the two-dimensional horizontal coordinate vector with components x and y. The function $\xi(\mathbf{x}_H,t)$ will be said to specify the sea state. A two-dimensional cross section of a portion of S is depicted in Fig. 1.

A. The sea state

Sea states are usually treated statistically[14,15] in the sense that $\xi(\mathbf{x}_H,t)$ is taken to be a stochastic process, commonly assumed to be Gaussian with mean zero. Let ξ_S be this Gaussian process and let $\langle\cdot\rangle_S$ be the expectation value of functions of ξ_S. If one writes

$$\xi_S(\mathbf{x}_H,t) = \mathrm{Re}\int \hat{\xi}_S(\mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{x}_H-\omega(\mathbf{k})t)}\, d^2k, \tag{1}$$

then, since it is Gaussian, the expectation value $\langle\cdot\rangle_S$ is completely specified by

$$\langle\hat{\xi}_S(\mathbf{k})\hat{\xi}_S(\mathbf{q})\rangle_S = \langle\hat{\xi}_S(\mathbf{k})^*\hat{\xi}_S(\mathbf{q})^*\rangle_S = 0$$

and

$$\langle\hat{\xi}_S(\mathbf{k})\hat{\xi}_S(\mathbf{q})^*\rangle_S = \mathcal{F}(\mathbf{k})\,\delta(\mathbf{k}-\mathbf{q}). \tag{2}$$

Here $\mathbf{k}$ in (1) is the two dimensional wave vector and $\mathcal{F}(\mathbf{k})$ is the wave number spectral density function;[14,15] $\mathcal{F}(\mathbf{k})$ is real valued. The physical picture is that, under the action of a fairly steady wind over long times and large areas of the sea surface, an approximate equilibrium is reached between the energy being deposited in the sea by the wind and the various linear and nonlinear loss mechanisms. The resulting steady state may be described as a superposition of linear waves, (1), whose statistics are specified by (2).[14,15]

Given a sea state dispersion relation $2\pi f=\omega(\mathbf{k})$, one can relate the wave vector density function $\mathcal{F}(\mathbf{k})$ to the directional spectral density function $F(f,\theta)$ by[14,15]

$$\mathcal{F}(\mathbf{k})d^2k = F(f,\theta)df d\theta. \tag{3}$$

Here f is frequency in Hz and θ indicates direction of propagation relative to some fixed direction. As a function of θ, $F(f,\theta)$ is generally strongly peaked at angles near those of the direction of the prevailing winds. The integrated spectrum,

$$\bar{F}(f) = \int_0^{2\pi} F(f,\theta)\, d\theta, \tag{4}$$

is known as the frequency spectrum. The "significant wave height"

$$H_S = 4\sqrt{\int \mathcal{F}(\mathbf{k})\, d^2k} = 4\sqrt{\int_0^\infty \bar{F}(f)\, df} \tag{5}$$

is commonly used as a measure of how excited the sea state is.

Ocean wave spectra generally have sharp low-frequency cutoffs and high-frequency tails. These features are due to nonlinear effects which cause energy to cascade from longer to shorter wavelengths. For energetic seas the frequency spectrum generally saturates at some limiting form strongly peaked in a narrow frequency band centered at around 0.1 Hz.[14,15] A quasi-empirical form for the frequency spectrum is given by the two-parameter JONSWAP[15,16] model for highly excited seas. JONSWAP spectra are shown in Fig. 2 for several significant wave heights and peak frequencies. Note that while observed microbarom spectra are peaked at about 0.2 Hz, the spectra of the ocean waves that produced them are peaked at about 0.1 Hz. The frequency doubling observed in the microbarom spectra is a consequence of the nonlinearity of the radiation mechanism.[2]

Assuming that the source region is over deep water one has[17] $\omega(\mathbf{k})=\sqrt{g|\mathbf{k}|}$, where g is the acceleration due to gravity at the earth's surface, $g=9.8\ \mathrm{m/s^2}$. At the peak frequency, 0.1 Hz, one has an ocean surface wavelength of $2\pi/|\mathbf{k}| \approx 156$ m. At the microbarom peak frequency of 0.2 Hz the acoustic wavelengths are about 1700 m in the air and about 7500 m in the water. Thus, in the frequency band in which microbarom radiation is significant, the ocean surface wavelengths are much shorter than the acoustic wavelengths. It follows that, if c_a and c_w are the speeds of sound in the atmosphere and ocean, respectively, the values of $\mathbf{k}$ for which $\mathcal{F}(\mathbf{k})$ is significant satisfy

$$\frac{\omega}{c_w} < \frac{\omega}{c_a} \ll |\mathbf{k}|. \tag{6}$$

Note as well that the effect of gravity is small in the sense that

$$\frac{g}{\omega c_w} < \frac{g}{\omega c_a} \ll 1. \tag{7}$$

The pointlike correlation between sea state wave vector components indicated by the δ function in (2) is equivalent to the translation invariance of the stochastic process. In reality the statistical properties of the sea state are neither temporally nor translationally invariant. However, these changes are negligible over times of several ocean periods (tens of seconds) and distances of several ocean wavelengths (hundreds of meters). It will be assumed that (2) remains valid if $\mathcal{F}$ is allowed to vary slowly with both time and horizontal position.

B. The equations of motion in the source region

The highly active region S of the ocean surface and a shallow layer of the ocean and atmosphere surrounding S will be referred to as the source region. The vertical extent of the source region, above and below the ocean surface, is greater than the significant ocean wavelengths; however, the height to which it extends in the atmosphere is much less than the acoustic wavelength in air and the depth to which it extends in the water is much less than the acoustic wavelength in water. In this region the effects of viscosity, thermal conduction and molecular relaxation can be ignored. Thus, in the source region the air and the water obey the equations of lossless fluid mechanics.[18] One has the equation of continuity

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \tag{8}$$

the Euler equation

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v}\right) + \nabla P = -\rho g \hat{\mathbf{z}}, \tag{9}$$

where g is the acceleration due to gravity, and the thermodynamic equation of state,

$$\rho = f_\sigma(P) \tag{10}$$

where f_σ, for $\sigma=a,w$, are the adiabatic equations of state for air, a, and water, w, respectively. Note that this notation will be used throughout for quantities which are discontinuous across the air/water interface: the subscript σ will be assumed to be a for air and w for water.

The pressure and normal components of the velocity field must be continuous at the air/water interface. Thus one has

$$P(\mathbf{x}_H, \xi + 0^+, t) = P(\mathbf{x}_H, \xi - 0^+, t), \tag{11}$$

and

$$\mathbf{n}(\mathbf{x}_H, t)\cdot(\mathbf{v}(\mathbf{x}_H, \xi + 0^+, t) - \mathbf{v}(\mathbf{x}_H, \xi - 0^+, t)) = 0 \tag{12}$$

where $\mathbf{n}$ is an upwardly pointing normal vector to the interface (see Fig. 1) chosen here to be

$$\mathbf{n}(\mathbf{x}_H, t) = \begin{pmatrix} -\nabla_H \xi \\ 1 \end{pmatrix}$$

with

$$\nabla_H = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{pmatrix}.$$

Further, the vertical component of the acceleration of the interface must equal that of the adjoining fluid elements,

$$v_z(\mathbf{x}_H, \xi \pm 0^+, t) = \frac{\partial \xi}{\partial t} + \mathbf{v}(\mathbf{x}_H, \xi \pm 0^+, t)\cdot\nabla\xi. \tag{13}$$

C. The ambient state

A solution to (11)–(13) with $\xi=0$ will be called an ambient state. The corresponding pressure, density, and velocity will be denoted P_0, ρ_0, and $\mathbf{v}_0$, respectively. We will assume that P_0, ρ_0, and $\mathbf{v}_0$ depend only on height/depth z and that $\hat{\mathbf{z}}\cdot\mathbf{v}_0=0$. With these assumptions the equations of fluid mechanics reduce to

$$\frac{dP_0}{dz} = -\rho_0 g \tag{14}$$

for which the general solution is given implicitly by

$$\int_{P_0(0)}^{P_0(z)} \frac{1}{f(P_0)}\, dP_0 = -gz.$$

Letting

$$\rho_\sigma = f_\sigma(P_0(0))$$

and introducing the small-signal sound speeds at the interface

$$c_\sigma = \sqrt{\frac{1}{f'_\sigma(P_0(0))}}$$

one has, for small z,

$$P_0(z) = P_0(0) - \rho_\sigma g z + \frac{\rho_\sigma g^2}{2c_\sigma^2} z^2 + \ldots. \tag{15}$$

It will be assumed here that

$$\mathbf{v}_0 = 0. \tag{16}$$

Note that the ambient density and sound speed profiles, $\rho_0 = f_\sigma(P_0(z))$ and $c_0(z) = 1/\sqrt{f'_\sigma(P_0(z))}$, are discontinuous at the air/water interface. By (7) and (15) both are approximately piecewise constant,

$$\rho_0 \approx \begin{cases} \rho_a & \text{if } z > 0, \\ \rho_w & \text{if } z < 0, \end{cases}$$

and

$$c_0 \approx \begin{cases} c_a & \text{if } z > 0, \\ c_w & \text{if } z < 0, \end{cases}$$

in the source region. For later reference note that $f''_\sigma(P_0)$ is related to the "B/A" parameter of nonlinearity[19] through

$$\frac{B}{A} = -\rho_0 c_0^4 f''_\sigma(P_0) \approx -\rho_\sigma c_\sigma^4 f''_\sigma(P_0(0)). \tag{17}$$

The parameter B/A has been tabulated for many gases and fluids.[19,20] For air one has $B/A \approx 0.4$ and for sea water $B/A \approx 5.25$.

D. The order expansion

Solving (11) and (13) in the case in which $\xi \neq 0$ requires some approximation method. It was pointed out in Ref. 2 that the nonlinearity of the equations can be treated using regular perturbation theory to second order.[20,21] The procedure is to perform an expansion about the ambient state,

$$P = P_0 + P_1 + P_2 + \cdots,$$

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots,$$

$$\rho = \rho_0 + \rho_1 + \rho_2 + \cdots,$$

and

$$\xi = \xi_1 + \xi_2 + \cdots,$$

where the subscript 1 indicates solutions of the linear approximation and the subscript 2 indicates terms quadratic in the linear solutions. In the expansion of the sea state ξ it will be assumed[3,5] that the linear approximation ξ_1 is given by the Gaussian process ξ_S specified by (1) and (2). One then substitutes the order expansions into the equations of fluid mechanics and interface conditions and expands. The solution is obtained order by order, beginning with the ambient state at zeroth order.

In expanding the interface conditions some care is required since the interface is in motion. The interface conditions for the full problem are imposed at $z = \xi(\mathbf{x}_H, t)$, however, for the ambient state they are imposed at $z = 0$. Thus the order expansion for the interface conditions involves both expanding in powers of ξ about $\xi = 0$ as well as in the order expanded variables.

The order parameter, or Mach number, for such a perturbation expansion is the ratio of a typical fluid velocity to the small-signal sound speed.[20] For this problem the order parameter can be taken to be $\omega_0 H_S/c_0$ where $\omega_0 = 2\pi f_0$ and f_0 is the peak frequency of the sea state, about 0.1 Hz. Note that this order parameter is in fact the Mach number at the interface times 2π and that the Mach numbers for air and water are different. The subscript n in the perturbation expansion indicates a term whose magnitude is proportional to $(\omega_0 H_S/c_0)^n$. In addition to Mach number one may exploit the small parameters $\omega_0/c_\sigma k_0$ and $g/\omega_0 c_\sigma$; here k_0 is equal to $|\mathbf{k}|$ evaluated at the peak frequency f_0. The second-order solutions are required, but only to lowest nonzeroth order in these small parameters.

E. The linear response

In this section the linear approximation will be obtained[17] and put into a suitable form. It has been emphasized[2,4–6] that in the linear approximation the sea does not radiate. Indeed, the first-order acoustic fields are vertically evanescent and are negligible outside of the source region itself.

Substituting the order expansions into the equations of fluid mechanics at first order one has the familiar equations of lossless acoustics. The equation of state, (10), becomes

$$\rho_1 = \frac{1}{c_\sigma^2} P_1,$$

the equation of continuity, (8), becomes

$$\frac{\partial P_1}{\partial t} + \rho_\sigma c_\sigma^2 \nabla \cdot \mathbf{v}_1 = 0,$$

and the Euler equation (9) becomes

$$\nabla P_1 + \rho_\sigma \frac{\partial \mathbf{v}_1}{\partial t} + g\rho_1 \hat{\mathbf{z}} = 0.$$

The interface conditions must be expanded about $\xi = 0$ to first order in ξ_1 as well as in the first-order variables. The first-order part of the pressure interface condition (11) is

$$P_1|_{z=-0^+}^{0^+} + \xi_1 \left.\frac{\partial P_0}{\partial z}\right|_{z=-0^+}^{0^+} = 0,$$

which, with (14) gives

$$(P_1 - \rho_\sigma g \xi_1)|_{z=-0^+}^{0^+} = 0. \tag{18}$$

The velocity interface condition, (12), together with the assumption, (16), of no mean flow, gives

$$v_{1z}|_{z=-0^+}^{0^+} = 0 \tag{19}$$

and (13) gives

$$\frac{\partial \xi_1}{\partial t} = v_{1z}(\mathbf{x}_H, 0^+, t) = v_{1z}(\mathbf{x}_H, -0^+, t). \tag{20}$$

Introduce a velocity potential ϕ_1 with $\mathbf{v}_1 = \nabla\phi_1 + \mathbf{w}_1$, with $\nabla \cdot \mathbf{w}_1 = 0$ and with

$$P_1(\mathbf{x}_H, z, t) = -\rho_\sigma \frac{\partial \phi_1}{\partial t}. \tag{21}$$

Choosing

$$\rho_\sigma \frac{\partial \mathbf{w}_1}{\partial t} = -\frac{g}{c_\sigma^2} P_1 \hat{\mathbf{z}} \tag{22}$$

the first-order Euler equation is solved. Substituting into the first-order equation of continuity one obtains the wave equation for the velocity potential,

$$\left(\nabla^2 - \frac{1}{c_\sigma^2}\frac{\partial^2}{\partial t^2}\right)\phi_1 = 0. \tag{23}$$

The system is driven by the interface motion through (18). As in (1) one has the horizontal wave vector expansion

$$\xi_1(\mathbf{x}_H, t) = \mathrm{Re}\int \hat{\xi}_1(\mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{x}_H - \omega(\mathbf{k})t)}\, d^2k \tag{24}$$

for the interface. The solutions to the first-order equations can be expanded similarly. Substituting in (23) one obtains

$$\phi_1(\mathbf{x}, t) = \mathrm{Re}\int \hat{\phi}_1^{(\sigma)}(\mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{x}_H - \omega(\mathbf{k})t) - \sqrt{|\mathbf{k}|^2 - \omega^2/c_\sigma^2}|z|}\, d^2k. \tag{25}$$

It follows from (22) that $w_{1z} \sim (g/\omega_0 c_\sigma)(1/\rho_\sigma c_\sigma)P_1$. Substituting (25) into (21) one finds that $\partial\phi_1/\partial z \sim (k_0 c_\sigma/\omega_0)(1/\rho_\sigma c_\sigma)P_1$. It follows that $w_{1z} \sim (g/\omega_0 c_\sigma) \times (\omega_0/k_0 c_\sigma)(\partial\phi_1/\partial z)$ so that, by (6) and (7), $\mathbf{w}_1$ is negligible as compared to $\nabla\phi_1$. It follows that one may write $\mathbf{v}_1 = \nabla\phi_1$.

Expanding with respect to the horizontal wave vector one has

$$P_1(\mathbf{x}, t) = \mathrm{Re}\int \hat{P}_1^{(\sigma)}(\mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{x}_H - \omega(\mathbf{k})t) - \sqrt{|\mathbf{k}|^2 - \omega^2/c_\sigma^2}|z|}\, d^2k \tag{26}$$

and

$$\mathbf{v}_1(\mathbf{x}, t) = \mathrm{Re}\int \hat{\mathbf{v}}_1^{(\sigma)}(\mathbf{k}) e^{i(\mathbf{k}\cdot\mathbf{x}_H - \omega(\mathbf{k})t) - \sqrt{|\mathbf{k}|^2 - \omega^2/c_\sigma^2}|z|}\, d^2k. \tag{27}$$

Substituting into (18)–(20) and letting $(-1)^\sigma$ be 1 in the air, $\sigma = a$, and -1 in the water, $\sigma = w$, one obtains

$$\hat{\phi}_1^{(\sigma)}(\mathbf{k}) = (-1)^\sigma \frac{i\omega}{\sqrt{|\mathbf{k}|^2 - \omega^2/c_\sigma^2}} \hat{\xi}_1(\mathbf{k}), \tag{28}$$

$$\hat{P}_1^{(\sigma)}(\mathbf{k}) = -(-1)^\sigma \frac{\rho_\sigma \omega^2}{\sqrt{|\mathbf{k}|^2 - \omega^2/c_\sigma^2}} \hat{\xi}_1(\mathbf{k}), \tag{29}$$

and

$$\hat{\mathbf{v}}_1^{(\sigma)}(\mathbf{k}) = \omega\left(-(-1)^\sigma \frac{\mathbf{k}}{\sqrt{|\mathbf{k}|^2 - \omega^2/c_\sigma^2}} - i\hat{\mathbf{z}}\right)\hat{\xi}_1(\mathbf{k}). \tag{30}$$

By (6) the first-order solutions decrease exponentially with either altitude or depth. At distances large compared to $1/|\mathbf{k}|$ from the interface they become negligible. In particular, the linear solutions do not radiate. The continuity of pressure (18) gives the surf dispersion relation

$$0 = (\rho_w - \rho_a)g + \frac{\omega^2}{|\mathbf{k}|}\left(\frac{\rho_w}{\sqrt{1 - \dfrac{\omega^2}{|\mathbf{k}|^2 c_w^2}}} + \frac{\rho_a}{\sqrt{1 - \dfrac{\omega^2}{|\mathbf{k}|^2 c_a^2}}}\right)$$

which gives

$$|\mathbf{k}| \approx \frac{\omega^2}{g}$$

so that

$$\omega(\mathbf{k}) \approx \sqrt{g|\mathbf{k}|}. \tag{31}$$

Note that with this dispersion relation the two order parameters $\omega_0/k_0 c_\sigma$ and $g/\omega_0 c_\sigma$ are identical.

Checking the orders of the first-order solutions one finds that

$$\mathbf{v}_1 \sim c_\sigma \frac{\omega_0 H_S}{c_\sigma}$$

while

$$P_1 \sim \rho_\sigma c_\sigma^2 \frac{\omega_0}{k_0 c_\sigma}\frac{\omega_0 H_S}{c_\sigma}.$$

Both are first order in Mach number as expected, however, the pressure P_1 is also first order in the small parameter $\omega_0/k_0 c_\sigma$.

The first-order solutions given by Eqs. (28) and (30) are well approximated by setting $\omega/|\mathbf{k}|c_\sigma = 0$. In the linear approximation this is equivalent to assuming both air and water to be incompressible. In this approximation the acoustic pressure fields, $\hat{P}_1^{(\sigma)}$, are zero and the velocity fields, $\hat{\mathbf{v}}_1^{(\sigma)}$, differ in the water and in the air only by phase: the vertical components are equal while the horizontal components have equal magnitudes but opposite signs. In Ref. 2 it was shown that to obtain the oceanic microbaroms it is sufficient to replace the first-order solutions with their incompressible approximations. It will be seen below that to obtain the atmospheric microbaroms the incompressible approximation to the first-order solutions is not sufficient.

F. The second-order acoustics

It is known[2,5] that in the second-order approximation the sea does radiate. In this section the leading-order corrections to the linear approximation, the terms of second order in Mach number, are obtained. The solution is simplified somewhat by obtaining only the leading-order terms in the small parameter $\omega_0/k_0 c_\sigma$. The general form of the second-order solution is quite complicated, even to leading order in $\omega_0/k_0 c_\sigma$, and will not be given here. Rather, in this section the solution will be presented as a superposition of pairs of plane waves. The explicit determination of the solution is postponed to Sec. III where it is shown that, to determine the statistical properties of the microbarom signal, only the terms corresponding to counter-propagating ocean waves of equal frequency are required.[3]

Continuing the order expansion, at second order one finds[21] that the equation of state, (10), can be written

$$\rho_2 - \frac{1}{c_\sigma^2}P_2 = \frac{1}{2}f''(P_0)P_1^2,$$

the equation of continuity, (8), can be written

$$\frac{\partial P_2}{\partial t}+\rho_\sigma c_\sigma^2\nabla\cdot\mathbf{v}_2=\frac{(1-\rho_\sigma c_\sigma^4 f''(P_0))}{\rho_\sigma c_\sigma^2}P_1\frac{\partial P_1}{\partial t}-\mathbf{v}_1\cdot\nabla P_1$$

and the Euler equation, (9), can be written

$$\nabla P_2+\rho_\sigma\frac{\partial\mathbf{v}_2}{\partial t}+g\rho_2\hat{\mathbf{z}}=-\frac{1}{c_\sigma^2}P_1\frac{\partial\mathbf{v}_1}{\partial t}-\frac{1}{2}\rho_\sigma\nabla(\mathbf{v}_1\cdot\mathbf{v}_1).$$

Consider the orders of the various terms in the second-order equations of state and continuity. Note that, recalling (17),

$$\frac{1}{2}f''(P_0)P_1^2\sim\frac{B}{A}\rho_\sigma\left(\frac{\omega_0}{k_0c_\sigma}\right)^2\left(\frac{\omega_0H_S}{c_\sigma}\right)^2$$

and

$$\frac{(1-\rho_\sigma c_\sigma^4 f''(P_0))}{\rho_\sigma c_\sigma^2}P_1\frac{\partial P_1}{\partial t}\sim\left(1+\frac{B}{A}\right)\rho_\sigma c_\sigma^2\omega_0\left(\frac{\omega_0}{k_0c_\sigma}\right)^2\left(\frac{\omega_0H_S}{c_\sigma}\right)^2.$$

while

$$\mathbf{v}_1\cdot\nabla P_1\sim\rho_\sigma c_\sigma^2\omega_0\left(\frac{\omega_0H_S}{c_\sigma}\right)^2.$$

All of these terms are second order in Mach number ω_0H_S/c_σ. However, the two terms involving the second derivative $f''(P_0)$ are second-order in the small parameter ω_0/k_0c_σ as well and thus can be dropped. Similarly, on the right side of the second-order Euler equation one has the terms

$$\frac{1}{c_\sigma^2}P_1\frac{\partial\mathbf{v}_1}{\partial t}\sim\rho_\sigma c_\sigma\omega_0\frac{\omega_0}{k_0c_\sigma}\left(\frac{\omega_0H_S}{c_\sigma}\right)^2$$

and

$$\frac{1}{2}\rho_\sigma\nabla(\mathbf{v}_1\cdot\mathbf{v}_1)\sim\rho_\sigma c_\sigma^2k_0\left(\frac{\omega_0H_S}{c_\sigma}\right)^2.$$

The first of these terms is again smaller by two orders of the small parameter ω_0/k_0c_σ than the second and can be dropped. One thus obtains the simpler set of equations

$$\rho_2-\frac{1}{c_\sigma^2}P_2=0,\tag{32}$$

$$\frac{\partial P_2}{\partial t}+\rho_\sigma c_\sigma^2\nabla\cdot\mathbf{v}_2=-\mathbf{v}_1\cdot\nabla P_1\tag{33}$$

and

$$\nabla P_2+\rho_\sigma\frac{\partial\mathbf{v}_2}{\partial t}+g\rho_2\hat{\mathbf{z}}=-\frac{1}{2}\rho_\sigma\nabla(\mathbf{v}_1\cdot\mathbf{v}_1).\tag{34}$$

The second-order part of the pressure interface condition (11) is

$$\left(P_2+\xi_2\frac{\partial P_0}{\partial z}+\xi_1\frac{\partial P_1}{\partial z}+\frac{1}{2}\xi_1^2\frac{\partial^2P_0}{\partial z^2}\right)\Bigg|_{z=-0^+}^{0^+}=0.$$

Using the first-order Euler equation one finds

$$\xi_1\frac{\partial P_1}{\partial z}=-\rho_\sigma\xi_1\frac{\partial v_{1z}}{\partial t}\sim\rho_\sigma c_\sigma^2\left(\frac{\omega_0H_S}{c_\sigma}\right)^2$$

and using (15) one finds

$$\frac{1}{2}\xi_1^2\frac{\partial^2P_0}{\partial z^2}=\frac{1}{2}\frac{\rho_\sigma g^2}{c_\sigma^2}\xi_1^2\sim\rho_\sigma c_\sigma^2\left(\frac{g}{\omega_0c_\sigma}\right)^2\left(\frac{\omega_0H_S}{c_\sigma}\right)^2.$$

The last term is second order in g/ω_0c_σ and can be dropped. Using (15), the second-order part of the pressure interface condition can be written

$$(P_2-\rho_\sigma g\xi_2)|_{z=-0^+}^{0^+}=\rho_\sigma\xi_1\frac{\partial v_{1z}}{\partial t}\Bigg|_{z=-0^+}^{0^+}.\tag{35}$$

The second order part of the velocity interface condition (12) is

$$v_{2z}|_{z=-0^+}^{0^+}=\left(-\xi_1\frac{\partial v_{1z}}{\partial z}+\mathbf{v}_1\cdot\nabla_H\xi_1\right)\Bigg|_{z=-0^+}^{0^+}\tag{36}$$

and the second-order part of (13) is

$$\frac{\partial\xi_2}{\partial t}=\left(v_{2z}+\xi_1\frac{\partial v_{1z}}{\partial z}-\mathbf{v}_1\cdot\nabla_H\xi_1\right)\Bigg|_{z=\pm0^+}.\tag{37}$$

The procedure is to substitute the linear solutions (25)–(30) into the second-order wave equations (32)–(34) and find the outgoing solution which satisfies the interface conditions (35)–(37). The solutions are simplified somewhat by using the inequality (6) to conclude that compressibility is insignificant in the linear approximation so that $\sqrt{|\mathbf{k}|^2-\omega(\mathbf{k})^2/c_\sigma^2}\approx|\mathbf{k}|$. This approximation is valid as long as it gives a nonzero result.

The source terms in (36) and (37), however, can be written

$$-\xi_1\frac{\partial v_{1z}}{\partial z}+\mathbf{v}_1\cdot\nabla_H\xi_1=-\xi_1\nabla\cdot\mathbf{v}_1+\nabla_H(\xi_1\mathbf{v}_1).\tag{38}$$

The term $\nabla\cdot\mathbf{v}_1$ is zero in the incompressible approximation. In the product $\xi_1\mathbf{v}_1$ the terms which are responsible for the microbarom radiation, those with wave numbers of equal magnitude (and thus equal frequency) but opposite direction,[2,3] are constant so that for these components the term $\nabla_H\cdot(\xi_1\mathbf{v}_1)$ is zero as well. Thus, while the right side of (36) appears to be large, of order $(\omega_0/k_0c_\sigma)^{-1}(\omega_0H_S/c_\sigma)^2$, in fact its contribution to the microbarom radiation will be seen to be small. For this reason some care must be taken in evaluating the inhomogeneous part of the velocity interface condition (36). In particular, the incompressible approximation cannot be used.

To solve the second-order Euler equation (34) one can write $\mathbf{v}_2=\nabla\phi_2+\mathbf{w}_2$ where $\nabla\cdot\mathbf{w}_2=0$,

$$\rho_\sigma\frac{\partial\mathbf{w}_2}{\partial t}+\frac{g}{c_\sigma^2}P_2\hat{\mathbf{z}}=0,$$

and ϕ_2 is given by

$$P_2 + \rho_\sigma \frac{\partial \phi_2}{\partial t} = -\frac{1}{2}\rho_\sigma \mathbf{v}_1 \cdot \mathbf{v}_1. \tag{39}$$

With these definitions (34) is satisfied. As in Sec. II E, $\mathbf{w}_2$ can be shown to be smaller than $\nabla\phi_2$, in this case by a single power of $g/\omega_0 c_\sigma$. Consequently, it can be ignored so that one has $\mathbf{v}_2 = \nabla\phi_2$.

Substituting (39) into the second-order equation of continuity (33) and using the first-order Euler equation to simplify, one obtains the second-order wave equation

$$\left(\nabla^2 - \frac{1}{c_\sigma^2}\frac{\partial^2}{\partial t^2}\right)\phi_2 = \frac{1}{c_\sigma^2}\frac{\partial}{\partial t}\mathbf{v}_1 \cdot \mathbf{v}_1. \tag{40}$$

Note that, since the linear solutions decrease rapidly with distance from the interface, the source term on the right side of (40) does so as well.

The second-order wave equation contains the effects of nonlinearities in the air and the water themselves. The second-order interface conditions contain the nonlinear effects due the fact that the motions of the fluids affect the motion of the interface. To separate the effects of the fluid media from those of the undulating interface one may write the solution of (40) as

$$\phi_2(\mathbf{x}_H, z, t) = \phi_p(\mathbf{x}_H, z, t) + \phi_h(\mathbf{x}_H, z, t), \tag{41}$$

where ϕ_p is a particular solution and ϕ_h is the solution of the homogeneous wave equation required so that the interface conditions (35)–(37) are satisfied.

To determine ϕ_p substitute (27) and (30), into (40). In evaluating the source term the incompressible approximation $\sqrt{|\mathbf{k}|^2 - \omega(\mathbf{k})^2/c_\sigma^2} \approx |\mathbf{k}|$ may be used here since it gives a nonzero result. The source term on the right of (40) becomes

$$\begin{aligned}
&\frac{1}{c_\sigma^2}\frac{\partial}{\partial t}\mathbf{v}_1 \cdot \mathbf{v}_1 \\
&\quad = \int\int [\mathcal{R}_\sigma^{(+)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q}) e^{i((\mathbf{k}+\mathbf{q})\cdot\mathbf{x}_H - (\omega(\mathbf{k})+\omega(\mathbf{q}))t)} \\
&\qquad + \mathcal{R}_\sigma^{(-)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q})^* e^{i((\mathbf{k}-\mathbf{q})\cdot\mathbf{x}_H - (\omega(\mathbf{k})-\omega(\mathbf{q}))t)}] \\
&\qquad \times e^{-(|\mathbf{k}|+|\mathbf{q}|)|z|}\, d^2k\, d^2q + \text{complex conjugate}
\end{aligned} \tag{42}$$

with

$$\mathcal{R}_\sigma^{(\pm)}(\mathbf{k},\mathbf{q}) = -\frac{i}{c_\sigma^2}\omega(\mathbf{k})\omega(\mathbf{q})(\omega(\mathbf{k}) \pm \omega(\mathbf{q}))\left(\frac{\mathbf{k}\cdot\mathbf{q}}{|\mathbf{k}||\mathbf{q}|} \mp 1\right).$$

Using (42) the particular solution ϕ_p can be chosen to be

$$\begin{aligned}
&\phi_p(\mathbf{x}_H, z, t) \\
&\quad = \int\int [\mathcal{Q}_\sigma^{(+)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q}) \\
&\qquad \times e^{i((\mathbf{k}+\mathbf{q})\cdot\mathbf{x}_H - (\omega(\mathbf{k})+\omega(\mathbf{q}))t) - (|\mathbf{k}|+|\mathbf{q}|)|z|} \\
&\qquad + \mathcal{Q}_\sigma^{(-)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q})^* \\
&\qquad \times e^{i((\mathbf{k}-\mathbf{q})\cdot\mathbf{x}_H - (\omega(\mathbf{k})-\omega(\mathbf{q}))t) - (|\mathbf{k}|+|\mathbf{q}|)|z|}]\, d^2k\, d^2q \\
&\qquad + \text{complex conjugate}
\end{aligned} \tag{43}$$

with

$$\mathcal{Q}_\sigma^{(\pm)}(\mathbf{k},\mathbf{q}) = \frac{\mathcal{R}_\sigma^{(\pm)}(\mathbf{k},\mathbf{q})}{[(\omega(\mathbf{k}) \pm \omega(\mathbf{q}))^2/c_\sigma^2] + 2(|\mathbf{k}||\mathbf{q}| \mp \mathbf{k}\cdot\mathbf{q})}.$$

The homogeneous solution ϕ_h is then of the form

$$\begin{aligned}
&\phi_h(\mathbf{x}_H, z, t) \\
&\quad = \int\int [C_\sigma^{(+)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q}) \\
&\qquad \times e^{i((\mathbf{k}+\mathbf{q})\cdot\mathbf{x}_H + \Omega_+(\mathbf{k},\mathbf{q})|z| - (\omega(\mathbf{k})+\omega(\mathbf{q}))t)} \\
&\qquad + C_\sigma^{(-)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q})^* \\
&\qquad \times e^{i((\mathbf{k}-\mathbf{q})\cdot\mathbf{x}_H + \Omega_-(\mathbf{k},\mathbf{q})|z| - (\omega(\mathbf{k})-\omega(\mathbf{q}))t)}]\, d^2k\, d^2q \\
&\qquad + \text{complex conjugate}
\end{aligned} \tag{44}$$

with

$$\Omega_\pm(\mathbf{k},\mathbf{q}) = \sqrt{\frac{(\omega(\mathbf{k}) \pm \omega(\mathbf{q}))^2}{c_\sigma^2} - (\mathbf{k} \pm \mathbf{q})^2}$$

and the second-order contribution to the sea state is given by

$$\begin{aligned}
\xi_2(x_H, t) = \int\int [&\hat{\xi}^{(+)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q}) \\
&\times e^{i((\mathbf{k}+\mathbf{q})\cdot\mathbf{x}_H - (\omega(\mathbf{k})+\omega(\mathbf{q}))t)} + \hat{\xi}^{(-)}(\mathbf{k},\mathbf{q})\hat{\xi}_1(\mathbf{k})\hat{\xi}_1(\mathbf{q})^* \\
&\times e^{i((\mathbf{k}-\mathbf{q})\cdot\mathbf{x}_H - (\omega(\mathbf{k})-\omega(\mathbf{q}))t)}]\, d^2k\, d^2q \\
&+ \text{complex conjugate}.
\end{aligned} \tag{45}$$

The coefficients $C_\sigma^{(\pm)}(\mathbf{k},\mathbf{q})$ and $\hat{\xi}^{(\pm)}(\mathbf{k},\mathbf{q})$ are determined by the condition that the interface conditions (35)–(37) must be satisfied. The explicit determination of $C_\sigma^{(\pm)}(\mathbf{k},\mathbf{q})$ and $\hat{\xi}^{(\pm)}(\mathbf{k},\mathbf{q})$ will be postponed until the next section where it is shown that, to determine the statistical properties of the microbarom signal, only the coefficients $C_\sigma^{(+)}(\mathbf{k},-\mathbf{k})$ and $\hat{\xi}^{(+)}(\mathbf{k},-\mathbf{k})$ are required. Note that these particular coefficients correspond to the interaction of ocean waves of equal frequency and opposite propagation direction.

The decomposition (41) with the choice (43) for the particular solution separates the second-order velocity potential into a term, the particular solution, which is negligible outside of the source region and a term, the homogeneous solution, which contains the part of the field which radiates into the atmosphere and into the ocean. The particular solution contains the nonlinear effects produced in the air or water in the bulk rather than at the interface. These volume contributions do not directly produce acoustic radiation. The radiation is produced by the requirement that the interface conditions be satisfied. Note that the homogeneous solution must both account for nonlinear contributions to as well as correct for the deviations of the particular solution from the interface conditions. Specifically, the air and the water must accommodate both the nonlinear contributions to the interface motion as well as the nonlinear contributions to the flow in the bulk.

III. THE RECEIVED MICROBAROM SIGNAL

Let $P_\sigma(\mathbf{x}_H,z,t)$ be the acoustic pressure at large distances from the source region, either in the atmosphere, $\sigma=a$ and $z\geqslant 0$, or in the ocean, $\sigma=w$ and $z\leqslant 0$. To determine P_σ one must find the outward-propagating solution to the equations for acoustic propagation in the atmosphere/ocean which reduces in the source region to the solution produced in Sec. II F. One may write

$$P_\sigma = P_{\sigma p} + P_{\sigma h},$$

where

(*i*) in the source region $P_{\sigma p}$ and $P_{\sigma h}$ reduce to the solutions produced in Sec. II F,

$$P_{\sigma p} = -\rho_\sigma \frac{\partial \phi_p}{\partial t} - \frac{1}{2}\rho_\sigma \mathbf{v}_1 \cdot \mathbf{v}_1$$

and

$$P_{\sigma h} = -\rho_\sigma \frac{\partial \phi_h}{\partial t},$$

where ϕ_p and ϕ_h are given by (43) and (44);

(*ii*) away from the source region $P_{\sigma p}$ is negligible, so that $P_\sigma = P_{\sigma h}$, and $P_{\sigma h}$ is taken to satisfy the rigid ground boundary condition at the atmospheric side of the air/water interface and the pressure release boundary condition at the oceanic side

$$0 = \left.\frac{\partial P_{\sigma h}(\mathbf{x}_H,z,t)}{\partial z}\right|_{z=0^+} = P_{\sigma h}(\mathbf{x}_H,z,t)|_{z=-0^+}$$

Let $G_a(\mathbf{x}_H,z,\mathbf{x}'_H,z',t)$ be the Green's function describing the propagation of sound in the atmosphere over a rigid surface and let $G_w(\mathbf{x}_H,z,\mathbf{x}'_H,z',t)$ be the Green's function describing the propagation of sound in the ocean under a pressure release surface. Let S be the part of the $z=0$ plane which is in the source region. Then the Helmholtz-Kirchoff integral theorem gives

$$P_{\sigma h}(\mathbf{x}_H,z,t) = -\rho_\sigma \int_{-\infty}^{\infty}\int_S G_\sigma(\mathbf{x}_H,z,\mathbf{y}_H,0,t-\tau) \times \frac{\partial v_\sigma(\mathbf{y}_H,\tau)}{\partial \tau}\, d^2\mathbf{y}_H\, d\tau, \tag{46}$$

where

$$v_\sigma(\mathbf{x}_H,t) = \left.\frac{\partial \phi_h(\mathbf{x}_H,z,t)}{\partial z}\right|_{(-1)^\sigma z=0^+} \tag{47}$$

is the normal component of the homogeneous part of the fluid velocity at the interface.

A. The statistics of the microbarom signal and the sea state

Let $\langle\cdot\rangle_T$ represent both the sea state ensemble average, $\langle\cdot\rangle_S$, given in (2) as well as, if required, an average, $\langle\cdot\rangle_P$, over fluctuations in the propagation medium. Of interest is the correlation

$$\langle P_\sigma(\mathbf{x}_H,z,0)P_\sigma(\mathbf{x}'_H,z',\tau)^*\rangle_T = \lim_{T\to\infty}\frac{1}{T}\int_0^T P_\sigma(\mathbf{x}_H,z,t)P_\sigma(\mathbf{x}'_H,z',t+\tau)^*\, dt \tag{48}$$

between the signals received at (possibly) spatially separated points (see Fig. 3). The Fourier transform of (48) with respect to τ is known as the cross spectral density; for $\mathbf{x}'_H=\mathbf{x}_H$ it is the power spectrum.[22] Substituting (46) and (47), into (48) one has

$$\begin{aligned}\langle P_\sigma(\mathbf{x}_H,z,0)P_\sigma(\mathbf{x}'_H,z',\tau)^*\rangle_T &= \rho_\sigma^2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_S\int_S \langle G_\sigma(\mathbf{x}_H,z,\mathbf{y}_H,0,-\tau_1) \\ &\times G_\sigma(\mathbf{x}'_H,z',\mathbf{y}'_H,0,\tau-\tau_2)^*\rangle_P \\ &\times \left\langle \frac{\partial v_\sigma(\mathbf{y}_H,\tau_1)}{\partial\tau_1}\frac{\partial v_\sigma(\mathbf{y}'_H,\tau_2)^*}{\partial\tau_2}\right\rangle_S d^2y_H\, d^2y'_H\, d\tau_1\, d\tau_2.\end{aligned} \tag{49}$$

To compute v_σ in (47) one may substitute from (44) for the velocity potential ϕ_h. Note that v_σ is a binomial in the linear sea state amplitudes $\hat{\xi}_1$. As discussed in Sec. II D, to compute the average over the sea state in (49) we assume that $\hat{\xi}_1$ is the Gaussian process $\hat{\xi}_S$ given by (2). Since $\langle\cdot\rangle_S$ is Gaussian[23]

$$\begin{aligned}&\langle\hat{\xi}_1(\mathbf{k}_1)\hat{\xi}_1(\mathbf{q}_1)\hat{\xi}_1(\mathbf{k}_2)^*\hat{\xi}_1(\mathbf{q}_2)^*\rangle_S \\ &=\langle\hat{\xi}_1(\mathbf{k}_1)\hat{\xi}_1(\mathbf{k}_2)^*\rangle_S\langle\hat{\xi}_1(\mathbf{q}_1)\hat{\xi}_1(\mathbf{q}_2)^*\rangle_S \\ &\quad+\langle\hat{\xi}_1(\mathbf{k}_1)\hat{\xi}_1(\mathbf{q}_2)^*\rangle_S\langle\hat{\xi}_1(\mathbf{q}_1)\hat{\xi}_1(\mathbf{k}_2)^*\rangle_S \\ &=\mathcal{F}(\mathbf{k}_1)\mathcal{F}(\mathbf{q}_1)(\delta(\mathbf{k}_1-\mathbf{k}_2)\delta(\mathbf{q}_1-\mathbf{q}_2)+\delta(\mathbf{k}_1-\mathbf{q}_2) \\ &\quad\times\delta(\mathbf{q}_1-\mathbf{k}_2))\end{aligned}$$

so that one has

$$\begin{aligned}\langle v_\sigma(\mathbf{y}_H,\tau_1)v_\sigma(\mathbf{y}'_H,\tau_2)^*\rangle_S &= 2\int\int \mathcal{F}(\mathbf{k})\mathcal{F}(\mathbf{q}) \\ &\times|C_\sigma^{(+)}(\mathbf{k},\mathbf{q})|^2|\Omega_+(\mathbf{k},\mathbf{q})|^2 \\ &\times e^{i((\mathbf{k}+\mathbf{q})\cdot(\mathbf{y}_H-\mathbf{y}'_H)-(\omega(\mathbf{k})+\omega(\mathbf{q}))(\tau_1-\tau_2))}\, d^2k\, d^2q \\ &+\text{complex conjugate}+\text{time-independent terms}\end{aligned} \tag{50}$$

for the correlation between the normal velocity field v_σ at different positions and times.

Substitute (50) into (49). The integrals over τ_1 and τ_2 are Fourier transforms in time of the G_σ. Let $\hat{G}_\sigma(\mathbf{x}_H,z,\mathbf{x}'_H,z',\nu)$ be these Fourier transforms at angular frequency ν. Since the Green's functions for the acoustic propagation are approximately constant over distances which are small compared to acoustic wavelengths and since the significant sea state wave vectors $\mathbf{K}$ satisfy (6), one has,[24] to leading order in ω_0/k_0c_σ,

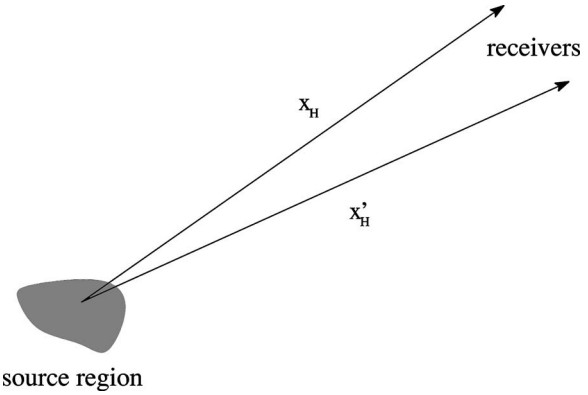


FIG. 3. The geometry. Microbaroms radiated from the source region are detected by distant sensors.

$$\int_S \int_S \hat{G}_\sigma(\mathbf{x}_H, z, \mathbf{y}_H, 0, \nu)\hat{G}_\sigma(\mathbf{x}'_H, z', \mathbf{y}'_H, 0, \nu)^* \times e^{i\mathbf{k}\cdot(\mathbf{y}_H - \mathbf{y}'_H)}\, d^2\mathbf{y}_H\, d^2\mathbf{y}'_H \approx (2\pi)^2 \delta(\mathbf{k}) \int_S \hat{G}_\sigma(\mathbf{x}_H, z, \mathbf{y}_H, 0, \nu) \times \hat{G}_\sigma(\mathbf{x}'_H, z', \mathbf{y}_H, 0, \nu)^*\, d^2\mathbf{y}_H.$$

It follows that

$$\langle P_\sigma(\mathbf{x}_H, z, 0) P_\sigma(\mathbf{x}'_H, z', \tau)^* \rangle_T = \frac{32\rho_\sigma^2 \pi^2}{c_\sigma^2} \int \left(\int_S \langle \hat{G}_\sigma(\mathbf{x}_H, z, \mathbf{y}_H, 0, 2\omega(\mathbf{k})) \times \hat{G}_\sigma(\mathbf{x}'_H, z', \mathbf{y}_H, 0, 2\omega(\mathbf{k}))^* \rangle_P\, d^2\mathbf{y}_H \right) \times \mathcal{F}(\mathbf{k})\mathcal{F}(-\mathbf{k}) |C_\sigma^{(+)}(\mathbf{k}, -\mathbf{k})|^2 \omega(\mathbf{k})^4 e^{-i2\omega(\mathbf{k})\tau}\, d^2k. \quad (51)$$

Equation (51) was first obtained by Hasselmann for the oceanic microbarom signal. It gives a direct relation between the spectral density function of the sea state and the statistical properties of the received microbarom signals.

The sharp low-frequency cutoff in the wave spectrum means that there is a smallest $|\mathbf{k}|$ and ω, corresponding to a longest wavelength and period in the ocean wave spectrum. At separations, $|\mathbf{y}_H - \mathbf{y}'_H|$, much longer than this longest wavelength or at time delays much greater than this longest period the motions of the sea surface become uncorrelated.[14] Thus, the normal velocity (50) becomes uncorrelated at separations large compared to the longest wavelength and at time delays long compared to the longest period. However, by (6) these separations are small compared to acoustic wavelengths. The physical picture which emerges from the stochastic sea state model is one of intermittent radiation, persisting for times long compared with the longest ocean wave period, from patches of the ocean surface randomly distributed across S, whose diameters are larger than the longest ocean wave wavelengths but smaller than acoustic wavelengths. This picture is consistent with the observations of individual microbarom wave trains reported in Ref. 25. Recall that radiation produced by patches which are small compared to acoustic wavelengths can be represented at large distances as a acoustic monopole fields. Thus the microbarom signal can be thought of as an incoherent superposition of fields produced by monopole sources at the sea surface (see also Ref. 5), precisely as presented by (51).

B. Solving the reduced problem

Taking the dispersion relation $\omega(\mathbf{k})$ to be given by the deep water dispersion relation, (31), all that remains is a determination of the coefficients $C_\sigma^{(\pm)}(\mathbf{k}, \mathbf{q})$ in (44). The task is greatly simplified by the fact that we need only $C_\sigma^{(+)}(\mathbf{k}, -\mathbf{k})$. Hence, only the sum frequency components for sea state wave vectors of equal magnitude and opposite direction are required. Referring to Eqs. (44) and (45), one sees that the restriction to interface waves of equal and opposite wave vectors reduces the acoustic problem to one of plane waves propagating normal to the air/water interface. The determination of the reduced set of coefficients is thus equivalent to the one-dimensional problem of finding the outward propagating velocity potential

$$\phi = C_\sigma^{(+)}(\mathbf{k}, -\mathbf{k}) e^{i[2\omega(\mathbf{k})/c_\sigma]|z| - i2\omega(\mathbf{k})t} \quad (52)$$

and uniformly vibrating surface displacement

$$\xi = \hat{\xi}^{(+)}(\mathbf{k}, -\mathbf{k}) e^{-i2\omega(\mathbf{k})t}, \quad (53)$$

which satisfy the interface conditions.

The interface conditions simplify as well. Substituting (39) and (41) into the pressure condition (35) one obtains

$$\left(\rho_\sigma \frac{\partial \phi_h}{\partial t} + \rho_\sigma g \xi_2 \right) \Bigg|_{z=-0^+}^{0^+} = \left(-\rho_\sigma \frac{\partial \phi_p}{\partial t} - \frac{1}{2}\rho_\sigma \mathbf{v}_1 \cdot \mathbf{v}_1 - \rho_\sigma \xi_1 \frac{\partial v_{1z}}{\partial t} \right) \Bigg|_{z=-0^+}^{0^+}.$$

Referring to (43), the relevant component of ϕ_p, to leading order in $\omega_0/k_0 c_\sigma$, is

$$i \frac{\omega(\mathbf{k})^3}{|\mathbf{k}|^2 c_\sigma^2} e^{-2|\mathbf{k}||z| - i2\omega(\mathbf{k})t}.$$

Substituting the linear form (27) in $\frac{1}{2}\rho_\sigma \mathbf{v}_1 \cdot \mathbf{v}_1 + \rho_\sigma \xi_1 \partial v_{1z}/\partial t$, one finds the relevant component to be

$$2\rho_\sigma \omega(\mathbf{k})^2 e^{i[2\omega(\mathbf{k})/c_\sigma]|z| - i2\omega(\mathbf{k})t}.$$

Thus, recalling that the density of water is much greater than that of air and noting that the contribution from ϕ_p is higher order in $\omega_0/k_0 c_\sigma$, the reduced velocity potential (52) and surface displacement (53) satisfy

$$\left(\rho_\sigma \frac{\partial \phi}{\partial t} + \rho_\sigma g \xi \right) \Bigg|_{z=-0^+}^{0^+} = 2\rho_w \omega(\mathbf{k})^2 e^{-i2\omega(\mathbf{k})t}. \quad (54)$$

Similarly, substituting (41) into the velocity condition (36) one obtains

$$\left.\frac{\partial \phi_h}{\partial z}\right|_{z=-0^+}^{0^+} = \left.\left(-\frac{\partial \phi_p}{\partial z} - \xi_1 \frac{\partial v_{1z}}{\partial z} + \mathbf{v}_1 \cdot \nabla_H \xi_1\right)\right|_{z=-0^+}^{0^+}$$

Using (38) one needs only the relevant component of $-\xi_1 \nabla \cdot \mathbf{v}_1 = (1/\rho_\sigma c_\sigma^2)(\partial P_1/\partial t)$. Referring to (29) one finds this component to be

$$(-1)^\sigma i \frac{\omega(\mathbf{k})^3}{|\mathbf{k}| c_\sigma} e^{i[2\omega(\mathbf{k})/c_\sigma]|z| - i2\omega(\mathbf{k})t}.$$

Adding the contribution from ϕ_p and dropping $1/c_w^2$ relative to $1/c_a^2$ one obtains

$$\left.\frac{\partial \phi}{\partial z}\right|_{z=-0^+}^{0^+} = 3i \frac{\omega(\mathbf{k})^3}{|\mathbf{k}| c_a^2} e^{-i2\omega(\mathbf{k})t} \tag{55}$$

and

$$\frac{\partial \xi}{\partial t} = \left.\left(\frac{\partial \phi}{\partial z} - 3i \frac{\omega(\mathbf{k})^3}{|\mathbf{k}| c_\sigma} e^{-i2\omega(\mathbf{k})t}\right)\right|_{z=\pm 0^+}. \tag{56}$$

Substitute (52) and (53), into (54)–(56) to obtain

$$\rho_a[-2i\omega(\mathbf{k}) C_a^{(+)}(\mathbf{k},-\mathbf{k}) + g\hat{\xi}^{(+)}(\mathbf{k},-\mathbf{k})] + \rho_w[2i\omega(\mathbf{k}) C_w^{(+)}(\mathbf{k},-\mathbf{k}) - g\hat{\xi}^{(+)}(\mathbf{k},-\mathbf{k})] = 2\rho_w \omega(\mathbf{k})^2,$$

$$\frac{2}{c_a} C_a^{(+)}(\mathbf{k},-\mathbf{k}) + \frac{2}{c_w} C_w^{(+)}(\mathbf{k},-\mathbf{k}) = 3\frac{\omega(\mathbf{k})^2}{|\mathbf{k}| c_a^2}$$

$$\hat{\xi}^{(+)}(\mathbf{k},-\mathbf{k}) = -\frac{1}{c_w} C_w^{(+)}(\mathbf{k},-\mathbf{k}) + \frac{3\omega(\mathbf{k})^2}{2|\mathbf{k}| c_w^2}.$$

Solving, one finds to leading order

$$C_a^{(+)}(\mathbf{k},-\mathbf{k}) = i\omega(\mathbf{k})\left(\frac{c_a}{c_w} - \frac{3}{2} i \frac{\omega(\mathbf{k})}{|\mathbf{k}| c_a}\right) \tag{57}$$

$$C_w^{(+)}(\mathbf{k},-\mathbf{k}) = -i\omega(\mathbf{k}), \tag{58}$$

and

$$\hat{\xi}(\mathbf{k},-\mathbf{k}) = i\frac{\omega(\mathbf{k})}{c_w}.$$

The atmospheric and oceanic velocity potential amplitudes (57) and (58) are of rather different forms. The oceanic microbarom velocity potential amplitude (58) is equivalent to that obtained by Longuet-Higgens[2] (it is also equivalent to that obtained, incorrectly, by Arendt and Fritts[6] for the atmospheric microbarom amplitude). The imaginary part of the atmospheric microbarom velocity potential amplitude (57) is equivalent to the form predicted by Brekhovskikh *et al.* in Ref. 5 for the normally propagating component of the atmospheric radiation.

The ratio of the velocity potential amplitudes (57) and (58), is given by the term $c_a/c_w - \frac{3}{2} i\omega(\mathbf{k})/|\mathbf{k}| c_a$. This term is plotted in Fig. 4 assuming the deep water dispersion relation (31). Note that for frequencies near 0.1 Hz the velocity potential amplitude in the atmosphere is smaller than in the ocean by about a factor of four. The real and imaginary parts, c_a/c_w and $-\frac{3}{2}\omega(\mathbf{k})/|\mathbf{k}| c_a$, respectively, of the ratio of (57) and (58), are of comparable magnitude. The real part c_a/c_w is the larger of the two, by about a factor of 3, and represents the contribution from sound radiated from the ocean into the atmosphere; that such a term is significant was pointed out in Ref. 5. The imaginary part $-\frac{3}{2}\omega(\mathbf{k})/|\mathbf{k}| c_a$ represents the compression of the air by the ocean waves.

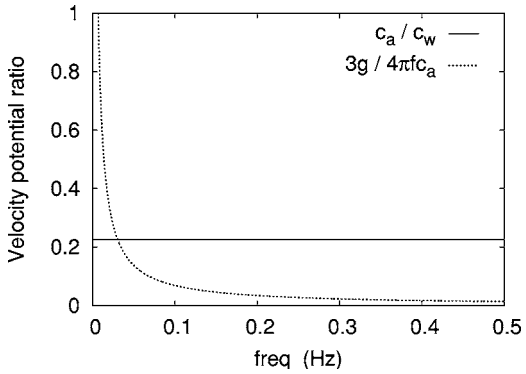


FIG. 4. The real and imaginary parts of the complex conjugate (so that both parts are positive) of the ratio, $c_a/c_w - \frac{3}{2} ig/2\pi f c_a$, of velocity potential amplitude in the atmosphere to that in the ocean for an infinitely deep ocean plotted as a function of frequency f.

Generally speaking, because of the enormous impedance contrast between air and water, the acoustic pressure radiated into the air by sound in the water is reduced by a factor of $\rho_a c_a/\rho_w c_w$. In this case, however, the effective interface pressure source given by the right side of (35) is proportional to the density of the water, ρ_w. This is the origin of the factor c_a/c_w in (57) and illustrates the basic physics of the problem. Once the interface is in motion the air and water in the source region must have roughly the same velocity fields [up to phase, see (30)]. However, the changes in pressure required to support changes in velocity are proportional to density. Thus the sound pressure fluctuations in the water associated with the motion of the interface are greater than those in the air by a factor of $\rho_w/\rho_a \approx 1000$.

The compressibility of the air arises in (57) because, as discussed above, the horizontally constant part of the effective interface velocity source given by the right side of (36) is proportional to $\nabla \cdot \mathbf{v}_1$, which is zero in the incompressible approximation. Indeed, to have a horizontally constant velocity source requires the fluid to be compressible.

C. The source spectra

Substituting (57) and (58), in (51), using the deep water dispersion relation (31) to relate the wave vector spectral density to the frequency direction spectral density through $\mathbf{k} = (4\pi^2/g) f^2(\cos\theta, \sin\theta)$ and

$$\mathcal{F}(\mathbf{k}) = \frac{g^2}{32\pi^4 f^3} F(f,\theta),$$

and choosing as integration variable the frequency of the received acoustic signal one finds

$$\langle P_\sigma(\mathbf{x}_H,z,0)P_\sigma(\mathbf{x}'_H,z',\tau)^*\rangle_T = \int_S \int_0^\infty \mathcal{Q}_\sigma(\mathbf{x}_H,\mathbf{x}'_H,z,z',\mathbf{y}_H,f)\mathcal{D}_\sigma(f)e^{-i2\pi f\tau}\,df\,d^2\mathbf{y}_H, \quad (59)$$

where $\mathcal{Q}_\sigma$ is the cross spectral density,[22]

$$\mathcal{Q}_\sigma(\mathbf{x}_H,\mathbf{x}'_H,z,z',\mathbf{y}_H,f) = \langle \hat{G}_\sigma(\mathbf{x}_H,z,\mathbf{y}_H,0,2\pi f)\hat{G}_\sigma(\mathbf{x}'_H,z',\mathbf{y}_H,0,2\pi f)^*\rangle_P, \quad (60)$$

for propagation from a point source at $\mathbf{y}_H \in S$ to receivers at $(\mathbf{x}_H,z)$ and $(\mathbf{x}'_H,z')$, and $\mathcal{D}_\sigma$ is the source strength spectrum squared

$$\mathcal{D}_\sigma(f) = \int_0^{2\pi} F\left(\frac{f}{2},\theta\right)F\left(\frac{f}{2},\theta+\pi\right)d\theta \cdot \begin{cases} \dfrac{4\rho_a^2 g^2\pi^4 f^3}{c_a^2}\left(\dfrac{9g^2}{4\pi^2 c_a^2 f^2}+\dfrac{c_a^2}{c_w^2}\right) \text{ in the atmosphere,} & \sigma = a \\ \dfrac{4\rho_w^2 g^2\pi^4 f^3}{c_w^2} \text{ in the ocean,} & \sigma = w. \end{cases} \quad (61)$$

A similar form for the source strength spectrum $\mathcal{D}_w$ for radiation into the ocean was first obtained by Hasselmann in Ref. 3. If $\hat{P}_\sigma$ is the Fourier transform in time of the received acoustic pressure P_σ, then[22] the cross spectral density of the received microbarom signals is given by

$$\langle \hat{P}_\sigma(\mathbf{x}_H,z,2\pi f)\hat{P}_\sigma(\mathbf{x}'_H,z',2\pi f)^*\rangle_T = \int_S \mathcal{Q}_\sigma(\mathbf{x}_H,\mathbf{x}'_H,z,z',\mathbf{y}_H,f)\mathcal{D}_\sigma(f)\,d^2\mathbf{y}_H. \quad (62)$$

The integral over S in (62) is required because, in general, as discussed above, $\mathcal{D}_\sigma$ is not constant over the source region, but varies slowly enough to be considered constant over distances of many ocean wavelengths.

Consider the ratio of the atmospheric to oceanic source strength spectra. Note that this ratio is independent of the sea state. One has

$$\sqrt{\frac{\mathcal{D}_a(f)}{\mathcal{D}_w(f)}} = \frac{\rho_a c_w}{\rho_w c_a}\sqrt{\frac{9g^2}{4\pi^2 c_a^2 f^2}+\frac{c_a^2}{c_w^2}}. \quad (63)$$

The source strength spectra ratio (63) is plotted in Fig. 5. For higher frequencies the ratio is asymptotic to the incompressible value $\rho_a/\rho_w \approx 10^{-3}$. In the frequency range of interest, 0.1 to 0.3 Hz, the ratio is always less than 2×10^{-3}. This explains the observation, pointed out in the Introduction, that atmospheric microbarom levels are generally three orders of magnitude lower than oceanic levels.

The integral,

$$\int_0^{2\pi} F\left(\frac{f}{2},\theta\right)F\left(\frac{f}{2},\theta+\pi\right)d\theta, \quad (64)$$

first obtained by Hasselmann,[3] is the only term in (61) which is not explicit. It is a factor common to both atmospheric and oceanic radiation and is a measure of the density of counter propagating waves at frequency f. Its determination is problematic. The largest component of a wave field consists of traveling waves propagating in the direction of the prevailing wind. Evaluating (64) requires knowing the tails of the angular part of the wave vector distribution. The determination of (64) thus depends on quantities which are difficult to determine, either experimentally or numerically.[7,15,26,27] It is possible that observed atmospheric and oceanic microbarom levels will ultimately be used to constrain ocean wave models.

IV. A SIMPLE MODEL FOR THE ATMOSPHERIC MICROBAROM SIGNAL FROM STORMS OVER THE OPEN OCEAN

In this section an example calculation is presented to illustrate an application of (61) and (62) to the atmospheric microbarom signal. A simple model is considered for the signal received from a distant isolated storm over the deep ocean far from any land masses. It is assumed that the source region is roughly centered at the origin $\mathbf{y}_H=\mathbf{0}$, that the horizontal extent of the source region is small compared with the distance $|\mathbf{x}_H|$ to the receiver, and that the source spectrum is constant over the source region.

To compute the power spectrum,

$$\langle|\hat{P}_\sigma(\mathbf{x}_H,z,2\pi f)|^2\rangle_T = \int_S \mathcal{Q}_\sigma(\mathbf{x}_H,\mathbf{x}_H,z,z,\mathbf{y}_H,f)\mathcal{D}_\sigma(f)\,d^2\mathbf{y}_H,$$

of the received signal both the power spectral density, $\mathcal{Q}_\sigma(\mathbf{x}_H,\mathbf{x}_H,z,z,\mathbf{y}_H,f)$, for acoustic propagation from $(\mathbf{y}_H,0)$ to $(\mathbf{x}_H,z)$ as well as the source spectrum, $\mathcal{D}_\sigma(f)$, are required. Let $|S|$ be the area of the source region interface S. As a consequence of the assumption that the source spectrum is constant one has

$$\langle|\hat{P}_\sigma(\mathbf{x}_H,z,2\pi f)|^2\rangle_T = |S|\mathcal{Q}_\sigma(\mathbf{x}_H,\mathbf{x}_H,z,z,\mathbf{0},f)\mathcal{D}_\sigma(f).$$

The term $\sqrt{\mathcal{Q}_a(\mathbf{x}_H,\mathbf{x}_H,0,0,\mathbf{y}_H,f)}$ will be referred to here as the propagation factor so that to obtain the received microbarom power spectral density one simply takes the product of source strength squared, propagation factor squared, and area $|S|$. Note that −20 times the logarithm of the propagation factor is generally called the transmission loss.[28]

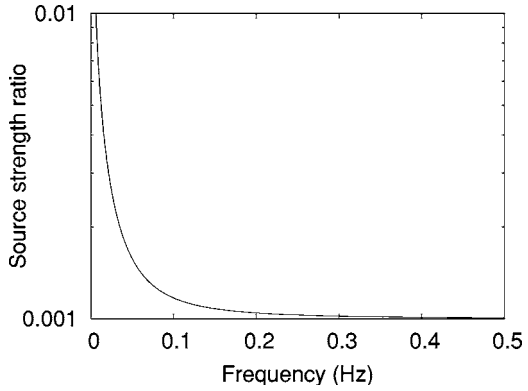


FIG. 5. The ratio of the microbarom source strength spectra for radiation into the atmosphere versus radiation into the ocean.

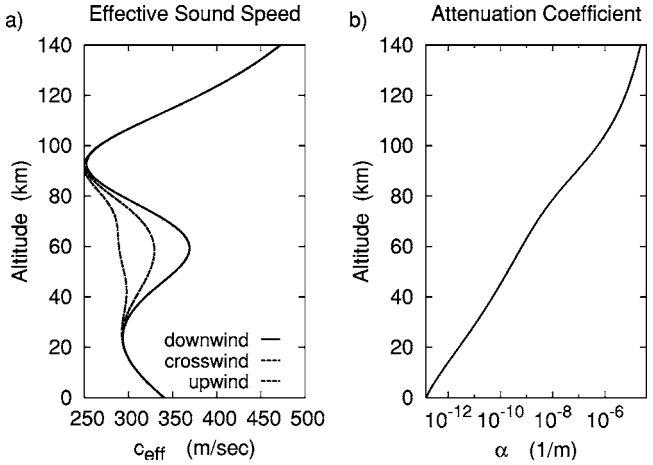


FIG. 6. (a) The model effective sound speed for downwind, crosswind, and upwind propagation. (b) The classical attenuation coefficient, as a function of altitude.

A. Propagation model

To model the propagation a stratified model will be used.[29] In this model the mean thermodynamic properties of the atmosphere, mean temperature, mean pressure, mean density ρ_a, and mean entropy, are assumed to depend only on altitude. It is assumed that the mean winds $\mathbf{v}_0$ have no vertical component and that the horizontal components again depend only on altitude.[30,31] The frequencies of interest here are much larger than the Brunt-Väisälä frequency in the atmosphere so that the effects of buoyancy can be ignored. Thus one may set $g=0$ in the equation for the Green's function. To simplify the model further the effective sound speed approximation, in which the wind velocity in the direction of horizontal propagation,

$$\hat{\mathbf{k}} = \frac{\mathbf{x}_H - \mathbf{y}_H}{|\mathbf{x}_H - \mathbf{y}_H|},$$

is added to the adiabatic sound speed, is used. The effective sound speed approximation is valid for low angle propagation.

Atmospheric attenuation is included by adding an imaginary part, an attenuation coefficient $\alpha(z)$,[32] to the wave number. Here only the classical attenuation coefficient, that due to thermo-viscous effects alone, will be used.

The resulting equation for the Green's function for atmospheric propagation, $\hat{G}_a$, is[30,31,33]

$$\left(\nabla_H^2 + \rho_a \frac{\partial}{\partial z}\frac{1}{\rho_a}\frac{\partial}{\partial z} + \left(\frac{\omega}{c_{\text{eff}}} + i\alpha(z)\right)^2\right)$$

$$\times \hat{G}_a(\mathbf{x}_H, z, \mathbf{y}_H, z', \omega) = \delta(\mathbf{x}_H - \mathbf{y}_H)\delta(z - z')$$

for $z>0$ with

$$c_{\text{eff}}(z) = c_a(z) + \hat{\mathbf{k}} \cdot \mathbf{v}_0(z)$$

and with rigid ground boundary conditions

$$\left.\frac{\partial P_a(\mathbf{x}_H, z, \omega)}{\partial z}\right|_{z=0} = 0$$

taken at $z=0$. For the purposes of this example the sound speed c_a is taken to be given by the polynomial fit to the "Standard Atmosphere"[34] given in Ref. 35. To simulate downwind versus upwind propagation a simple model for wind speed is used: $\hat{\mathbf{k}} \cdot \mathbf{v}_0(z)$ is assumed to have a Gaussian profile centered at the stratopause (an altitude of 60 km in the model used here) and 17.5 km wide. Explicitly,

$$\hat{\mathbf{k}} \cdot \mathbf{v}_0(z) = \pm 40 \cdot \exp\left[-\left(\frac{z - 6.0 \times 10^4}{1.75 \times 10^4}\right)^2\right] \text{m/s}.$$

The effective sound speeds that result for downwind, upwind, and crosswind propagation, as well as the classical attenuation coefficient, are plotted in Fig. 6.

To solve for $\hat{G}_a$ one may expand in vertical normal modes.[28,36] At long ranges from the source ($\mathbf{x}_H$ and $\mathbf{y}_H$ widely separated) one has

$$\hat{G}_a(\mathbf{x}_H, z, \mathbf{y}_H, z', \omega) = \frac{1}{\rho_a(z')}\sqrt{\frac{i}{8\pi|\mathbf{x}_H - \mathbf{y}_H|}}\sum_j \frac{e^{ik_j|\mathbf{x}_H - \mathbf{y}_H|}}{\sqrt{k_j}}\psi_j(z)\psi_j(z') \tag{65}$$

with

$$\left(\rho_a \frac{\partial}{\partial z}\frac{1}{\rho_a}\frac{\partial}{\partial z} + \left(\frac{\omega}{c_{\text{eff}}} + i\alpha(z)\right)^2 - k_j^2\right)\psi_j(z) = 0$$

and $\psi_j'(0)=0$. The modes satisfy the bi-orthogonality (no complex conjugation) condition

$$\int_0^\infty \psi_j(z)\psi_n(z)\frac{1}{\rho_a(z)}\,dz = \delta_{jn}.$$

The condition that this integral converges for $j=n$ uniquely determines the mode numbers.

Rather than the Green's function, the cross spectral density $\mathcal{Q}$ from (60) is required. For a stationary atmosphere the cross spectral density is the product of two Green's functions. In this case the propagation factor is simply the magnitude $|G_a(\mathbf{x}_H, 0, \mathbf{y}_H, 0, 2\pi f)|$ of the Green's function. For a fluctuating atmosphere a simple model for the average over the propagation medium $\langle\cdot\rangle_P$ is obtained by assuming that the different modes have statistically independent, uniformly distributed phases. Under such assumptions the square of the propagation factor for surface to surface propagation is given by the "incoherent modal sum"[28]

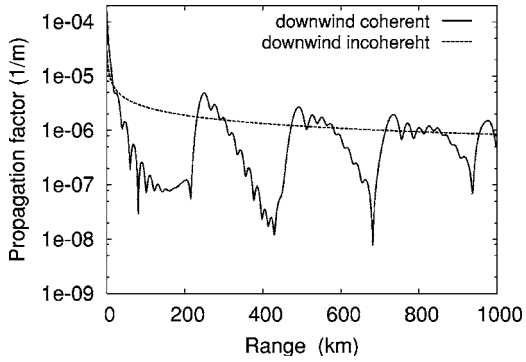


FIG. 7. The propagation factor for downwind propagation. Both the coherent modal sums, representing propagation in a stationary atmosphere, and the incoherent modal sum, representing the mean for propagation in a fluctuating atmosphere, are shown.

$$\mathcal{Q}_a(\mathbf{x}_H,\mathbf{x}_H,0,0,\mathbf{0},f)=\frac{1}{8\pi\rho_a(0)^2|\mathbf{x}_H|}\cdot\sum_j\frac{e^{-2\ \mathrm{Im}\ k_j|\mathbf{x}_H|}}{|k_j|}|\psi_j(0)^4|. \tag{66}$$

The mode functions ψ_j and wave numbers k_j have been determined using the finite difference methods described in Ref. (28). The modal attenuation coefficients, $\mathrm{Im}\,k_j$, are determined perturbatively: the modes are determined for the lossless case $\alpha=0$ and then first-order perturbation theory[36] is used to approximate the imaginary parts of the modal wave numbers. Both upwind and crosswind the surface-to-surface propagation paths must pass through the thermosphere (the region above 120 km). In these cases there is a severe reduction in predicted levels as a consequence of the increased attenuation in the thermosphere. While this is consistent with observations[1,9,11,12,37] the numerical values for the propagation as predicted by such linear models should not be taken literally since, in the thermosphere, the density ρ_a decreases dramatically so that the accuracy of the linear approximation to the atmospheric response is doubtful.[38] For downwind propagation, however, the sound gets trapped in the duct formed by the stratosphere (below 60 km) so that received levels are not greatly influenced by the thermosphere. Thus, for downwind propagation a linear propagation model is reasonable. In Fig. 7 the propagation factor for the downwind model is plotted as a function of range $|\mathbf{x}_H|$ at frequency $f=0.2$ Hz.

B. Sea state model

Now consider the sea state angular integral (64). It is common to write $F(f,\theta)$ as[15]

$$F(f,\theta)=\bar{F}(f)a(f,\theta)$$

where $\bar{F}(f)$ is the frequency spectrum given by (4). For purposes of illustration it will be assumed here that the frequency spectrum $\bar{F}(f)$ can be taken to be given by the JONSWAP[15,16] model of highly excited seas, that both $\bar{F}(f)$ and $a(f,\theta)$ are independent of position in the source region, and that $a(f,\theta)=a(\theta)$ is independent of frequency. Then one has

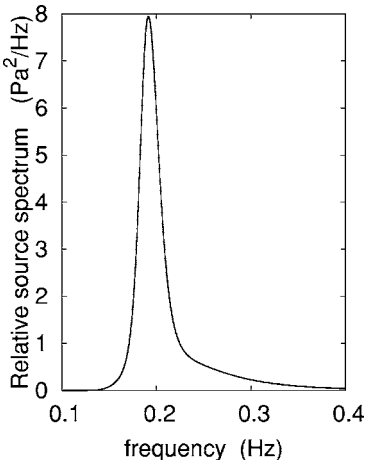


FIG. 8. The relative source spectrum for atmospheric microbaroms, $\mathcal{D}_a(f)/b$.

$$\int_0^{2\pi}F\left(\frac{f}{2},\theta\right)F\left(\frac{f}{2},\theta+\pi\right)d\theta=\bar{F}\left(\frac{f}{2}\right)^2\int_0^{2\pi}a(\theta)a(\theta+\pi)\,d\theta.$$

If $\int_0^{2\pi}a(\theta)a(\theta+\pi)d\theta=b$, then

$$\mathcal{D}_a(f)=4b\bar{F}\left(\frac{f}{2}\right)^2\frac{\rho_a^2g^2\pi^4f^3}{c_a^2}\left(\frac{9g^2}{4\pi^2c_a^2f^2}+\frac{c_a^2}{c_w^2}\right).$$

In Fig. 8 the atmospheric microbarom source strength squared divided by b, $\mathcal{D}_a(f)/b$, is plotted for a significant wave height of 10 m and peak frequency of 0.95 Hz. Referring to Figs. 8 and 7, one finds that to produce a microbarom signal with amplitude of 0.1 $\mathrm{Pa}/\sqrt{\mathrm{Hz}}$ at the peak frequency 1000 km from the source region would require

$$\left(10^{-12}\frac{1}{\mathrm{m}^2}\right)\left(8\frac{\mathrm{Pa}^2}{\mathrm{Hz}}\right)b|S|=0.01\frac{\mathrm{Pa}^2}{\mathrm{Hz}}.$$

This gives $b|S|\approx10^9$ m^2. Assuming the source region to be a disk of radius 200 km gives $b\approx0.01$. For comparison, note that assuming the sea to be isotropic gives $b=2\pi$. It should be emphasized, however, that the value for b obtained here is illustrative only. The forms used for both the propagation factor and the source strength are based on simplifying assumptions: for the propagation factor these are the use of the standard model for the atmospheric temperature and the exponential profile for the atmospheric wind; for the source strength these are the use of the JONSWAP model for the frequency spectrum and the assumption that the sea state statistics are constant over the source region.

V. DISCUSSION OF THE RESULTS

A complete solution of the problem of the radiation of infrasound by ocean waves, the so-called microbarom radiation, has been presented. A direct connection is made between the statistical properties of the microbarom signal and the stochastic models commonly used to describe ocean wave systems. Since the acoustic wavelengths are so much longer than the correlation lengths for the ocean waves the received microbarom signal can be described as an incoherent superposition of fields produced by monopole sources at

the sea surface. The problem of finding the source strength of the radiation has been reduced to oceanography: the determination of the density of counter-propagating waves of equal frequency (standing waves) on the ocean surface, as given by the integral (64), is required.

The source strength spectral density function for the radiation of atmospheric microbaroms has been derived and compared to that for oceanic microbaroms. The primary radiation is into the ocean; the source strength for radiation into the atmosphere is three orders of magnitude smaller than that for radiation into the ocean. This is in accord with observations: typical peak atmospheric microbarom levels are measured in the tenths of Pascals while typical peak oceanic microbarom levels can reach 100 Pascals. The reason for this difference lies in the fact that sea water is three orders of magnitude denser than air. The velocity of the air/water interface is fixed by the ocean waves. Consequently, the air and water have similar velocity fields near the interface. However, the variations in pressure required to maintain variations in velocity are proportional to the density of the fluid. Thus the pressure fluctuations in the ocean are three orders of magnitude larger than those in the air. Indeed, the ratio of the atmospheric to the oceanic microbarom source strengths is shown to be approximately equal to the ratio of the atmospheric to the oceanic densities.

The full atmospheric microbarom signal is shown to be the sum of two terms. The first term, responsible for about 80% of the radiation, represents sound radiated from pressure fluctuations produced in the ocean by the motion of the interface plus a term, responsible for the remaining 20% of the radiation, due to the compression of the air by the ocean waves. The single unknown factor, the standing wave density (64), is common to both the atmospheric and oceanic source strength spectra.

ACKNOWLEDGMENTS

We thank Allan Pierce and Konstantine Naugol'nykh for comments which led to a strengthening of the manuscript.

[1]W. L. Donn and B. Naini, "Sea wave origin of microbaroms and microseisms," J. Geophys. Res. **78**, 4482–4488 (1973).
[2]M. S. Longuet-Higgins, "A theory of the origin of microseisms," Philos. Trans. R. Soc. London **243**, 1–35 (1950).
[3]K. Hasselmann, "A statistical analysis of the generation of microseisms," Rev. Geophys. **1**, 177–210 (1963).
[4]E. S. Posmentier, "A theory of microbaroms," Geophys. J. R. Astron. Soc. **13**, 487–501 (1967).
[5]I. M. Brekhovskikh, V. V. Goncharov, V. M. Kurtepov, and K. A. Naugol'nykh, "The radiation of infrasound into the atmosphere by surface waves in the ocean," Atmos. Oceanic Phys. **9**, 899–907 (1973).
[6]S. Arendt and D. C. Fritts, "Acoustic radiation by ocean surface waves," J. Fluid Mech. **415**, 1–21 (2000).
[7]S. C. Webb, "The equilibrium oceanic microseism spectrum," J. Acoust. Soc. Am. **92**, 2141–2158 (1992).
[8]S. C. Webb and C. S. Cox, "Observations and modeling of seafloor microseisms," J. Geophys. Res. **91**, 7343–7358 (1986).
[9]W. L. Donn and D. Rind, "Microbaroms and the temperature and wind of the upper atmosphere," J. Atmos. Sci. **29**, 156–172 (1973).
[10]M. Garces, M. Willis, C. Hetzer, A. L. Pinchon, and D. Drob, "On using ocean swells for continuous infrasonic measurements of winds and temperature in the lower, middle and upper atmosphere," Geophys. Res. Lett. **31**, L018614 (2004).
[11]M. Willis, M. Garces, C. Hetzer, and S. Busnger, "Infrasonic observations of open ocean swells in the Pacific," Geophys. Res. Lett. **31**, L019303 (2004).
[12]A. Le Pinchon, V. Maurer, D. Raymond, and O. Hyvernaud, "Infrasonic from ocean waves observed in Tahiti," Geophys. Res. Lett. **31**, L019103 (2004).
[13]J. R. Bowman, G. E. Baker, and M. Bahavar, "Ambient infrasound noise," Geophys. Res. Lett. **32**, L09803 (2005).
[14]B. Kinsman, *Wind Waves* (Prentice Hall, Englewood Cliffs, NJ, 1965).
[15]G. J. Komen, L. Cavaleri, M. Dronelan, K. Hasselmann, S. Hasselmann, and P. A. E. M. Janssen, *Dynamics and Modeling of Ocean Waves* (Cambridge, U.P., Cambridge, 1996).
[16]K. Hasselmann, T. P. Barnett, E. Bouws, H. Carlson, D. E. Cartwright, K. Enke, J. A. Ewing, H. Gienapp, D. E. Hasselmann, P. Kruseman, A. Meerburg, P. Müller, D. J. Olbers, K. Richter, W. Sell, and H. Walden, "Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap)," Dtsch. Hydrogr. Z. Suppl. A **8**, 95 (1973).
[17]H. Lamb, *Hydrodynamics* (Dover, New York, 1945).
[18]L. D. Landau and I. M. Lifshitz, *Fluid Mechanics* (Pergamon, Oxford, 1959).
[19]R. T. Beyer, "The parameter b/a," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), Chap. 2.
[20]J. H. Ginsberg, "Perturbation methods," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), Chap. 10, pp. 281–282.
[21]M. F. Hamilton and C. L. Morfey, "Model equations," in *Nonlinear Acoustics*, edited by M. F. Hamilton and D. T. Blackstock (Academic, New York, 1998), Chap. 3.
[22]J. S. Bendat and A. G. Piersol, *Random Data, Analysis and Measurement Procedures*, 3rd ed. (Wiley, New York, 2000).
[23]J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).
[24]This may be verified by writing $\mathbf{K}=k_0\mathbf{Q}$ and changing integration variables from y_H and y'_H to $y_H+y'_H$ and $(\omega_0/c_\sigma)(y_H-y'_H)$.
[25]J. V. Olsen and C. A. L. Szuberla, "Distribution of wave packet sizes in microbarom wave trains observed in Alaska," J. Acoust. Soc. Am. **117**, 1032–1037 (2005).
[26]M. A. Donelan, J. Hamilton, and W. H. Hui, "Directional spectra of wind generated waves," Philos. Trans. R. Soc. London, Ser. A **315**, 509–562 (1985).
[27]C. W. Wright, E. J. Walsh, D. Vandemark, W. Krabill, A. W. Garcia, S. H. Houston, M. D. Powell, P. G. Black, and F. D. Marks, "Hurricane directional wave spectrum spatial variation in the open ocean," J. Phys. Oceanogr. **31**, 2472–2488 (2001).
[28]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, New York, 2000).
[29]A. D. Pierce, "Propagation of acoustic-gravity waves in a temperature- and wind-stratified atmosphere," J. Acoust. Soc. Am. **37**, 218–227 (1965).
[30]V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E & FN Spon, London, 1997).
[31]D. C. Pridmore-Brown, "Sound propagation in a temperature- and wind-stratified medium," J. Acoust. Soc. Am. **34**, 438–443 (1962).
[32]L. C. Sutherland and H. E. Bass, "Atmospheric absorption in the atmosphere up to 160 km," J. Acoust. Soc. Am. **99**, 1012–1032 (2004).
[33]S. N. Kulichkov, K. V. Avilov, O. E. Popov, A. I. Otrezov, G. A. Bush, and A. K. Baryshnikov, "Some results of simulation of long-range infrasonic propagation in the atmosphere," Izv. Russ. Acad. Sci. Atmos. Oceanic Phys., Eng. Transl., **40**, 202–215 (2004).
[34]*U. S. Standard Atmosphere* (U.S. Government Printing Office, Washington, DC, 1976).
[35]J. F. Lingevitch, M. D. Collins, and W. L. Siegmann, "Parabolic equations for gravity and acousto-gravity waves," J. Acoust. Soc. Am. **105**, 3049–3056 (1999).
[36]A. D. Pierce, "Guided infrasonic modes in a temperature- and wind-stratified atmosphere," J. Acoust. Soc. Am. **41**, 597–611 (1967).
[37]D. Rind and W. L. Donn, "Further use of natural infrasound as a continuous monitor of the upper atmosphere," J. Atmos. Sci. **32**, 1694–1704 (1975).
[38]S. Kulichkov, "Nonlinear acoustic phenomena in atmosphere," in *Nonlinear Acoustics at the Beginning of the 21st Century*, edited by O. V. Rudenko and O. A. Sapozhnikov, 16th International Symposium on Nonlinear Acoustics (ISNA, Moscow, 2002), pp. 251–258.

Time-domain impedance boundary conditions for surfaces with subsonic mean flows

X. D. Li[a)]
School of Jet Propulsion, BeiHang University, Xueyuan Road 37, Beijing, 100083, People's Republic of China

C. Richter[b)] and F. Thiele[c)]
Hermann-Föttinger-Institute of Fluid Mechanics, Berlin University of Technology, Strasse des 17. Juni 135, 10623, Berlin, Germany

(Received 21 September 2005; revised 8 March 2006; accepted 9 March 2006)

The time-domain impedance boundary conditions of Tam and Auriault are extended for impedance boundaries with subsonic mean flows. This extension requires an effective impedance. A model for the effective impedance is proposed, which guarantees the continuity of particle displacement over an infinitely thin shear layer connecting the fluid to the impedance boundary as described by Myers. Together with the effective impedance model, the time domain impedance boundary condition renders the requirement of measuring the effective impedance under flow conditions unnecessary. To demonstrate the feasibility of the extended boundary conditions, a numerical validation using the NASA Grazing Incidence Tube measurements of Jones *et al.* is carried out. The assumed velocity on the infinitely thin shear layer connecting the flow over the surface to the locally reacting wall is shown to have a strong influence on the performance of the liner. Another important parameter identified is the shape of the mean flow profile, which not only causes convective effects but also may excite possible flow instabilities. It is demonstrated that satisfactory results can be obtained using the extended time domain impedance boundary condition based on the correct effective impedances and mean flow profiles. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2191610]

PACS number(s): 43.28.Py [AJZ] Pages: 2665–2676

I. INTRODUCTION

Acoustic lining is a key technology in the reduction of sound radiation from flow ducts. It is found in simple air conditioning systems as well as in commercial aeroengines. Consequently, the development of tools suitable for the prediction of sound propagation in lined flow ducts has been a very important subject in acoustics for some time. The liner impedance arising from a simple mechanical or acoustic model for the locally reacting wall is defined in the frequency domain and normally used for frequency-domain analysis. According to Hamilton and Astley[1] as well as Li *et al.*,[2] for a large number of aeroacoustics problems, e.g., large 3D, broadband, nonlinear, and transient problems, it is preferable to use time-domain methods.

Time-domain impedance boundary conditions (TDIBCs) appear separately in electromagnetism and acoustics. So far, several different TDIBCs have been proposed for computational aeroacoustics (CAA). Tam and Auriault[3] proposed both a single-frequency and a broadband TDIBC and have proved the well-posedness for cases without flow on porous surfaces. Özyörük and Long[4] proposed the implementation of a convective-modified impedance in the time domain via z-transformation. Fung and Ju[5,6] constructed stable TDIBCs using a reflection coefficient instead of the impedance. The detailed history and theoretical and numerical issues of TDIBCs have been recently reviewed by Fung and Ju.[7]

All three kinds of TDIBCs have been validated using the NASA Grazing Incidence Tube (GIT) data.[8] Özyörük and Long[4] and Fung and Ju[5,6] validated their TDIBC using several flow conditions showing good agreement.[9–11] Very good agreement compared to the cases measured in the NASA GIT with zero mean flow on the impedance surface[12] and for multi-dimensional problems[13] has been demonstrated by Zheng and Zhuang using TDIBCs of Tam and Auriault.[3] However, for some cases Özyörüz *et al.*[9] as well as Ju and Fung[11] reported that impedance jumping between the hard and soft walls could result in amplified numerical instabilities. This type of instability was first found by Tester,[14] who showed the convective impedance boundary condition could lead to an unstable solution in the study of a lined duct with mean flow. Tam and Auriault[3] proved that the existence of a Kelvin-Helmholtz-type instability renders the initial value problem ill-posed for the time-domain solution by the direct implementation of the convective impedance boundary condition. To overcome this difficulty, they suggested a direct iteration model by measuring the effective impedance in the presence of mean flow. However, possibly due to the problems in measuring the effective impedance described below, no comprehensive proof was provided in their paper. Özyörük and Long[10] found that the numerical instabilities can be removed using a convective impedance with a shear

[a)]Electronic mail: lixd@buaa.edu.cn
[b)]Electronic mail: christoph.richter@cfd.tu-berlin.de
[c)]Electronic mail: frank.thiele@cfd.tu-berlin.de

layer assumption. Ju and Fung[11] proposed and validated that the use of an effective plane-wave impedance in their time-domain method was unconditionally stable even for a slip flow. However, the drawbacks of the use of effective plane-wave impedance[11] are that the incident angle of sound waves must be assumed as known before the calculation and the curvature effects cannot be accounted for. In fact, the use of an effective impedance is a very useful idea which can be traced back to even earlier works by Ingard,[15] Goldstein and Rice,[16] and Nayfeh *et al.*[17] The present work is motivated by several works referring to the effective impedance[3,11,15–17] and the widely used impedance boundary condition by Myers based on the continuity of particle displacement.[18]

The main objective of this paper is to extend Tam and Auriault's TDIBCs[3] for arbitrary subsonic mean flow profiles using the effective impedance and to provide a method to calculate the effective impedance Z' from the flow field and the flow impedance Z. To overcome the requirement of measuring the effective impedance, the impedance boundary condition by Myers[18] based on the continuity particle displacement is resorted. The next section will give the mathematical formulation of the extended TDIBCs and provide a stability analysis on their well-posedness. In Sec. III the mathematical model and numerical algorithm for the NASA GIT using the extended TDIBCs are presented. The corresponding numerical results are presented in Sec. IV and compared against the latest GIT measurement results by Jones *et al.*[19] The concluding remarks are finally given in Sec. V.

II. TDIBCS FOR SURFACES WITH SUBSONIC MEAN FLOWS

Throughout this paper a $e^{-i\omega t}$ convention applies and the surface normal vector $\mathbf{n}$ is defined positive when pointing into the surface. The effective impedance for arbitrary subsonic mean flow profiles is used to provide a TDIBC based on the idea of Tam and Auriault,[3] which is shown to be stable for subsonic mean flows. A detailed analysis of the stability is given in Appendix A. Both the single-frequency TDIBC and the broadband time domain impedance boundary condition are used in this way, extending the applicability of the original TDIBC, while the well-posedness remains guaranteed.

A. Single-frequency TDIBC for surfaces with subsonic mean flows

Using the effective impedance within the definition of the single-frequency one-dimensional TDIBCs by Tam and Auriault[3] to allow subsonic mean flow profiles leads to

$$\frac{\partial p}{\partial t} = R' \frac{\partial v_n}{\partial t} - X' \omega v_n, \quad X' < 0, \tag{1}$$

$$p = R' v_n + \frac{X'}{\omega} \frac{\partial v_n}{\partial t}, \quad X' \geqslant 0. \tag{2}$$

Equations (1) and (2) are equivalent to the frequency-domain effective impedance boundary condition $Z' = p/v_n$, with $(Z' = R' - iX')$ for a sound field of the single frequency ω. Contrary to the zero mean flow assumption required on the impedance wall used by Tam and Auriault, the effective impedance Z' is defined under subsonic mean flow conditions. Thus it is necessary to measure the effective impedance in the presence of mean flow, or preferably to try to represent it by a model with mean flow effects accounted for. In general Z' can be written as a function of the flow impedance Z, angular frequency ω, effective mean flow Mach number $\mathbf{M_e}$, and the intensity of sound: $Z' = Z'(Z, \omega, \mathbf{M_e}, \mathrm{SPL})$, where the flow impedance only depends on the magnitude of the effective flow Mach number on the surface. For zero mean flow on the impedance surface, Eqs. (1) and (2) reduce to the original single-frequency TDIBCs by Tam and Auriault and the well-posedness has already been proved[3] in their investigation. The analysis on the well-posedness of Eqs. (1) and (2) is addressed in Appendix A.

Considering only an effective axial mean flow Mach number $\mathbf{M_e}$, the nondimensionalized linearized Euler equations (LEEs) on the surface of the liner can be written as

$$\frac{\partial v}{\partial t} + \mathbf{M_e} \frac{\partial \mathbf{v}}{\partial x} = -\nabla p \tag{3}$$

$$\frac{\partial p}{\partial t} + \mathbf{M_e} \frac{\partial p}{\partial x} + \frac{\partial \mathbf{v}}{\partial x} = 0. \tag{4}$$

The extended single-frequency TDIBC of Eqs. (1) and (2) can also be formulated in the following form for numerical implementation through the combination of Eq. (3) with Eq. (1), and Eq. (4) with Eq. (2):

$$\nabla_n p = \frac{1}{R'} \left(\frac{\partial \mathbf{v}}{\partial x} - X' \omega v_n \right) + \frac{\mathbf{M_e}}{R'} \frac{\partial p}{\partial x} - \mathbf{M_e} \frac{\partial v_n}{\partial x}, \quad X' < 0, \tag{5}$$

$$\nabla_n p = \frac{\omega}{X'} (R' v_n - p) - \mathbf{M_e} \frac{\partial v_n}{\partial x}, \quad X' \geqslant 0. \tag{6}$$

Given $\mathbf{M_e} = \mathbf{0}$ on the impedance wall, Eqs. (5) and (6) will be reduced to the original implementation form of Tam and Auriault[3] as

$$\nabla_n p = \frac{1}{R} \left(\frac{\partial \mathbf{v}}{\partial x} - X \omega v_n \right), \quad X < 0, \tag{7}$$

$$\nabla_n p = \frac{\omega}{X} (R v_n - p), \quad X \geqslant 0. \tag{8}$$

B. Broadband TDIBC for surfaces with subsonic mean flows

The three-parameter broadband TDIBC by Tam and Auriault[3] can be applied for subsonic mean flows and written as

$$\frac{\partial p}{\partial t} = R'_0 \frac{\partial v_n}{\partial t} - X'_{-1} v_n + X'_1 \frac{\partial^2 v_n}{\partial t^2}. \tag{9}$$

This is equivalent to a frequency-domain impedance boundary condition with the effective impedance Z' defined by

$$Z' = R'_0 - i[X'_{-1}/\omega + X'_1\omega], \tag{10}$$

where the valid range of the three parameters is given by $R'_0 > 0$, $X'_{-1} < 0$ and $X'_1 > 0$ from the transfer of the causal impedance and reactance conditions by Tam and Auriault[3] to the effective impedance.

In the same way as for the extended single-frequency TDIBC [Eqs. (1) and (2) and Eqs. (5) and (6)], the effective impedance can be defined as $Z' = Z'(Z, \omega, \mathbf{M_e}, \mathrm{SPL})$. For $\mathbf{M_e} = \mathbf{0}$, the extended broadband TDIBC will be reduced to the form given by Tam and Auriault.[3] It can also be proved that the proposed effective impedance in the broadband TDIBC Eq. (9) would lead to well-posed problems with the detailed stability analysis given in Appendix B. In the following, two different numerical implementation forms are presented, where the first is used for the computations of this paper and the second is chosen to show the link to the original form used by Tam and Auriault.[3]

Divided by $-i\omega$, then transformed back to the time domain using the auxiliary variable q, $\partial q/\partial t = v_n$, the broadband TDIBC Eq. (9) can be formulated into the implementation form used in this paper:

$$\frac{\partial v_n}{\partial t} = \frac{1}{X'_1}[p - R'_0 v_n + X'_{-1} q]. \tag{11}$$

The condition for the time derivative of v_n is used to calculate the pressure derivative normal to the wall by inserting Eq. (11) into the LEE multiplied by the impedance wall normal vector,

$$\boldsymbol{\nabla}_n p = -\left(\mathbf{M_e} \cdot \boldsymbol{\nabla}_n \mathbf{v} + \frac{\partial v_n}{\partial t}\right). \tag{12}$$

The original method by Tam and Auriault used $Q = \partial v_n/\partial t$, and to consider the LEE of Eqs. (3) and (4), in which case the extended broadband TDIBC can also be written as

$$\frac{\partial}{\partial t}\begin{bmatrix} v_n \\ Q \end{bmatrix} = \begin{bmatrix} Q \\ \dfrac{1}{X'_1}(-\mathbf{M_e} \cdot \boldsymbol{\nabla} p - \boldsymbol{\nabla} \cdot \mathbf{v} - R'_0 Q + X'_{-1} v_n) \end{bmatrix}. \tag{13}$$

Again Eq. (12) will be used to calculate the pressure derivative with a given Q. Assuming a zero effective mean flow on the impedance wall, Eqs. (12) and (13) can be simplified to the original forms by Tam and Auriault[3] as

$$\frac{\partial}{\partial t}\begin{bmatrix} v_n \\ Q \end{bmatrix} = \begin{bmatrix} Q \\ \dfrac{1}{X_1}(-\boldsymbol{\nabla} \cdot \mathbf{v} - R_0 Q + X_{-1} v_n) \end{bmatrix}, \tag{14}$$

$$\boldsymbol{\nabla}_n p = -Q. \tag{15}$$

C. An effective impedance model based on the Myers boundary condition

It is shown that the extended single-frequency TDIBC of Eqs. (1) and (2) or Eqs. (5) and (6) and the extended broadband TDIBC of Eq. (9) or Eqs. (12) and (13) can admit subsonic mean flows without losing the well-posedness. However, the problem has only been shifted to the determination of the effective impedance considering the mean flow effects. The Myers boundary condition[18] is designed to describe the mean flow effects if only the thickness of the mean flow and the acoustic boundary layers are small compared to the acoustic wavelength, which can be written into the following dimensionless form:

$$\hat{v}_n = \hat{p}/Z + (i\omega)^{-1}\mathbf{M} \cdot \boldsymbol{\nabla}(\hat{p}/Z) - (i\omega Z)^{-1}\mathbf{n} \cdot (\mathbf{n} \cdot \boldsymbol{\nabla}\mathbf{M})\hat{p}, \tag{16}$$

where $\hat{v}_n$ denotes the normal velocity (which is, by the definition of $\mathbf{n}$, positive when pointing into the impedance wall), $\mathbf{M}$ is the grazing mean flow Mach number, $\hat{p}$ is the pressure fluctuation, and Z denotes the flow impedance. Equation (16) can be rewritten into the following form:

$$Z = \frac{\hat{p}}{\hat{v}_n} + \frac{Z}{i\omega\hat{p}}\frac{\hat{p}}{\hat{v}_n}\mathbf{M} \cdot \nabla\frac{\hat{p}}{Z} - \frac{\mathbf{n} \cdot (\mathbf{n} \cdot \boldsymbol{\nabla}\mathbf{M})}{i\omega}\frac{\hat{p}}{\hat{v}_n}. \tag{17}$$

Defining the effective impedance as $Z' = \hat{p}/\hat{v}_n$ in the presence of the mean flows, Eq. (17) can be written as

$$Z' = \frac{\hat{p}}{\hat{v}_n} = Z \Bigg/ \left[1 + \frac{Z}{i\omega\hat{p}}\mathbf{M} \cdot \boldsymbol{\nabla}\frac{\hat{p}}{Z} - \frac{\mathbf{n} \cdot (\mathbf{n} \cdot \boldsymbol{\nabla}\mathbf{M})}{i\omega}\right], \tag{18}$$

where Z' is the effective acoustic impedance and equals $R' - iX'$, where R' and X' represent the effective acoustic resistance and reactance, respectively. Z is the flow impedance according to Myers.[18] To account for the refraction effects from mean flows, an effective mean Mach number $\mathbf{M_e}$ is introduced. Using $\mathbf{M_e}$ Eq. (18) is formulated into

$$Z' = \frac{\hat{p}}{\hat{v}_n} = Z \Bigg/ \left[1 + \frac{Z}{i\omega\hat{p}}\mathbf{M_e} \cdot \boldsymbol{\nabla}\frac{\hat{p}}{Z} - \frac{\mathbf{n} \cdot (\mathbf{n} \cdot \boldsymbol{\nabla}\mathbf{M_e})}{i\omega}\right]. \tag{19}$$

In many cases a slip condition is employed in the acoustic calculations with the frequency domain boundary condition, which leads to a so-called plug flow with $\mathbf{M_e} = \mathbf{M}$ at the liner surface. For a two-dimensional plane impedance wall with curvature effects neglected, Eq. (19) can be simplified as

$$Z' = Z \Bigg/ \left[1 + \frac{Z}{i\omega\hat{p}}\mathbf{M_e} \cdot \boldsymbol{\nabla}(\hat{p}/Z)\right], \tag{20}$$

which can also be derived directly from the Ingard boundary condition.[15] Equations (19) and (20) constitute the basis for the calculation of effective impedances in this paper.

If assuming a plane harmonic wave in a plug flow in the x direction with unity amplitude ($e^{[i(k_x x - \omega t)]}$), Eq. (20) can be further simplified into the following form verified by Ju and Fung,[11]

$$\hat{v}_n = \hat{p}/Z', \tag{21}$$

where $Z' = Z/(1 - \mathrm{M}_x \cos\theta)$ denotes the effective plane-wave impedance with the effective incident wave angle $\theta = \cos^{-1}[k_x/\omega]$ depending on ω and M through k_x. The main advantage of using Eqs. (19) and (20) over Eq. (21) is that the incident angle of a sound wave has been trans-

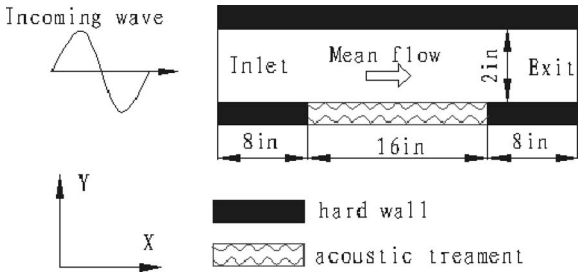


FIG. 1. Configuration of NASA Langley 2D grazing flow impedance tube.

ferred to an impact angle as part of the solution of the effective impedance.

D. Implementation of the extended TDIBC

Both the extended single-frequency and broadband TDIBCs can be validated. Here we only focus on the implementation of the broadband TDIBC. The key element of the implementation is on the calculation of the effective impedance by Eq. (19) in the time domain. Assuming that the impedance Z under flow conditions is known, only the denominator of Eq. (19) has to be calculated. Considering there are no curvature effects involved in the validation configuration of the NASA GIT as shown in Fig. 1, this paper will focus on the convective term in Eq. (19). However, the extension of the presented method to the curvature term is straightforward. The effective impedance is initially defined in the frequency domain; the key idea is to do the transfer to the time domain after calculating the complex amplitude ratio within the convective term. As the spectra are very simple for a single frequency, it is possible to calculate the complex amplitude ratio within the convective term by a ratio of amplitudes obtained in the time domain. The complex pressure amplitude is present in both numerator and denominator. As the time and length scales are coupled in acoustic waves, the ratio is real if the wave number vector is real. Unfortunately the wave number vector along the impedance wall is complex, describing the decay along the liner. Therefore the ratio may contain a phase shift. If no cutoff or close to cutoff modes are present, the phase shift is small and may be neglected. The convective term includes $(i\omega\hat{p})^{-1}\nabla\hat{p}$, which is interpreted as the gradient of the time integral of $\hat{p}$ over $\hat{p}$. The convective term in Eq. (20) can be written in the following form:

$$\frac{Z}{i\omega\hat{p}}\mathbf{M_e}\cdot\nabla(\hat{p}/Z)=\frac{1}{\hat{p}}\mathbf{M_e}\cdot\nabla(\hat{P}), \tag{22}$$

where $\hat{p}/i\omega$ has been replaced by the auxiliary variable $\hat{P}$, which is calculated in the time domain by an integration:

$$\frac{\partial\hat{P}}{\partial t}=-\hat{p}. \tag{23}$$

Among the many possibilities, this is numerically the easiest to use, as the broadband TDIBC without flow would also require the integral. In general integral quantities add stability to the computation, as they arise from filtered field variables, which contain no short waves. For the presented results, the gradient operator is applied after the integration and requires the overhead of calculating $\hat{P}$ in the whole computational domain. If the gradient is applied first and the time integration afterwards, the computational effort could be significantly reduced especially for large node numbers.

As discussed above, for a single frequency it is possible to replace the complex amplitude ratio of Eq. (22) by the amplitudes of numerator and denominator. Assuming the phase angle of the convective term is neglectable, the root mean squares (rms) may be used. This yields the following replacement for the convective term:

$$\frac{1}{\hat{p}}\mathbf{M_e}\cdot\nabla(\hat{P})=\left|\mathbf{M}\cdot\frac{\mathbf{k}}{\omega}\right|=\frac{[\mathbf{M}\cdot\nabla\hat{P}]_{\mathrm{rms}}}{\hat{p}_{\mathrm{rms}}}. \tag{24}$$

As the rms operation allows only arcs between 0 and π, the scalar product has to be multiplied by -1 or 1 for downstream and upstream propagating waves, respectively. The root mean square values of both parts are calculated with the time simulation. As the simulation starts from $t=0$ with the undisturbed average flow field, the calculation of the rms values also starts with zero. The rms value for the new time level $n+1$ is obtained by the value at the old time level n for both parts, e.g.,

$$\hat{p}_{\mathrm{rms}}^{n+1}=\sqrt{(n[\hat{p}_{\mathrm{rms}}^{n}]^2+[\hat{p}^n]^2)/(n+1)}. \tag{25}$$

Only the two root mean square values have to be stored; the whole time history is included. Inserting the discussed denominator into Eq. (20) one obtains

$$Z'=Z\Bigg/\left(1+\frac{[\mathbf{M}\cdot\nabla\hat{P}]_{\mathrm{rms}}}{\hat{p}_{\mathrm{rms}}}\right), \tag{26}$$

which is used in connection with Eqs. (25) and (23) to calculate the effective impedance in the time domain. At the present time, the presented extension is theoretically limited to single frequency and mode of wave propagation only.

Together with Eq. (11) the effective impedance is used to calculate $\partial v_n/\partial t$. Inserted into the momentum equation, this provides a condition to the pressure gradient:

$$-\mathbf{n}\cdot\nabla p+\mathbf{n}\cdot\nabla p\Big|_{p_{\mathrm{ghost}}=0}=\frac{1}{X_1'}[p-R_0'v_n+X_{-1}'q]-\left.\frac{\partial v_n}{\partial t}\right|_{p_{\mathrm{ghost}}=0}. \tag{27}$$

The gradient is used in connection to the definition of the finite difference scheme at the wall to obtain $p_{\mathrm{ghost}}=0$:

$$\frac{1}{X_1'}[p-R_0'v_n+X_{-1}'q]-\left.\frac{\partial v_n}{\partial t}\right|_{p_{\mathrm{ghost}}=0}=-a_7^{51}p_{\mathrm{ghost}}. \tag{28}$$

In the special case of plane waves it is simple to prove the method. In this special case, the free field wave number k is dependent on the effective flow Mach number M_e and linearly dependent on $i\omega$. The convective term discussed above represents the cosine of the impact angle of the acoustic wave with respect to the mean flow direction in the frequency domain. The liner can change this propagation direc-

tion and **k** is a vector in general. The dot product must be obtained with the calculation of the acoustic field.

III. NUMERICAL ALGORITHM

A. Configuration of the NASA grazing incidence tube

The test section of the NASA Langley GIT[8] is selected to validate Eqs. (5) and (6) and (11)–(13). Since the shortest acoustic wavelength considered is much larger than the span of the test sections, the three-dimensional flow-impedance tube can be simplified into a two-dimensional configuration as shown in Fig. 1. The GIT consists of hard-walled inflow and outflow sections with the impedance wall section between. Due to this setup, the impedance jumps from the inlet to the liner and from the liner to the outlet section. In the simulation this jump is not considered in terms of derivatives within the formulation of the effective impedance. The effects due to the mismatch of hard-walled modes and soft-walled modes at these two interfaces are considered by a block interface in the computation.

The inflow and outflow regions are assumed anechoic. The sound source impedance is not known from the experiment, therefore a sponge layer in combination with the sound source is used as an anechoic termination. The outflow impedance was measured and it is shown that the outflow is only approximately anechoic under flow conditions.[19] Jones *et al.* state a strong influence of the assumed termination impedance on the educed impedance.[19] However, the conditions of the impedance eduction would approximately be reproduced by a anechoic termination. This avoids errors due to interactions of the imperfectly anechoic sponge layer at the inflow and the termination impedance. Therefore, at the outflow, a nonreflective perfectly matched layer[20] boundary condition is used.

B. Governing equations

The two-dimensional linearized Euler equations for nonuniform flow are utilized as the governing equations which can be written in the following form,

$$\frac{\partial \mathbf{U}}{\partial t} + \mathbf{A_x}\frac{\partial \mathbf{U}}{\partial \mathbf{x}} + \mathbf{A_y}\frac{\partial \mathbf{U}}{\partial \mathbf{y}} + (\mathbf{B_x} + \mathbf{B_y})\mathbf{U} = \mathbf{S}, \tag{29}$$

where

$$\mathbf{U} = \begin{bmatrix} \rho \\ u \\ v \\ p \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0 \\ -\frac{\rho}{\bar{\rho}}\left(\bar{u}\frac{\partial \bar{u}}{\partial x} + \bar{v}\frac{\partial \bar{u}}{\partial y}\right) \\ -\frac{\rho}{\bar{\rho}}\left(\bar{u}\frac{\partial \bar{v}}{\partial x} + \bar{v}\frac{\partial \bar{v}}{\partial y}\right) \\ 0 \end{bmatrix},$$

$$\mathbf{A_x} = \begin{bmatrix} \bar{u} & \bar{\rho} & 0 & 0 \\ 0 & \bar{u} & 0 & \frac{1}{\bar{\rho}} \\ 0 & 0 & \bar{u} & 0 \\ 0 & \gamma\bar{p} & 0 & \bar{u} \end{bmatrix}, \quad \mathbf{A_y} = \begin{bmatrix} \bar{v} & 0 & \bar{\rho} & 0 \\ 0 & \bar{v} & 0 & 0 \\ 0 & 0 & \bar{v} & \frac{1}{\bar{\rho}} \\ 0 & 0 & \gamma\bar{p} & \bar{v} \end{bmatrix},$$

$$\mathbf{B_x} = \frac{\partial}{\partial \mathbf{x}}\begin{bmatrix} \bar{u} & \bar{\rho} & 0 & 0 \\ 0 & \bar{u} & 0 & 0 \\ 0 & \bar{v} & 0 & 0 \\ 0 & \bar{p} & 0 & \gamma\bar{u} \end{bmatrix}, \quad \mathbf{B_y} = \frac{\partial}{\partial \mathbf{y}}\begin{bmatrix} \bar{v} & 0 & \bar{\rho} & 0 \\ 0 & 0 & \bar{u} & 0 \\ 0 & 0 & \bar{v} & 0 \\ 0 & 0 & \bar{p} & \gamma\bar{v} \end{bmatrix},$$

and $\bar{u}, \bar{v}$ and u, v denote the velocity components of mean flow and fluctuations, respectively. $\bar{\rho}$ and ρ represent the mean flow and fluctuation densities, respectively. γ is the specific heat ratio. In this paper, all the variables are nondimensionalized by the following scales: length L, velocity c_∞, time L/c_∞, density ρ_∞, pressure $\rho_\infty c_\infty^2$, and impedance $\rho_\infty c_\infty$.

C. Discretization schemes

For solving the 2D LEE of Eq. (29), a computational aeroacoustics (CAA) approach is adopted and extended based on our previous works on the azimuthal sound mode propagation in hard-walled axisymmetric flow ducts[21] and the generation of supersonic jet screech tones.[22] The dispersion-relation-preserving (DRP) scheme[23] is employed for spatial discretization. The optimized coefficients of the centralized stencil are obtained for wave numbers up to the effective wave number $\eta = 1.1$. The required spatial resolution for the scheme is given by the optimized range to be approximately 5.7 points per wavelength (PPW). However, 10 PPW is usually required for the application to practical problems with high demands of accuracy. The alternating 5/6 step low-dissipation and low-dispersion Runge-Kutta (LDDRK) scheme of Hu *et al.*[24] in the 2*N* storage form[25] is used for time integration. Since the centralized DRP scheme has no inherent dissipation, a set of explicit standard filters with a stencil of up to 11 points is used in the numerical computations to eliminate the spurious waves and grid odd-even oscillations.

D. Boundary conditions

The implementation of the extension of the time-domain impedance boundary condition for subsonic flow on the surface as the main contribution of this paper were discussed in Sec. II D. Only the special case of hard wall boundary conditions is discussed here. Computational boundary conditions are used for the sound source with inflow and the anechoic outflow of the GIT, which are also briefly described below.

1. Hard wall boundary condition

The hard wall regions of the duct are treated as a special case of the impedance wall with $Z'(i\omega) = (\infty, i\infty)$. The single-frequency TDIBC Eqs. (5) and (6) or the broadband TDIBC Eqs. (11)–(13) can then be used as the hard wall boundary conditions. If the three-parameter model is applied for the impedance of the hard wall as well as the lined parts, the hard wall is obtained setting $1/X_+ = 0$. This will return the impedance boundary condition given by Eq. (11) back to the hard wall condition:

$$\frac{\partial v_n}{\partial t} = 0. \tag{30}$$

2. Inflow boundary condition

In the inflow region, a nonreflecting boundary condition permits the upstream propagating acoustic waves to leave the domain without reflection and generates an incoming plane waves simultaneously. Assuming a uniform mean flow in the x direction, the upstream $[\rho_{us}, u_{us}, v_{us}, p_{us}]^T$ and downstream $[\rho_{ds}, u_{ds}, v_{ds}, p_{ds}]^T$ propagating acoustic waves can be expressed in the following form,[26]

$$\begin{bmatrix} \rho_{us} \\ u_{us} \\ v_{us} \\ p_{us} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} F\left(\frac{x}{1-M} + t \right), \tag{31}$$

$$\begin{bmatrix} \rho_{ds} \\ u_{ds} \\ v_{ds} \\ p_{ds} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} G\left(\frac{x}{1+M} - t \right). \tag{32}$$

By differentiation and eliminating the unknown functions F and G, the acoustic disturbances would satisfy the following radiation boundary equations,

$$\left[\frac{\partial}{\partial t} - (1-M)\frac{\partial}{\partial x} \right] \begin{bmatrix} \rho \\ u \\ v \\ p \end{bmatrix} = -2\frac{\partial}{\partial x} \begin{bmatrix} \rho_{ds} \\ u_{ds} \\ v_{ds} \\ p_{ds} \end{bmatrix}, \tag{33}$$

where the downstream propagating wave at the inflow region $[\rho_{ds}, u_{ds}, v_{ds}, p_{ds}]^T$ will be excited as the sound source.

3. Perfectly matched layer outflow boundary condition

At the exit region of the duct, reflection waves may arise due to their termination not being fully anechoic in real situations. The complex impedance measured by Jones *et al.*[19] for the termination at the outflow may be used as a boundary condition. However, they represent only one value in a one-dimensional model and to reproduce a measured standing wave pattern the impedance of the inlet was also required. Therefore a nonreflecting boundary condition is prescribed at the outflow region.

This means there is no reflection of acoustic waves from the outflow boundary. At the same time hydrodynamic (vorticity and entropy) modes are allowed to leave the computational domain without sound generation. In order to allow acoustic, entropy and vorticity mode waves to leave the computational domain, a complex stretching of the spatial variable towards the outflow is applied. This stretching factor controls the imaginary part of the spatial variable and produces decreasing amplitudes without any change to the dispersion relation. In a limited zone the stretching factor for the axial direction of the flow duct (σ_x) is increased exponentially with decreasing distance to the outflow. All waves fade away before they reach the boundary of the computational domain. For the boundary point a zero-perturbation boundary condition is applied.

TABLE I. Impedance values for the three parameter model.

	$M_{avg}=0.255$			$M_{avg}=0.335$		
f (Hz)	R_0	$1/X_1$	X_{-1}	R_0	$1/X_1$	X_{-1}
500	0.57	20.8	−14.5	0.61	31.9	−8.1
1000	0.27	12.9	−24.6	0.675	13.1	−23.5
1500	1.26	12.9	−24.6	1.18	13.1	−23.5
2000	5.32[a]	123[a,b]	−34.4[a,b]	4.4	36.1[b]	−98.5[b]
2500	1.02	16.3	−198	0.93	17.8	−186
3000	0.73	16.3	−198	0.73	17.8	−186

[a]Original value from Ref. 19, but data table communicated by Jones to the authors shows the given value.
[b]Negative slope; no fit possible for $X_1 > 0$ and $X_{-1} \leqslant 0$. 2000 and 3000 Hz are used for the nonlinear curve fit to get X_1 and X_{-1}.

As the dispersion relation is perfectly matched between the boundary zone and the inner domain, this technique is called perfectly matched layer (PML) boundary condition[20] and can be expressed as

$$\frac{\partial \mathbf{U}}{\partial t} = \underbrace{-\mathbf{A}_x \cdot \frac{\partial \mathbf{U}}{\partial x} - \mathbf{A}_y \cdot \frac{\partial \mathbf{U}}{\partial y} - [\mathbf{B}_x + \mathbf{B}_y] \cdot \mathbf{U}}_{\text{Euler-equation}} - \underbrace{\sigma_x \mathbf{U}}_{\text{sponge layer}}$$
$$\underbrace{-\sigma_x \mathbf{A}_y \cdot \frac{\partial \mathbf{U}_1}{\partial y}}_{\text{perfect match}} - \underbrace{\frac{M_x}{1-M_x^2}\sigma_x \mathbf{A}_x \cdot \mathbf{U}}_{\text{Prandtl-Glauert-transform}}. \tag{34}$$

In this equation $\mathbf{U_1}$ denotes the auxiliary vector $\partial \mathbf{U_1}/\partial t = \mathbf{U}$, which makes it possible to allow a wave propagation along the nonreflective boundary, whereas the friction in the sponge layer removes any wave traveling into the boundary. The last term was obtained by Hu[20] in order to ensure the PML remains well-posed under homogeneous mean flow conditions.

IV. NUMERICAL RESULTS FOR THE NASA GIT

The extended TDIBCs are validated by the latest experimental results of Jones *et al.*[19] measured at the NASA GIT. A single-frequency 130-dB sound source excites plane waves at frequencies in the range of 0.5–3.0 kHz with 0.5-kHz increments, which fits the experiments published by Jones *et al.*[19]

The educed flow impedance of the ceramic tubular liner used by Jones *et al.*[19] changes under flow conditions. The liner is not a so-called linear liner, at least in the 2D model assumed here. Therefore the impedance educed with the particular mean flow of two selected mean flow Mach numbers is used to validate the new TDIBCs. The impedances used for the calculations are shown in Table I for the selected cases.[19] The corresponding dimensionless angular frequency ω is defined with the length unit $L=1$ m as $\omega = 2\pi L f_e / c_\infty$.

Then the incoming sound waves at the inflow boundary condition Eq. (33) can be written as

$$\begin{bmatrix} \rho_{ds} \\ u_{ds} \\ v_{ds} \\ p_{ds} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix} A \sin\left[\omega\left(\frac{x}{1+M} - t\right) + \phi\right], \tag{35}$$

where A denotes the amplitude of the incoming acoustic waves.

The computational domain is $x \in [-0.197, 1.000]$ and $y \in [0, 0.0508]$ with unit length $L=1$ m. Since the estimated shortest wavelength is approximately 0.1147 m in the x direction given by the highest frequency without mean flow ($f_e=3.0$ kHz at $M=0.0$) and the DRP scheme has an accuracy limit of about 10 PPW, the grid size in the x direction could be chosen at about 11.3 mm. This would mean only five points in the height direction of the duct are to be used ending up with a quasi-one-dimensional computation.

To reduce possible errors due to an irregular aspect ratio in the grid, a uniform Cartesian mesh is used in all computations with $\Delta x \approx \Delta y \approx 4$ mm. The CFL number is taken to be 0.5 for all cases. The computational domain is split into three multi-blocks for the application of hard wall and impedance boundary condition in the inlet, outlet, and duct sections, respectively. The total number of grid points including two exchange blocks of 6-points overlap is 322×15. This yields a grid resolution higher than 28 PPW for all cases. The overall computation time for 20 000 iterations with a time step size of $\Delta t \approx 4 \times 10^{-6}$ s (the results do not change essentially after 10 000 iterations) is about 10.5 min on a Pentium 4 with 3.06 GHz.

A. Effects of the mean flow

1. Influence of the mean flow velocity on the liner surface

First of all, numerical tests are carried out to find a good approximation of the mean flow velocity on the liner surface. The experimental data of Jones *et al.*[19] also includes mean flow profiles for the investigated Mach numbers at some positions.

As the model of Myers[18] considers an infinitely thin shear layer between the oscillating surface of the liner and the grazing flow, we need an velocity on the flow side of this shear layer. However, the experimental data does not provide this information, as the pitot probe method used by Jones *et al.*[19] does not allow us to measure velocities in the extreme vicinity to the wall.

The missing data can be reconstructed from the flow profile using the same assumption as Özyörük and Long[10] and Ju and Fung.[11] A fully developed laminar channel flow is assumed to evaluate the refraction effects as

$$\mathrm{M}(y) = (\mathrm{M}_{\max} - \mathrm{M}_e)(y/h)(1 - y/h) + \mathrm{M}_e \tag{36}$$

where h denotes the height of the duct and M_e is introduced for the missing velocity on the surface of the ceramic tubular liner, in order to study the influence of an assumed mean flow velocity on the decay observed in the simulation of the liner. This is done in the first step for the frequency of $f=2500$ Hz.

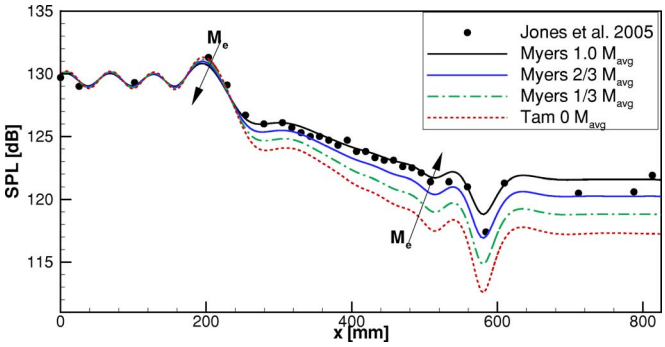


FIG. 2. (Color online) Parametric study for the effective Mach number on the liner surface for $f=2.5$ kHz in a constant mean flow with average Mach number $\mathrm{M}_{avg}=0.335$. The arrows denote the change with increasing M_e.

The result of this test is shown in Fig. 2. The Mach number is varied in the range of 0 to M_{avg} in the increment of $\frac{1}{3}\mathrm{M}_{avg}$. The mean flow profile for this test is a plug flow at the average Mach number $\mathrm{M}_{avg}=0.335$, only the convection correcture of Eq. (19) is varied. The results of the numerical experiment in Fig. 2 show a strong influence of the Mach number used for the convection correction on the decay rate.

As the waves propagate downstream with the mean flow, the effective impedances decrease with the assumed effective Mach number, which finally leads to the increase of sound pressure levels. The convective effect of the mean flow remains constant for the constant plug flow, only the Myers boundary condition is changed from no shear layer ($\mathrm{M}_e = \mathrm{M}_{avg}$) to a full shear layer assumption ($\mathrm{M}_e=0$).[3] The results suggest not to use the full average flow velocity, as the decay rate is too small. A good compromise is found with about $\frac{2}{3}$ of the peak Mach number, which is $0.796\mathrm{M}_{avg}$.

2. Influence of the mean flow profile

The measurement of the mean flow provided by Jones *et al.*[19] allows a more precise input of the real flow profile. The average measured flow profile is parabolic as it is typical for a laminar duct flow (see Fig. 3). As stated before, the information on the flow velocity on the perforated surface is missing in these data. Averaged over the liner there will be a

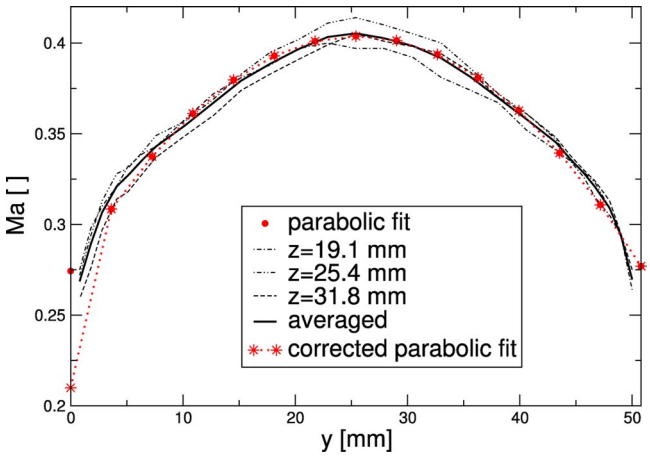


FIG. 3. (Color online) Parabolic fit of the averaged measured mean flow profiles at the average Mach number $\mathrm{M}_{avg}=0.335$. The corrected parabolic-fitted 15 data points are used in the following validation (measured profile).

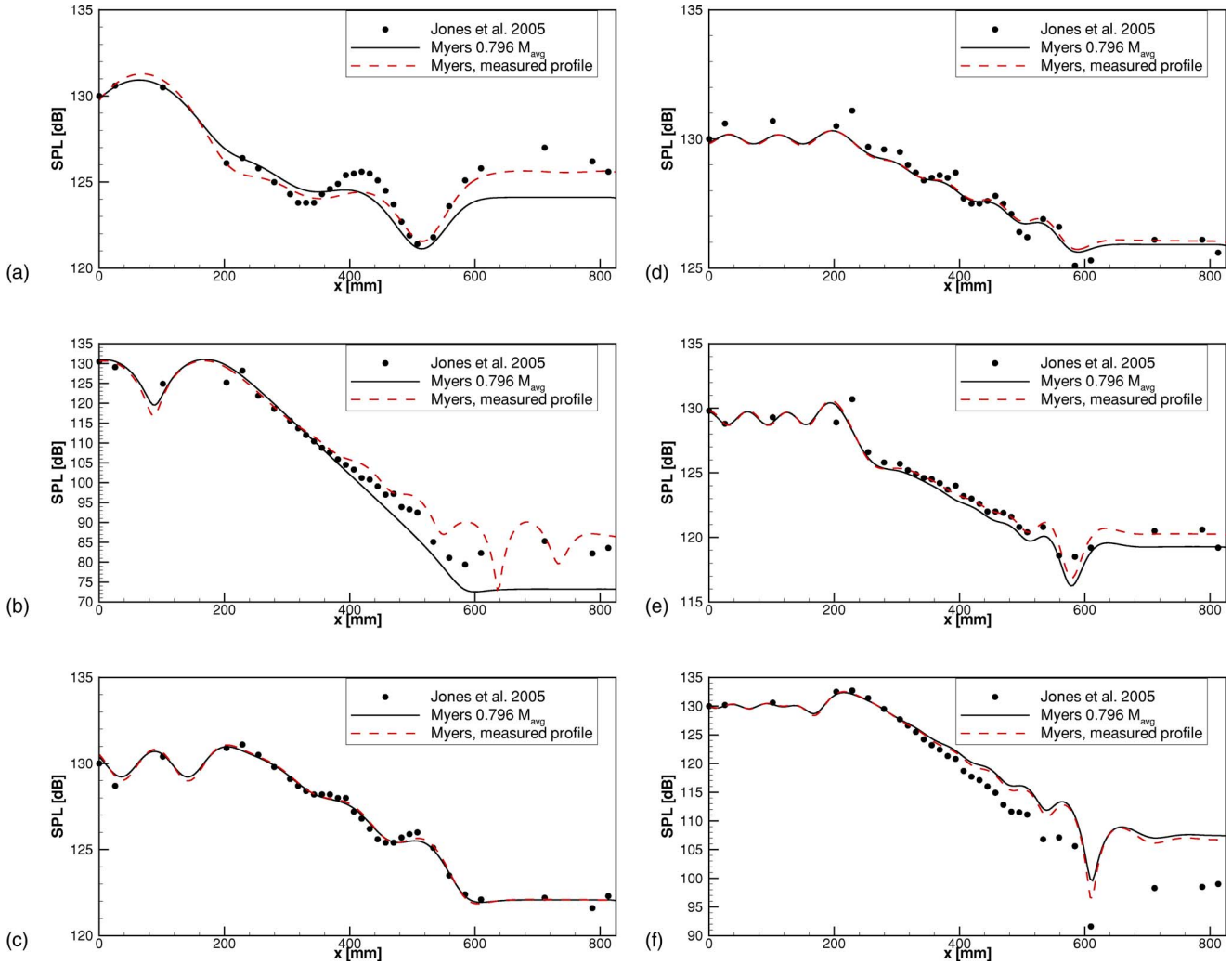


FIG. 4. (Color online) Sound pressure level along the opposite wall with flow (M_{avg}=0.255). The experimental data (bullets) are compared to the CAA results based on a plug flow (M_e=0.796M_{avg}; solid line) and the measured profile of Jones *et al.*[19] (M_e=0.6M_{avg}, dashed line). (a) f=500 Hz, (b) f=1000 Hz, (c) f=1500 Hz, (d) f=2000 Hz, (e) f=2500 Hz, and (f) f=3000 Hz.

velocity, as the orifices of the perforated liner allow grazing flow, while a no-slip condition should apply on the clean solid surface.

The velocity on the liner has to be based on some assumption, as it is not available from the measured data. Hot wire measurements could provide this data, however they must be averaged over orifices and solid clean parts depending on the perforated ratio of the liner. In the case shown in Fig. 3 the value at the surface of the liner has been fixed to M_e=0.21, which is around 60% of the average flow velocity. This value is obtained by a linear extrapolation of the velocity towards the liner starting from the last two measured values in the profile. For the effective impedance boundary condition this value acts as the effective velocity on the fluid side of the infinitely thin shear layer. Different from the plug flow investigated before, the convection is also described by the same value at the surface.

B. Validation with the NASA GIT experimental data

1. General comparison

Numerical results are presented against the experimental data as shown in Figs. 4 and 5, in which the solid and dashed lines represent the assumed plug flow with an effective Mach number of 79.6% of the average flow Mach number and the measured parabolic flow profile with an effective Mach number of 60% of the average flow Mach number, respectively. The first value assumed for the plug flow is obtained taking $\frac{2}{3}$ of the maximum Mach number of the laminar flow profile which is the extrapolated [see Eq. (36)] flow velocity arising at the 57% perforated surface of the liner. The second value is obtained by a linear extrapolation of the measured velocity values close to the liner.

The numerical results agree very well with the experimental data. Especially the measured flow profile leads to a very small difference of the SPL predicted and measured in the outlet duct except for the two cases at 3.0 kHz. In this case the first higher duct mode is very close to cut-on in the hard-walled outlet duct as well as in the lined section, which could possibly violate the assumption of the effective impedance calculation or even the eduction procedure by Jones *et al.*[19]

The numerical results obtained by the TDIBC using the plug flow assumption agree very well for the test frequencies around 2.0 kHz. However, for the low frequencies 500 Hz,

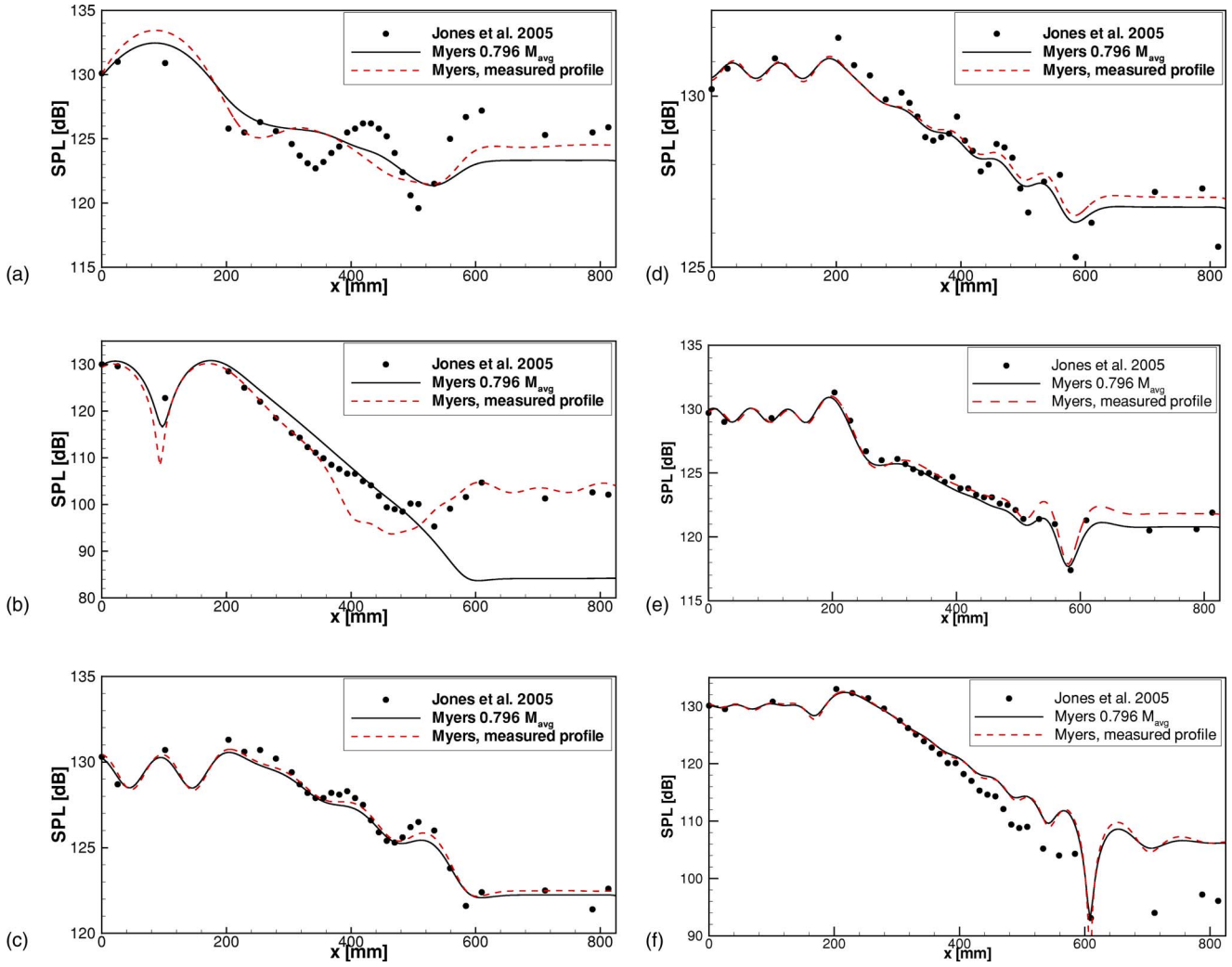


FIG. 5. (Color online) Sound pressure level along the opposite wall with flow (M_{avg}=0.335); The experimental data (bullets) is compared with the CAA results based on a plug flow (M_e=0.796M_{avg}; solid line) and the measured profile of Jones *et al.*[19] (M_e=0.6M_{avg}, dashed line). (a) f=500 Hz, (b) f=1000 Hz, (c) f=1500 Hz, (d) f=2000 Hz, (e) f=2500 Hz, and (f) f=3000 Hz.

1.0 kHz, and again the 3.0-kHz cases, the sound pressure levels in the outlet duct are mispredicted. This different behavior compared to the realistic flow assumption is discussed in detail in the next section.

2. Detailed investigation of the flow instability

The numerical results based on the plug flow assumption consistently underpredict the sound pressure levels for the frequencies f_e=0.5 and 1.0 kHz for the two test Mach numbers. The measured flow profile leads to a remarkably better prediction for these frequencies. The look at the instantaneous pressure contours in Fig. 6 can display the difference. A spatially growing hydrodynamic instability is excited at the trailing edge of the hard-walled section and, due to flow profile in the lined section, the amplitude increases linearly in space, but does not increase in time.

Hydrodynamic instabilities are inherent to the linearized Euler equations. In this case it is excited from the shear layer of the mean flow at the wall supported by the resonant frequency of the liner. The liner allows a vorticity wave to pass in the vicinity of the liner surface through the sheared flow, which allows a self-excited oscillation transferring energy from the shear layer to the growing hydrodynamic surface wave. It is also observed with the standard TDIBC of Tam and Auriault under sheared mean flow conditions and depends only on the flow profile. There is no instability observed with a plug flow. Once the instability wave hits the trailing edge of the wall downstream of the liner, it leads to increased near-field pressures on the upper wall as shown in Fig. 6. In real applications the growth of the instabilities is

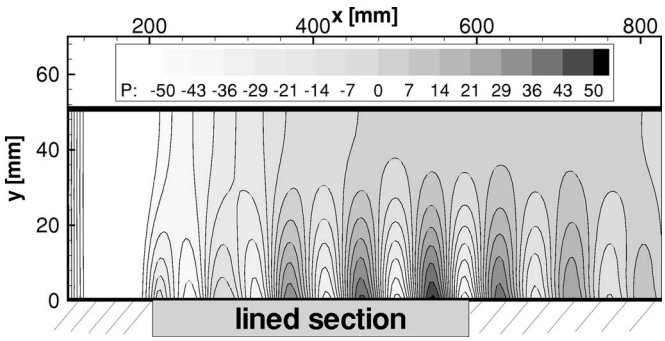


FIG. 6. Instantaneous pressure contours of the instability wave excited by the real flow profile at f_e=1000 kHz.

usually limited by nonlinear and viscous effects. If an instability is physically possible, it should also be observed in the experimental data. And, in fact, the decay along the opposite wall of the liner stops significantly earlier at 1.0 kHz than for all other frequencies. This observation could be connected to the acoustic field produced by the instability interacting with the trailing edge of the liner. As the impedance is modeled the impedance wall is similar to a compliant wall, which may attenuate or amplify boundary layer instabilities. The flow of the NASA GIT is laminar, therefore we assume the hydrodynamic instability is a Tollmien-Schlichting instability. For the interpolated flow profile shown in Fig. 3, the frequency of the Tollmien-Schlichting instability can be estimated. With the following connection of the perpendicular wave length λ in the shear layer and the momentum loss gage δ_1 obtained from $u_{\text{eff}}=0.335c$, $u_{\max}=0.4c$ and the channel height $h=0.051\ m$,

$$\lambda = 21\delta_1 = \frac{21}{2}\left(h - \sqrt{h^2\frac{u_{\text{eff}}}{u_{\max}}}\right) = 45.437\ \text{mm},$$

the typical frequency for this kind of flow instability is given by

$$f = \frac{u_c}{\lambda} = \frac{0.4u_{\text{eff}}}{\lambda} = 1002.71\ \text{Hz}.$$

As the Reynolds number of the average flow obtained with the boundary layer thickness is about 20 000, the flow can be unstable in general. This means that the flow instability observed in the computation at 1000 Hz is physically explainable as Tollmien-Schlichting instability that is amplified due to the resonance of the ceramic tubular liner described by the wall impedance. However, the missing exact flow profile at the wall leads to a overprediction of the amplitudes by the proposed effective impedance model.

With the assumption of parabolic mean flow, the numerical results improved significantly for most of these test cases. This shows that the refraction effects due to sheared mean flows become stronger at higher Mach numbers, which must be considered appropriately.

V. CONCLUSION

The single-frequency and broadband time-domain impedance boundary condition by Tam and Auriault have been shown to be suitable to include the mean flow convection effects without the requirement of a zero mean flow assumption on the impedance wall. The extended boundary conditions have been proved to be well-posed via a stability analysis. Furthermore, the traditional Myers boundary condition is connected to the extended time-domain impedance boundary conditions, which provides the effective acoustic impedance under flow conditions. This coupling releases the effective impedance from the requirement to be measured under given special acoustic conditions. The latest measurement data from the NASA greasing incidence tube is used for numerical validations. It is demonstrated that the extended boundary conditions can be used for slip and parabolic mean flow profiles, and the numerical results agree very well with the measurement data. The numerical validations demonstrate that the extended TDIBCs are suitable for different mean flow profiles. Providing a correct impedance and an appropriate mean flow profile, the sound propagation in lined ducts can be predicted accurately. It should be pointed out that it is not so easy to give the correct impedance in the presence of mean flow for nonplane incident waves. More general impedance models as well as more realistic mean flow profiles are still highly desirable. An instability on the liner surface is observed using the measured flow profile at the excitation frequency of 1000 Hz. This instability is connected to the presence of a Tollmien-Schlichting instability for the flow profile in nearly exact concurrence with the liner resonance. In addition, further numerical validation is required for the presented time-domain impedance boundary conditions with complex geometries and for higher mode propagation.

ACKNOWLEDGMENTS

The work was supported by NSFC-10472010, ASFC-05C51030, the DFG project group FOR-486 and LiMiT. The authors also would like to thank reviewers for their constructive comments and suggestions.

NOMENCLATURE

$\boldsymbol{\nabla}$ = gradient in Cartesian coordinates
$\boldsymbol{\nabla}_n$ = normal derivative at the surface
a_i^{51} = backward stencil coefficients of the DRP scheme
$\bar{c}$ = speed of sound
f_e = exciting frequency of sound waves
F = auxiliary function for stability analysis
f = auxiliary function for stability analysis
g = auxiliary function for stability analysis
k = wave number
k_x = wave number in x direction
$\mathbf{n}$ = normal vector pointing into the surface
$\mathbf{M}$ = dimensionless velocity (Mach number) vector
$\mathbf{M_e}$ = effective Mach number vector
M_e = effective Mach number
p = pressure fluctuation
q = auxiliary integral variable
Q = auxiliary derivative variable
R = acoustic resistance
R' = effective acoustic resistance
t = time
$\bar{u},\bar{v}$ = components of mean flow velocity
u,v = components of velocity fluctuation
v_n = normal fluctuation velocity
$\mathbf{v}$ = velocity vector
X = acoustic reactance
X' = effective acoustic reactance
Z = impedance
Z' = effective plane-wave impedance
γ = specific heat ratio
$\bar{\rho}$ = mean flow density
ρ = density fluctuation

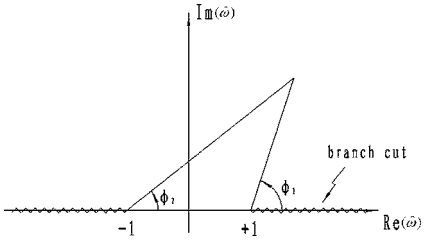


FIG. 7. Branch cut configuration for $(\hat{\omega}^2-1)^{1/2}=\frac{1}{2}(\phi_1+\phi_2)$.

Ω, ω = angular frequency

APPENDIX A: THE WELL-POSEDNESS OF THE SINGLE-FREQUENCY TDIBC FOR SURFACES WITH SUBSONIC MEAN FLOWS

Assuming the impedance surface located in the $x-z$ plane and applying the Fourier-Laplace transform to Eqs. (3) and (4), it can be found that the solution satisfies the outgoing wave condition at $y\to\infty$:

$$\begin{bmatrix}\tilde{u}\\ \tilde{v}\\ \tilde{w}\\ \tilde{p}\end{bmatrix}=A\begin{bmatrix}\alpha/(k\hat{\omega})\\ (\hat{\omega}^2-1)^{1/2}/\hat{\omega}\\ \beta/(k\hat{\omega})\\ 1\end{bmatrix}\exp\left[k(\hat{\omega}^2-1)^{1/2}y\right], \tag{A1}$$

where $\hat{\omega}=(\Omega-M_e\alpha)/k$, $k=(\alpha^2+\beta^2)^{1/2}$. The branch cuts of the function $(\hat{\omega}^2-1)^{1/2}$ are taken to be $0\leqslant\arg(\hat{\omega}^2-1)^{1/2}\leqslant\pi$ as shown in Fig. 7.

The Fourier-Laplace transforms of Eqs. (1) and (2) are

$$\tilde{p}=\left[-R'+\frac{iX'\omega}{\hat{\omega}k+\alpha \mathrm{M}_e}\right]\tilde{v},\quad X'<0, \tag{A2}$$

$$\tilde{p}=\left[-R'+\frac{iX'}{\omega}(\hat{\omega}k+\alpha \mathrm{M}_e)\right]\tilde{v},\quad X'\geqslant 0, \tag{A3}$$

respectively. Substitution of Eq. (A1) into Eqs. (A2) and (A3) leads to the following dispersion relations:

$$(\hat{\omega}k+\alpha \mathrm{M}_e)\left(R'+\frac{\hat{\omega}}{\sqrt{\hat{\omega}^2-1}}\right)=iX'\omega,\quad X'<0, \tag{A4}$$

$$\frac{\hat{\omega}}{(\hat{\omega}^2-1)^{1/2}}-\frac{i(\hat{\omega}k+\alpha \mathrm{M}_e)X'}{\omega}=-R',\quad X'\geqslant 0. \tag{A5}$$

The extended single-frequency TDIBC is well posed if only Eqs. (A4) and (A5) have no solutions in the upper half $\hat{\omega}$ plane. Tam and Auriault[3] have proved the well-posedness for cases without mean flow. For cases with mean flows, let the left-hand side of Eqs. (A4) and (A5) be expressed by $f(\hat{\omega})$ and $g(\hat{\omega})$, respectively:

$$f(\hat{\omega})=(\hat{\omega}k+\alpha \mathrm{M}_e)\left(R'+\frac{\hat{\omega}}{\sqrt{\hat{\omega}^2-1}}\right), \tag{A6}$$

$$g(\hat{\omega})=\frac{\hat{\omega}}{(\hat{\omega}^2-1)^{1/2}}-\frac{i(\hat{\omega}k+\alpha \mathrm{M}_e)X'}{\omega}. \tag{A7}$$

Assuming a subsonic mean flow $(0<\mathrm{M}_e<1)$, the stability problem can be discussed for $\alpha>0$, $\alpha=0$, and $\alpha<0$, respectively. For $\alpha=0$, the single-frequency TDIBCs are also well-posed since Eqs. (A4) and (A5) have the same forms as for cases of zero mean flow. Figures 8 and 9 show the maps of the upper half $\hat{\omega}$ plane in the f and g planes for cases of $\alpha>0$. The shaded regions are the image points of the upper half $\hat{\omega}$ plane. For the case of $X'<0$, there is no value of $\hat{\omega}$ in the upper half $\hat{\omega}$ plane that can satisfy dispersion relation of Eq. (A4) since the right-hand side of Eq. (A4) is imaginary and negative. For the case of $X'\geqslant 0$, the right-hand side of Eq. (A5) is real and negative, and no value of $\hat{\omega}$ in the upper

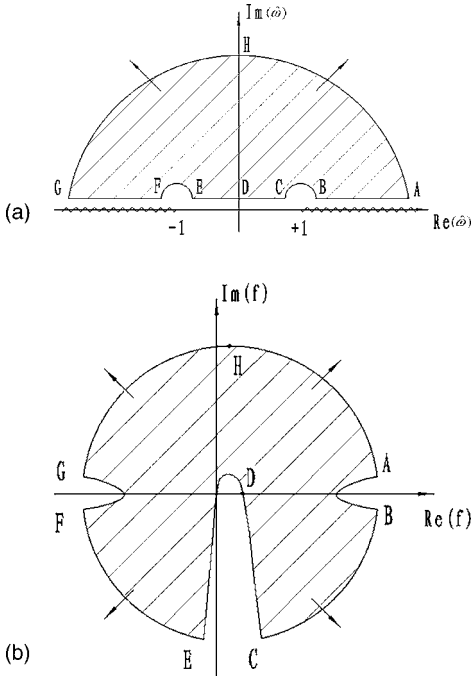


FIG. 8. Map of the upper half $\hat{\omega}$ in the f plane. (a) $\hat{\omega}$ plane. (b) f plane.

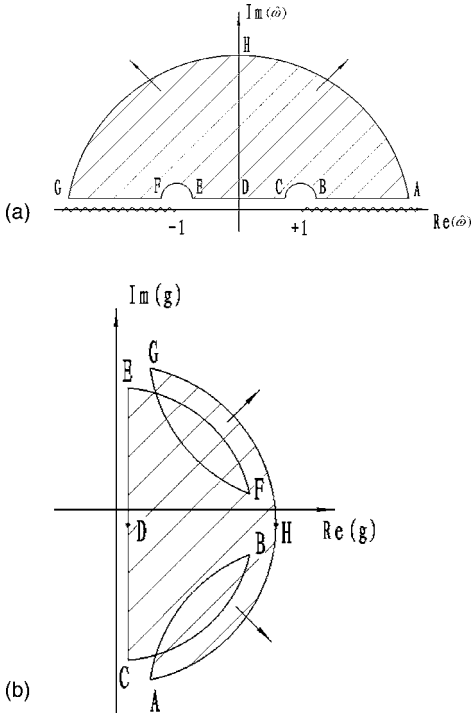


FIG. 9. Map of the upper half $\hat{\omega}$ in the g plane. (a) $\hat{\omega}$ plane. (b) g plane.

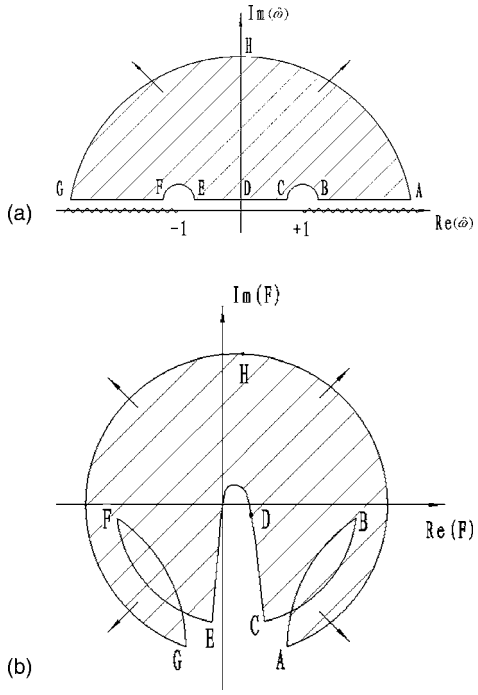


FIG. 10. Map of the upper half $\hat{\omega}$ in the F plane. (a) $\hat{\omega}$ plane. (b) F plane.

half $\hat{\omega}$ plane can satisfy the dispersion relation of Eq. (A5). Similarly, it can also be proved that there is no stability problem for the cases of $\alpha<0$. Detailed discussion is neglected here for conciseness. And it is easy to find that the above discussion is still valid for $-1<M_e<0$ through the sign change of α. Therefore, the extended single-frequency TDIBC of Eqs. (1) and (2) are well-posed.

APPENDIX B: THE WELL-POSEDNESS OF THE BROADBAND TDIBC FOR SURFACES WITH SUBSONIC MEAN FLOWS

The well-posedness of the extended TDIBC Eq. (9) is analyzed by a Fourier-Laplace transform

$$-\omega\tilde{p}=[\omega R_0'-i(X_{-1}'+\omega^2X_1')]\tilde{v}_n. \tag{B1}$$

Substitution of Eq. (A1) into Eq. (B1) yields the following dispersion relation,

$$(\hat{\omega}k+\alpha\mathrm{M}_e)\left[R_0'-iX_1'(\hat{\omega}k+\alpha\mathrm{M}_e)+\frac{\hat{\omega}}{\sqrt{\hat{\omega}^2-1}}\right]=iX_{-1}'. \tag{B2}$$

Figure 10 shows the map of the upper half $\hat{\omega}$ plane in the F plane for $\alpha>0$. Since the right-hand side of Eq. (B2) is imaginary and negative, no value of $\hat{\omega}$ in the upper half plane can satisfy the dispersion relation Eq. (B2). A similar proof can be given for the opposite case of $\alpha<0$. Thus the extended broadband TDIBC for subsonic mean flows leads to well-posed problems.

[1] J. Hamilton and R. Astley, "Acoustic Propagation on Irrotational Mean Flow Using Time-domain Finite and Infinite Elements," AIAA Paper No. 2003–3208 (2003).
[2] X. D. Li, N. Schönwald, J. P. Yan, and F. Thiele, "A Numerical Study on the Acoustic Radiation from a Scarfed Intake," AIAA Paper No. 2003-3245 (2003).
[3] C. K. W. Tam and L. Auriault, "Time-domain Impedance Boundary Conditions for Computational Aeroacoustics," AIAA J. **34**, 917–923 (1996).
[4] Y. Özyörük and L. N. Long, "A Time-Domain Implementation on Surface Acoustic Impedance Condition With and Without Flow," AIAA Paper No. 1996-1663 (1996).
[5] K.-Y. Fung, H. B. Ju, and B. Tallapragada, "Impedance and Its Time-Domain Extensions," AIAA J. **38**, 30–38 (2000).
[6] K.-Y. Fung and H. B. Ju, "Broadband Time-Domain Impedance Models," AIAA J. **39**, 1449–1454 (2001).
[7] K.-Y. Fung and H. Ju, "Time Domain Impedance Boundary Conditions," Int. J. Comput. Fluid Dyn. **18**, 503–511 (2004).
[8] T. L. Parrott, W. R. Watson, and M. G. Jones, "Experimental Validation of a Two-Dimensional Shear-Flow Model for Determining Acoustic Impedance," NASA TP-2679 (1987).
[9] Y. Özyörük, L. N. Long, and M. G. Jones, "Time-domain Numerical Simulation of a Flow-Impedance Tube," J. Comput. Phys. **146**, 29–57 (1998).
[10] Y. Özyörük and L. N. Long, "Time-domain Calculation of Sound Propagation in Lined Ducts with Sheared Flows," AIAA J. **38**, 768–773 (2001).
[11] H. Ju and K.-Y. Fung, "Time-Domain Impedance Boundary Conditions with Mean Flow Effects," AIAA J. **39**, 1683–1690 (2001).
[12] S. Zheng and M. Zhuang, "Verification and Validation of Time-Domain Impedance Boundary Condition in Lined Ducts," AIAA J. **43**, 306–313 (2005).
[13] S. Zheng and M. Zhuang, "Three-Dimensional Benchmark Problem for Broadband Time-Domain Impedance Boundary Conditions," AIAA J. **42**, 405–407 (2004).
[14] B. J. Tester, "The Propagation and Attenuation of Sound in Lined Ducts Containing Uniform or Plug Flow," J. Sound Vib. **28**, 151–203 (1973).
[15] U. Ingard, "Influence of Fluid Motion Past a Plane Boundary on Sound Reflection, Absorption, and Transmission," J. Acoust. Soc. Am. **31**, 1035–1036 (1959).
[16] M. Goldstein and E. Rice, "Effect of Shear on Duct Wall Impedance," J. Sound Vib. **30**(1), 79–84 (1973).
[17] A. H. Nayfeh, J. E. Kaiser, and D. P. Telionis, "Acoustics of Aircraft Engine-Duct Systems," AIAA J. **13**(2), 130–153 (1975).
[18] M. K. Myers, "On the Acoustic Boundary Condition in the Presence of Flow," J. Sound Vib. **71**, 429–434 (1980).
[19] M. G. Jones, W. R. Watson, and T. L. Parrott, "Benchmark Data for Evaluation of Aeroacoustic Propagation Codes with Grazing Flow," AIAA Paper No. 2005-2853 (2005).
[20] F. Q. Hu, "A Stable Perfectly Matched Layer For Linearized Euler Equations in Unsplit Physical Variables," J. Comput. Phys. **173**, 455–480 (2001).
[21] X. D. Li, C. Schemel, U. Michel, and F. Thiele, "Azimuthal Sound Mode Propagation in Axisymmetric Flow Ducts," AIAA J. **42**, 2019–2027 (2004).
[22] X. D. Li and J. H. Gao, "Numerical Simulation of the Generation Mechanism of Axisymmetric Supersonic Jet Screech Tones," Phys. Fluids **17**(8), 085105 (2005).
[23] C. K. W. Tam and C. Webb, "Dispersion-relation-preserving Finite Difference Schemes for Computational Aeroacoustics," J. Comput. Phys. **107**, 262–281 (1993).
[24] F. Q. Hu, M. Y. Hussaini, and J. Manthey, "Low-Dissipation and Low-Dispersion Runge-Kutta Schemes for Computational Acoustics," J. Comput. Phys. **124**, 177–191 (1996).
[25] D. Stanescu and W. Habashi, "2N-storage Low-Dissipation and Low-Dispersion Runge-Kutta Schemes for Computational Aeroacoustics," J. Comput. Phys. **143**, 674–681 (1998).
[26] C. K. W. Tam and A. Aganin, "Computation of Transonic Nozzle Sound Transmission and Rotor Problems by the Dispersion-Relation-Preserving Scheme," NASA CP 2000-209790 (2000).

Measurement and simulation of the channel intensity impulse response for a site in the East China Sea

Jee Woong Choi and Peter H. Dahl
Applied Physics Laboratory, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105-6698

(Received 20 September 2005; revised 16 February 2006; accepted 2 March 2006)

A model for the channel intensity impulse response $I^c(t)$ is presented that is generally applicable for source-receiver ranges less than ten water depths. The separate impulse response functions from each arrival, such as the single surface bounce or surface-to-bottom bounce, are modeled using bistatic scattering concepts and are incoherently summed for the total response function. The expression $I^c(t)$ is equivalent to a time-averaged response and embodies the boundary scattering and reflection physics corresponding to the center frequency at which computations are made. To compare with observations, $I^c(t)$ is convolved with representations of the 8- and 16-kHz continuous wave (CW) pulses, and an 8–16-kHz frequency-modulated (FM) pulse, that were used in the Asian Sea International Acoustics Experiment conducted in the East China Sea (depth 105 m). For the FM case the computation frequency is 12 kHz, the center frequency of the FM pulse. It is found that six primary arrivals dominate the response for ranges less than about 1 km. With modeling of $I^c(t)$ limited to these paths, the basic structure of $I^c(t)$ is set by bottom properties and acquisition geometry with some changes in intrapath time spreading that depend on sea surface conditions.
 [DOI: 10.1121/1.2189449]

PACS number(s): 43.30.Zk, 43.30.Gv, 43.30.Hw [DRD] Pages: 2677–2685

I. INTRODUCTION

Sound propagation in shallow water is greatly influenced by multiple interactions of the transmitted waveform with the sea surface and seabed boundaries producing time and frequency spread as well as energy loss. Here the channel impulse response is investigated for a shallow water waveguide in which the frequency range, water depth (D), and nominal source-receiver range (R) are 8–16 kHz, 105 m, and 500 m, respectively. In this regime, it makes sense to describe the channel in terms of identifiable eigenrays[1] corresponding to the direct path, sea surface path, bottom path, and surface-bottom path (and so on). In this paper a channel intensity impulse response function (IIRF) $I^c(t)$ is constructed using ray theory to estimate the arrival time and scattering and reflection models (discussed below) that characterize the boundary interaction.

To compare with observations, $I^c(t)$ is convolved with functions representing the continuous wave (CW) and frequency-modulated (FM) pulses that were used in the Asian Seas International Acoustics Experiment (ASIAEX) conducted in the East China Sea.[2] For the CW signals (3-ms duration, center frequencies of 8 and 16 kHz), this function is a simple boxcar function of the same duration. For the frequency-modulated (FM) signal (linear frequency-modulated between 8 and 16 kHz, duration 20 ms), this function is the square of the autocorrelation function of the FM pulse. Both CW and FM cases are compared with measurements; for the FM case, the received signals are matched filtered with the FM pulse replica and, after travel time alignment, the data are intensity averaged over 20 independent realizations. In the context of underwater acoustic communications, this represents (in an averaged sense) the front end of an optimal receiver.[3] Intersymbol interference associated with multipath propagation is an important and limiting feature in underwater acoustic communication,[4,5] motivating study of alternate techniques, e.g., involving time-reversal,[6,7] to overcome this effect. Also note that the IIRF, because of its strong relation to bottom properties, has been used in geoacoustic parameter inversion.[8]

This paper is organized as follows. A description of the field measurements is given in Sec. II. Construction of the model for the IIRF is presented in Sec. III. Results of convolving the model with representations of the CW and FM pulses and comparison with field measurements are presented in Sec. IV. Section V discusses the contribution from individual paths, and their incoherent sum that yields $I^c(t)$. A summary is given in Sec. VI.

II. FIELD MEASUREMENTS

The measurements were part of a set of nominally 0.5-km range propagation measurements designed to study sea surface scattering[9] and bottom reflection.[10] The measurements were made from the R/V *Melville* at a site 350 km east of Shanghai off the Chinese continental margin (29° 39′N, 126° 40′E) in waters 105 m deep (Fig. 1). The acoustic source system was deployed at depths of 25 and 50 m from the stern of the R/V *Melville*, which maintained position by dynamic positioning. (The positioning accuracy of the reference point, in this case *Melville*'s stern and location of source, was within 10 m.) Received signals were recorded with two collocated vertical line arrays, each of length 1 m, with one array at depth 26 m and the other array at depth 52 m, referenced to the array center (see Ref. 9 for discussion of array measurement). The location of the vertical line

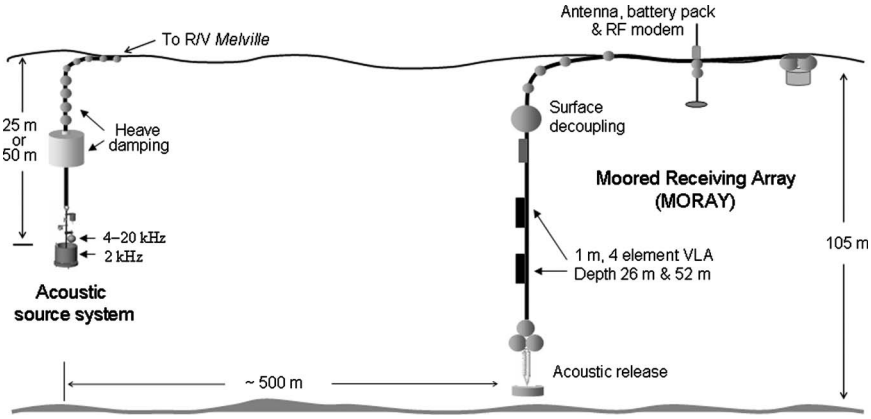


FIG. 1. Experimental layout for measuring the IIRF. The data presented in this paper originate from a single hydrophone on the receiving arrays at 26 and 52 m depth.

array was maintained by a stiff mooring. The three waveforms (two CW, one FM) were transmitted sequentially, separated by 200 ms, with a given waveform transmitted every 100 s and repeated 20 times to form an experimental set. The received signals were sampled at 50 kHz and intensity averaged over 20 pulses.

In terms of bottom properties, measurements of the fine scale sediment roughness[11] and bottom loss,[10] made simultaneously with those reported here, are applied. Specifically, the bottom roughness consists of small features with a root-mean-square (rms) height of about 5 mm and there is a surficial sediment layer of thickness $O(1)$ m within which sound speed increases from 1557 m/s at the water-sediment interface to 1625 m/s.

Figure 2 summarizes one measurement set for range 416 m taken at 13:54 UTC May 31 during which the wind speed was 7 m/s and rms waveheight was ~0.3 m. The right-hand side of Fig. 2(a) shows eigenrays corresponding to the first six arrivals: direct (D), sea surface (S), bottom (B), surface-bottom (S-B), bottom-surface (B-S), and surface-bottom-surface (S-B-S) paths in this order, and are computed using a ray-based propagation code based on the sound speed profile [Fig. 2(a), left]. The received intensity level for the CW pulse [Fig. 2(b)] and the amplitude of the matched filter response for the FM pulse [Fig. 2(c)] are referenced to those of the peak value for the direct path. The FM case is plotted in amplitude space for ready comparison with other results of this kind.[8,12] At this ~0.5-km range, there are essentially six paths that dominate the response, corresponding to the eigenrays in Fig. 2(a); all have no or only one interaction with the bottom (a range of approximately 1.7 km is required for bottom loss to be significantly reduced by precritical angle effects). The standard deviations for both data sets is ~5.6 dB and is consistent with the stan-

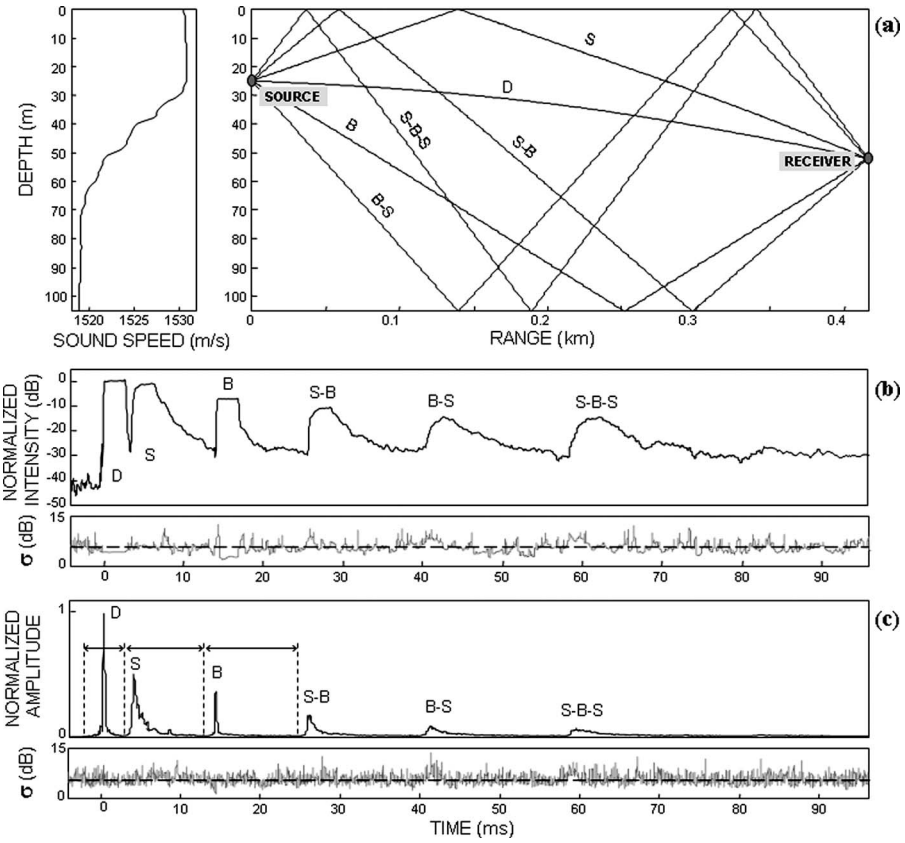


FIG. 2. First six eigenrays (right-hand side) for source depth 25 m, receiver depth 52 m, and range 416 m based on the average sound speed profile (left-hand side) corresponding to time of acoustic measurement. (b) Averaged intensity ($N=20$ realizations) and the standard deviation expressed in decibels for the 16-kHz CW signal. (c) Averaged amplitude ($N=20$ realizations) of the matched filter envelope response for the FM signal (averaging performed in intensity domain and square root taken) and its standard deviation expressed in decibels. The dashed line in the standard deviation plots indicates 5.6 dB. Measurements were taken on 31 May at 13:54 UTC during which the wind speed and rms waveheight were 7 m/s and ~0.3 m, respectively. Labels D, S, and B, etc. denote paths referred to in text as direct, surface, bottom, and their combinations.

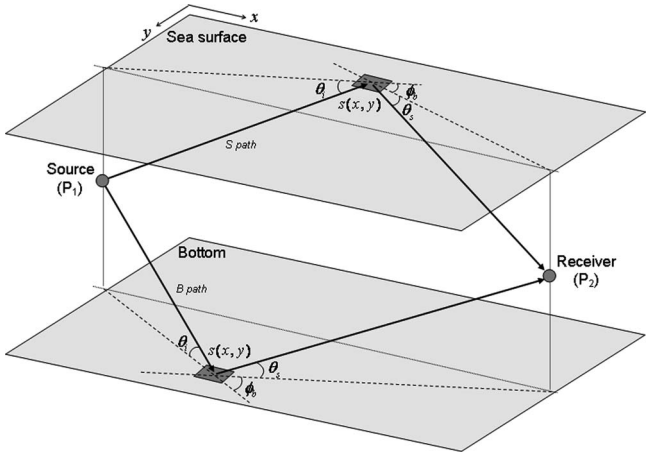


FIG. 3. The scattering geometry for sea surface (S) path and bottom (B) path. The angles θ_i, θ_s, and ϕ_b are the incident grazing angle, scattered angle, and bistatic angle, respectively.

dard deviation for an approximately instantaneous intensity measurement.[13,14] The D and B paths shown in the CW case are an exception; they are not subject to fluctuations of the kind imposed by the sea surface and have standard deviations of 4.1 and 2.5 dB, respectively. The corresponding portions for the FM case are not distinguishable owing to the impulselike shape of the matched filter response.

III. CONSTRUCTION OF CHANNEL INTENSITY IMPULSE RESPONSE WITH MULTIPATH ARRIVALS

The IIRF $I^c(t)$ is constructed using separate response functions, labeled generally as $I^p(t)$, with p designating each path eigenray. For example, the path interacting once with the sea surface is $I^s(t)$. This has been modeled by Dahl[15] utilizing the sea surface incoherent bistatic cross section for the case $\chi \gg 1$, where χ is the Rayleigh parameter; based on the grazing angle for the nominal specular path θ_g, then χ equals $2kH \sin \theta_g$, where k is acoustic wave number and H is rms waveheight. Recently Culver and Bradley[16] relate $I^s(t)$ based on this same modeling approach to signal bandwidth and frequency correlation. The path interacting with the bottom once is $I^b(t)$, which has been modeled by Williams and Jackson.[17] Here, the approaches largely similar to Refs. 15 and 17 are used to estimate $I^s(t)$ and $I^b(t)$ with some modifications discussed subsequently. The expression $I^c(t)$ gives the reflected, plus scattered and time-spread, intensity originating from (quoting from Ref. 17) a "short-duration or 'impulse' pulse." Upon convolving $I^c(t)$ with the squared envelope of the transmitted signal (CW case) or the squared autocorrelation function (FM case), a prediction for mean intensity level is obtained that is comparable to measurements.

The single boundary interaction case (Fig. 3) conveys the basic approach for obtaining an expression for the intensity reflected and scattered from a short-duration pulse transmitted at position P_1 and measured at the receiver position P_2. This interaction is described by the distribution of the bistatic cross section over the boundary surface as a function of position $s=s(x,y)$, with position s_{sp} corresponding to the specular point. (These positions are taken to be on the plane associated with the mean surface height.) Define $I^p(t)$ [applying to either the sea surface $I^s(t)$ or seabed $I^b(t)$ case] as follows:

$$I^p(t) = I_{\text{coh}}(t)\exp(-\chi^2) + I_{\text{scat}}(t). \tag{1}$$

The first term is the coherently reflected intensity where

$$I_{\text{coh}}(t) = \frac{I_0 BxBr|\Re(\theta_g)|^2}{(R_1+R_2)^2} 10^{-\alpha_w(R_1+R_2)/10}\delta(t-t_g), \tag{2}$$

with I_0 being the source intensity at range 1 m from the transducer (source level), Bx and Br are the intensity beam pattern weightings for the transmitter and receiver, respectively, $\Re(\theta_g)$ is the reflection coefficient, and α_w is the sea water attenuation coefficient in dB/m. The time dependence $\delta(t-t_g)$ depends on the path delay associated with the specular point t_g, which is set by the slant ranges $R_1=P_1s_{sp}$ and $R_2=P_2s_{sp}$, and the depth-averaged sound speed in the water column through which the ray propagates. The Rayleigh parameter is an important descriptor for the ASIAEX geometries, environmental conditions, and frequency range. For the sea surface, χ generally exceeds unity and the incoherent intensity dominates; the exception is the lowest frequency (8 kHz) and calmer conditions. For the seabed, χ is $\sim$0.1 and reflection dominates.

The second term is the result of incoherently scattered intensity and is computed as follows:[15] The travel time from the nth scattering patch located at s_n is computed from geometry and grouped within the interval defined by $t_g+k\Delta t$ and $t_g+(k+1)\Delta t$, $k=0,1,2,\ldots$, where $\Delta t=0.02$ ms. The scattered intensity for this path is I_{scat_n} and is given by

$$I_{\text{scat}_n} = I_0 \frac{Bx(s_n)Br(s_n)\sigma(s_n)10^{-\alpha_w[(P_1s_n)+(P_2s_n)]/10}dA}{(P_1s_n)^2(P_2s_n)^2}, \tag{3}$$

whereas the scattering patch area dA is 1 m^2, and σ is the incoherent bistatic scattering cross section per unit area, which is expressed as a function of angles θ_i, θ_s, and ϕ_b (Fig. 3). All scattered intensities whose travel times fall within the same kth interval associated with an equal travel time ellipse are summed, with result identified as I_k, and the scattering component of the intensity impulse response [i.e., in Eq. (1)] is given by

$$I_{\text{scat}}(t) = \sum_k I_k \delta(t-t_k). \tag{4}$$

A. Bistatic scattering cross section for the sea surface

Forward scattering from the sea surface is modeled using the small slope approximation for the incoherent bistatic cross section plus a bubble scattering component to give $\sigma_{sfc}=\alpha_b\sigma_{sr}+\sigma_b$, where σ_{sr} is the bistatic scattering cross section associated with the air-sea interface, α_b accounts for attenuation due to near-surface bubbles, and σ_b is the scattering cross section from near-surface bubbles.[18] Computing σ_{sr} is achieved by way of directional averaging of the sea surface autocorrelation function, since there was no significant improvement in model-data comparisons upon incorpo-

TABLE I. Input parameters for the bottom bistatic scattering model.

	Parameter							
	Density ratio	Sound speed ratio	Loss parameter	Spectral exponent	Spectral strength	Fluctuation ratio	Inhomogeneity exponent	Inhomogeneity strength
Symbol	ρ	ν	δ	γ	ω_2	μ	γ_3	ω_3
Value	2.0	1.025	0.00704	2.99	0.00379	−1	3	0.0001558

rating directional wave properties. The directionally averaged cross section requires evaluation of the integral

$$\sigma_{sr} \sim \int_0^\infty \rho J_0(2\kappa\rho)[e^{-\chi^2[1-C(\rho)]} - e^{-\chi^2}]d\rho, \tag{5}$$

where $C(\rho)$ is the isotropic autocorrelation function of sea surface waveheight variation, which is a function of the radial lag coordinate ρ, and κ is the magnitude of the horizontal component of the incident minus scattered wave vector.[15,18] The second term contributing to the integral can be ignored for $\chi \gg 1$ as a result of higher frequencies; inclusion here is needed for the mid frequencies (8 kHz) and moderate wind speeds that put $e^{-\chi^2}$ equal to ~ 0.1.

B. Bistatic scattering cross section for the seabed

The APL-UW bistatic bottom scattering model[17,19] is, upon incorporating a small modification, used to predict the bottom interaction of the acoustic energy. Note first that this model calculates the total cross section rather than the incoherent scattering cross section, so the first term in Eq. (1) is not used.[19] The modification entails inserting the reflection coefficient model from Ref. 10 into the bistatic scattering model that is based on a two-layered sediment for which the sound speed in the surficial sediment layer is allowed to vary. The sediment sound speed and density estimated from the geoacoustic inversion work from ASIAEX reported in Ref. 10 are also used. The one-dimensional roughness spectrum $W_1(K_x)$ for the bottom can be represented as follows,[19]

$$W_1(K_x) = \frac{w_1}{|K_x|^{\gamma_1}}, \tag{6}$$

where K_x is the spatial wave number of bottom relief features. For the ASIAEX site γ_1 and w_1 were estimated to be 1.99 and 0.0076 $\mathrm{cm}^{3-\gamma_1}$, respectively,[11] and these are converted to their isotropic two-dimensional counterparts. The input parameters for the APL-UW bistatic scattering model used in this work are given in Table I.

C. Multiple boundary interactions

For multiple boundary interactions two cases are addressed: (1) single bottom plus single surface interaction or the S-B and B-S paths for which $I^{sb}(t)$ and $I^{bs}(t)$ are constructed, and (2) single bottom plus double surface interaction or the S-B-S path for which $I^{sbs}(t)$ is constructed. In the first case the ansatz is taken that the impulse response for such a path can be described by a surface scattering process plus a bottom reflection process.[20] This approach (Fig. 4) invokes either an image source or receiver, depending on the path. The data show that time spreading associated with bottom scattering is minimal, which is consistent with $\chi \ll 1$ for the bottom, and it is assumed that any time spreading associated with the bottom will be not observable upon subsequent interaction with the sea surface which dominates the spreading. This point is quantified as follows using an integrated measure of time spread. The results shown in Fig. 2(c) represent a band-limited approximation to the time-averaged intensity impulse response; let us compute an integrated measure (defined in Ref. 15 and called L) of this response for its separate components, D, S and B, as delineated by the markers in Fig. 2(c). The result for the D, S, and B paths are 0.29, 1.16, and 0.34 ms, respectively. Consider the estimate for the D path as the baseline or minimum value given the FM pulse used in the measurements and system bandwidth, and effective time spread is in excess of this value. Thus, the effective time spread for the B path is minimal and that for the S path exceeds it by more than one order of magnitude. Bottom loss, however, remains an important feature, and this is accounted for by multiplying each I_{scat_n} by $|\Re_b(\theta)|^2$, where θ equals θ_i at the air-sea interface for the B-S path, and θ_s at the air-sea interface for the S-B path (see Fig. 4).

For the second case, $I^{sbs}(t)$ is constructed via the convolution of impulse response functions that correspond to each sea surface interaction, $I^{sbs'}(t)$ for the first interaction and $I^{s'bs}(t)$ for the second interaction, where $I^{sbs'}(t)$ is computed as if both the seabed and second sea surface interaction are replaced by reflection processes, and $I^{s'bs}(t)$ is computed in an analogous manner (Fig. 5). Bottom loss is not computed for $I^{sbs'}(t)$ because it is already included in the computation of $I^{s'bs}(t)$, and the intensity impulse response function $I^{sbs}(t)$ results from the following convolution

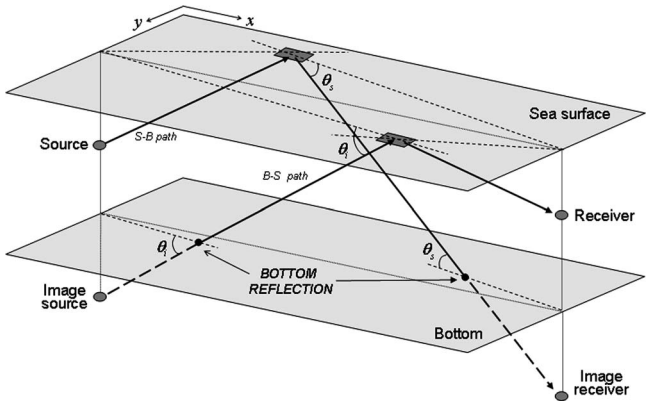


FIG. 4. The scattering geometry for S-B path and B-S path. An image receiver and image source are used for the S-B path and B-S path, respectively.

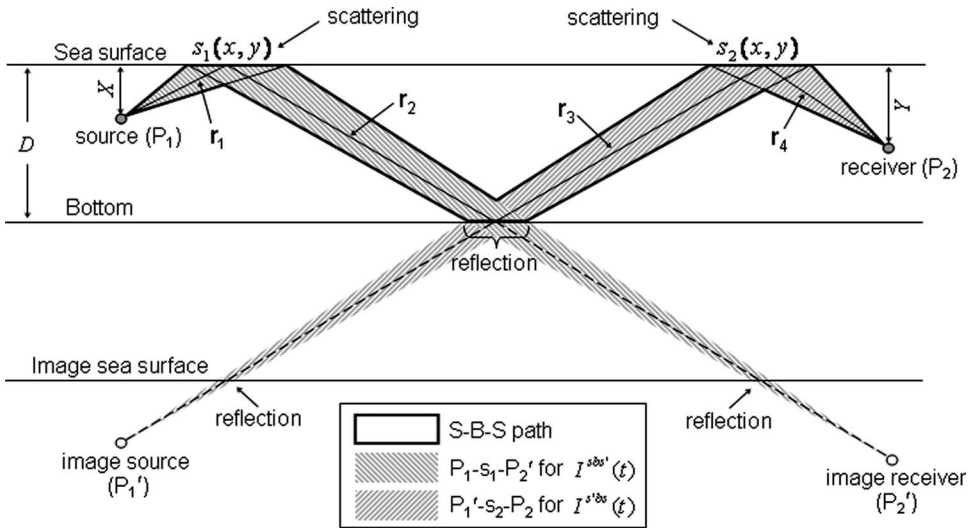


FIG. 5. Illustration of method to model the intensity impulse response for the S-B-S path, $I^{sbs}(t)$, which is constructed by convolving $I^{sbs'}(t)$ and $I^{s'bs}(t)$. These correspond to the paths P_1-s_1-P_2' and P_1'-s_2-P_2, respectively. An image receiver of depth $(2D+Y)$ and an image source of depth $(2D+X)$ are used for $I^{sbs'}(t)$ and $I^{s'bs}(t)$, respectively, where D, X, and Y are the depths shown in the figure. The ranges for specular paths are given by $r_{i=1,4}$.

$$I^{sbs}(t) = I^{sbs'} * I^{s'bs}\left(\sum_{i=1}^{4} r_i\right)^2, \tag{7}$$

where the factor in parentheses is equal to total range for the specular reflected path. This factor is required to compensate for a double counting of spreading loss that results from the convolution operation.

D. Modeling the channel intensity impulse response with multipath arrivals and comparison with CW intensity level and the matched filter envelope response

The intensity impulse response function for a direct path is calculated as $I_0 10^{-TL_d/10}\delta(t-t_d)$, where TL_d and t_d are the transmission loss and travel time, respectively, for a direct path computed using a ray-based propagation code. Combining this with the other five response functions results in an IIRF $I^c(t)$ along a common time base.

Figure 6 shows the modeled (and noise free) $I^c(t)$ that corresponds to the measurement geometry and conditions in Fig. 2, with modeling frequency equal to 8 kHz. Only the D and B paths are impulselike, whereas paths interacting with the sea surface undergo significant time spreading. For reference, the small horizontal mark near the label of each path gives the intensity impulse level when intrapath time spreading associated with scattering is ignored, and only bottom reflection loss is operative. That there is very little difference between this mark and the full-model result for the B path further suggests that reflection and consequent bottom loss dominate bottom interaction.

To compare with the ensemble-average intensity level for CW pulse, the IIRF $I^c(t)$ is convolved with a magnitude-squared representation of the transmit CW pulse. A prediction for the matched filter squared envelope $m(t)$ is derived using $I^c(t)$ computed at 12 kHz, the center frequency of the FM signal. The prediction for the FM signal is given by

$$m(t) = |H[s(t) * s(-t)]|^2 * I^c(t), \tag{8}$$

where $H[\cdot]$ denotes the Hilbert transform and $s(t)$ is the transmitted FM signal.

IV. COMPARISON OF MODEL PREDICTIONS WITH MEASUREMENTS

Figure 7 shows a comparison between the average intensity (over 20 pings) for CW signals (thick gray line), for which the center frequency of the 3-ms pulse is 8 kHz, and a model prediction (black line) also computed at 8 kHz. Surrounding the model-data comparison are maps of the distribution of bistatic cross section per unit area σ for the sea surface and seabed interfaces that were used to compute the intensity impulse response for a given ray path (diagrammed below each distribution), as outlined in the previous section. A total of six paths are modeled, yet in this 8-kHz data set, expressed on a decibel scale, prominent features of the time series are clearly reproduced by the model, with notable ex-

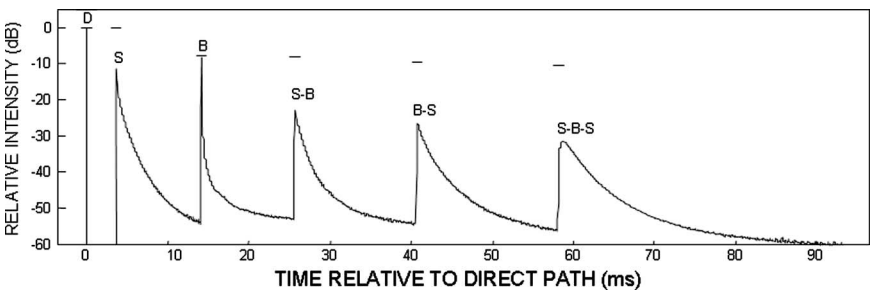


FIG. 6. Predicted (and noise free) IIRF $I^c(t)$ for frequency 8 kHz showing the primary arrivals: D, S, B, S-B, B-S, and S-B-S paths, with level referenced to that of the peak value for the direct path (D). The acquisition geometry and environmental conditions correspond to those described in Fig. 2. The small horizontal mark near each path label denotes the intensity impulse level were intrapath time spreading associated with scattering to be ignored, and only bottom reflection loss is operative.

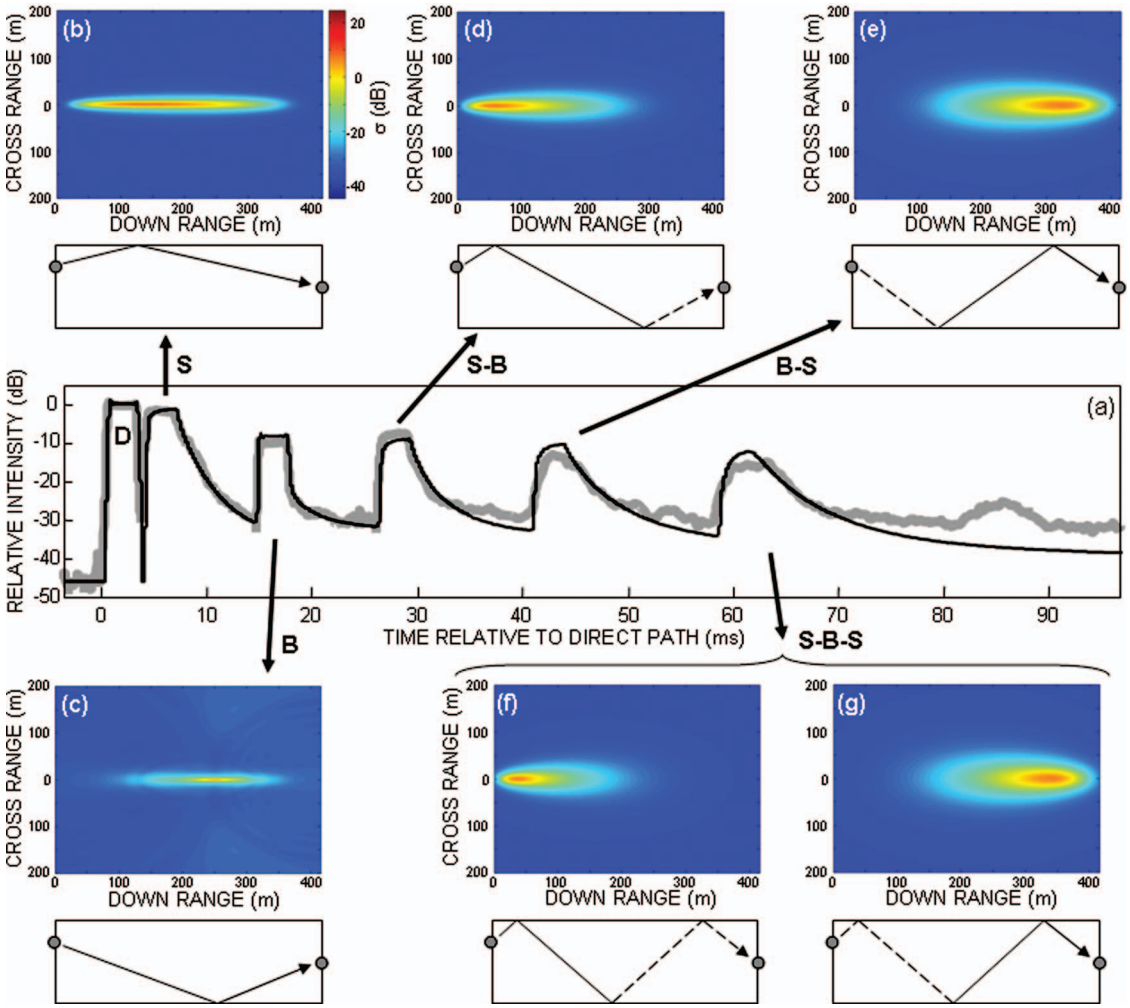


FIG. 7. (a) Comparison between the predicted intensity level (black line) and measured averaged intensity (thick gray line) for the 8-kHz CW signal from the same experimental set described in Fig. 2. The intensity level is referenced to that of the mean peak value for the direct path. Panels (b), (d), and (e) are maps of the sea surface bistatic scattering cross section for the S, S-B, and B-S paths, respectively, and (c) is a map of the bistatic cross section (incoherent +coherent) for the B path. Panels (f) and (g) show maps of the sea surface bistatic scattering cross section for computing $I^{sbs'}(t)$ and $I^{s'bs}(t)$, respectively, which are subsequently convolved to predict the impulse response for the S-B-S path. The ray diagram below each bistatic cross section map shows the corresponding specular ray path, for which a dashed line indicates a specular reflection.

ception of the seventh path that is not modeled. However, note that 99% of the time-integrated intensity is transmitted by the first six paths. Other differences of order 1 to 3 dB, tending to be located at the intensity tails of the S-B, B-S, and S-B-S paths, could result from several causes, as discussed in Sec. V.

A comparison for a 16-kHz CW pulse also shows reasonable agreement with the model prediction (Fig. 8). Recall that in this work bottom loss is computed using a two-layered sediment reflection coefficient model derived from ASIAEX measurements at the same site.[10] This model shows significant oscillations in bottom loss as a function of grazing angle for angles between 14° and 23° at 8 kHz, while beyond this range angular dependence is smoother and bottom loss at 8 and 16 kHz are nearly the same (Fig. 9). For the data set shown in Figs. 7 and 8, the nominal grazing angle for the bottom bounce (B) path is 18°, with this angle being greater than 24° for subsequent paths involving the bottom (S-B, B-S, etc.) The predicted bottom loss for the B path is 8.1 and 6.5 dB for 8 and 16 kHz, respectively, and this is consistent with the difference in intensity level between the D and B paths observed in the data (8.7 dB for 8 kHz and 7.2 dB for 16 kHz).

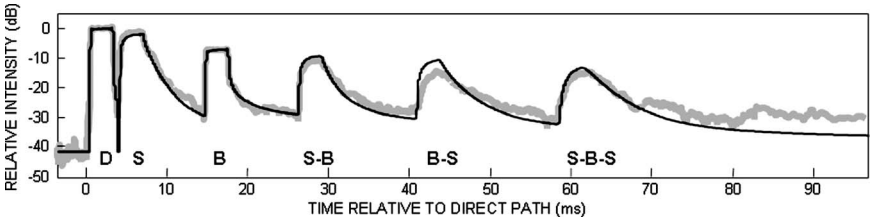


FIG. 8. Comparison between the predicted intensity level (black line) and measured averaged intensity (thick gray line) for the 16-kHz CW signal from the same experimental set described in Fig. 2. The intensity level is referenced to that of the mean peak value for the direct path.

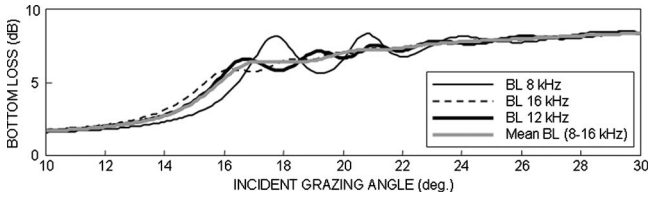


FIG. 9. Bottom loss as a function of grazing angle θ_g, or $-20\log|\Re(\theta_g)|$, based on the two-layered sediment model for the reflection coefficient $\Re$ in Ref. 10 for 8, 12, 16 kHz, and 8–16 kHz average.

An intensity average of bottom loss model curves for frequencies between 8 and 16 kHz is reasonably close to the model curve at 12 kHz (Fig. 9), which supports the approach of using 12 kHz as the modeling frequency for the 8–16-kHz FM signal. Figure 10 shows the averaged envelope (over 20 pings) of the matched filter response for the FM signals from two experimental runs, wind speed 7 m/s with range 416 m, (i.e, the May 31st data from Fig 2), and wind speed 5 m/s with range 456 m, taken at 01:05 UTC June 1. The measurements compare well with their model predictions including fair agreement in the relative amplitudes between the direct and boundary interaction paths, and in the relative amplitudes between the two experimental runs. In terms of the latter, both model and data for the 416-m range measurements (upper figure) display lower amplitudes relative to the direct path compared to the 456-m range measurements (lower figure). It is postulated that this difference arises from a combination of (1) the greater time spreading and subsequently lower peak amplitude in matched filtering seen in the 416-m measurements made at 7 m/s wind speed involving the sea surface (~1 dB effect) and (2) the greater transmission loss (in excess of geometric spreading) in the direct path owing to refraction effects for the 456-m range measurements, to which follow-on paths are not subjected because of their steeper propagation angles (~2 dB effect).

V. DISCUSSION

It is of interest to examine further the relation among each of the arrivals in the modeling process, including the small influence of near-surface bubbles in forward scattering from the sea surface. This is best illustrated with a CW model result (Fig. 11), shown previously in comparison with measurements (Fig. 8), but shown here with and without the influence of bubbles and with individual paths prior to their incoherent superposition. The contribution from near-surface bubbles for the paths associated with sea surface interaction (e.g., S, S-B, B-S, and S-B-S paths) becomes significant approximately 7 ms after the arrival onset (nominal specular arrival), a point at which scattering from bubbles exceeds Bragg scattering from the rough surface. For the S path, bubble scattering extends intrapath time spread significantly beyond that due to rough surface scattering, albeit at intensity levels ~30 dB below peak level. This effect is not clearly evident in observations (Fig. 8) owing to the arrival of the B path; however, in other geometries wherein these paths have greater time separation, the effect is observable.[21]

The intensity level of the B path shows little time spread, consistent with the Rayleigh parameter χ being ~0.2 at 16 kHz, and thus reflection dominates. However, after one pulse duration (3 ms) the time decay is similar to that of bubble scattering but with higher levels. Such decay is due to the contribution from both sustained Bragg scattering and scattering from sediment volume inhomogeneities as described in the APL-UW model,[19] and results here suggest that bottom scattering is a major contributor to the background floor for multipath return.

There exist 1- to 3-dB differences between the model and data curves displayed in Figs. 7 and 8 that show up primarily at the intensity tails of the S-B, B-S, and S-B-S paths. Possible sources for these differences are (1) late-arriving scattered returns from buoys and other experimental equipment (see Fig. 1), (2) multiple scattering (via surface-bottom and bottom-surface paths) not accounted for in the modeling, and (3) scattering from the fine-scale sediment layers, which is the cause of the disagreement between model and bottom scattering and reflection measurements.[22]

In this work model predictions for multipath propagation at ranges of order 0.5 km were limited to the sixth arrival, in view of the fact that subsequent arrivals were severely attenuated owing to bottom loss. However, as the source-receiver range increases, the bottom loss will be lessened, i.e., the bottom grazing angle goes precritical. Still, the limitation to the sixth arrival is a reasonable assumption even out to longer ranges. This is illustrated with a set of measurements of the arrival structure at ranges varying from 485 to 1024 m made with the same FM probe pulse on 6

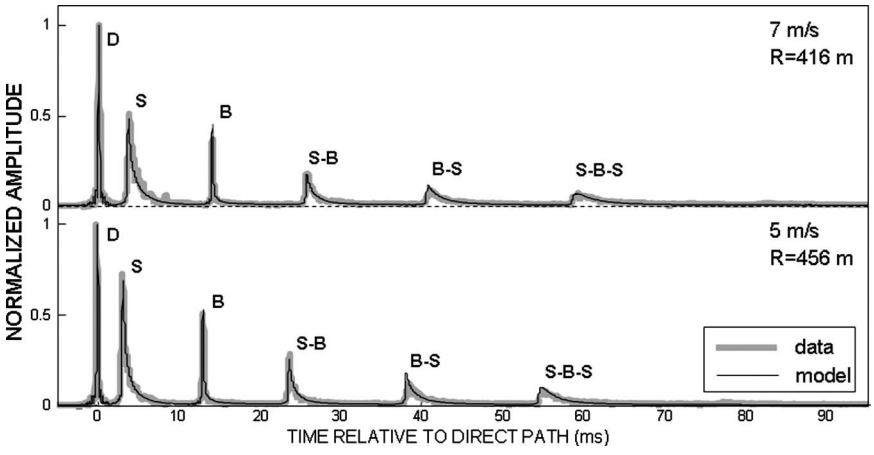


FIG. 10. Averaged envelope for the matched filter response over 20 pulses for measurements made 31 May at 13:54 UTC (upper plot) during which wind speed was 7 m/s (same measurement period and geometry as data shown in Figs. 2, 7, and 8), and measurements made 1 June at 01:05 UTC (lower plot) during which wind speed was 5 m/s, compared with equivalent model. The source and receiver depths for the lower plot data are 25 and 52 m, respectively. The averaged envelope is the square root of a linear intensity average over 20 transmissions, and the amplitude is referenced to that of the peak value for the direct path.

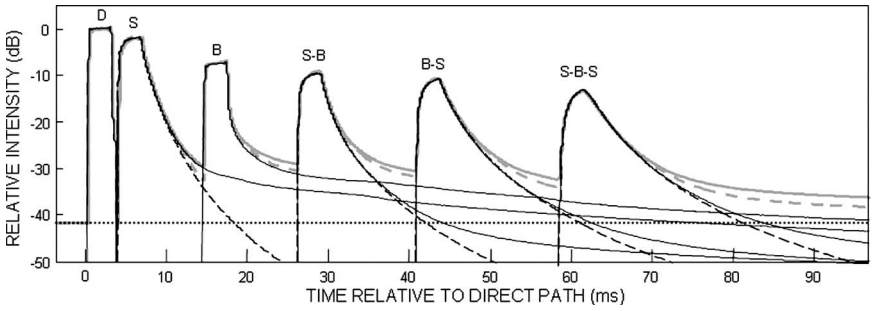


FIG. 11. Simulated total intensity level (gray line) and the contribution from each arrival (black line) for a 16-kHz CW pulse. Solid lines and dashed lines (except B path) represent the intensity level predicted with and without the effect of bubbles, respectively. The dotted line shows the noise level corresponding to the measurement set shown in Fig. 8, which is intensity summed for the total intensity level.

June beginning at 00:55 UTC and lasting about 2 h (Fig. 12). For these measurements the source and receiver depths were 50 and 26 m, respectively, wind speed was 1–2 m/s, and rms waveheight was 0.3 m. Note here that each signal represents a single transmission, and the sixth arrival, corresponding to the S-B-S path, is shaded for each case. The gross multipath time spreading clearly reduces with increasing range, yet six paths continue to dominate the arrival structure out to ranges of order 1 km (~97% of the energy in the received waveform at range 1024 m is transmitted by six paths[23]).

VI. SUMMARY

Results of measurements and modeling of the IIRF for the ASIAEX site in the East China Sea have been presented. The measurements were made in waters 105 m deep and at ranges between approximately 500 and 1000 m. The modeled IIRF, $I^c(t)$, is equivalent to a time-averaged response to an impulse and embodies the boundary scattering and reflection physics corresponding to the center frequency at which the computations are made. Upon convolving $I^c(t)$ with, for example, the squared envelope of a CW pulse, the result can be compared with field measurements of averaged intensity.

The numerical model representation of $I^c(t)$ consists of the incoherent addition of separate response functions, $I^p(t)$, associated with each eigenray, e.g., the direct path, the single-surface bounce (S), single-bottom bounce (B), plus paths involving multiple boundary interaction out to the sixth-arriving path. The impulse responses for S and B paths are described by the distribution of the bistatic cross section over the boundary surface. Responses for paths involving multiple boundary interactions such as the surface-bottom (S-B) path are computed in a similar manner but involve image reflection at the seabed; the S-B path invokes an image receiver and the B-S path invokes an image source. The sixth-arriving multipath is the surface-bottom-surface (S-B-S) path; by hypothesis, its response is constructed via the convolution of two partial impulse responses corresponding to the first and second interactions with the sea surface.

The model predictions compared reasonably well with the measurements involving CW pulses with center frequency of 8 and 16 kHz, and a linear frequency-modulated (FM) pulse of frequency range 8–16 kHz and duration 20 ms. For the FM pulse the matched filter response from measurements represents a band-limited approximation to the IIRF and these results compared well with a model representation generated by convolving $I^c(t)$ with the squared autocorrelation function of the transmitted waveform and taking the square root of the result.

It is found that six paths generally dominate the response at ranges 500 to 1000 m for this East China Sea channel, as shown by the measurements out to range ~1000 m, with the number of paths effectively limited by bottom loss. The basic structure of $I^c(t)$ is therefore set by seabed properties and acquisition geometry (source depth, receiver depth, range) with some differences in intrapath spreading dependent on sea surface conditions.

For the ASIAEX acquisition geometries studied here, for which the S path arrives before the B path, the intrapath time spread for the B path appears to be the major contributor to the background floor for multipath return. Note, however, that if the source and receiver are close to the seabed, such

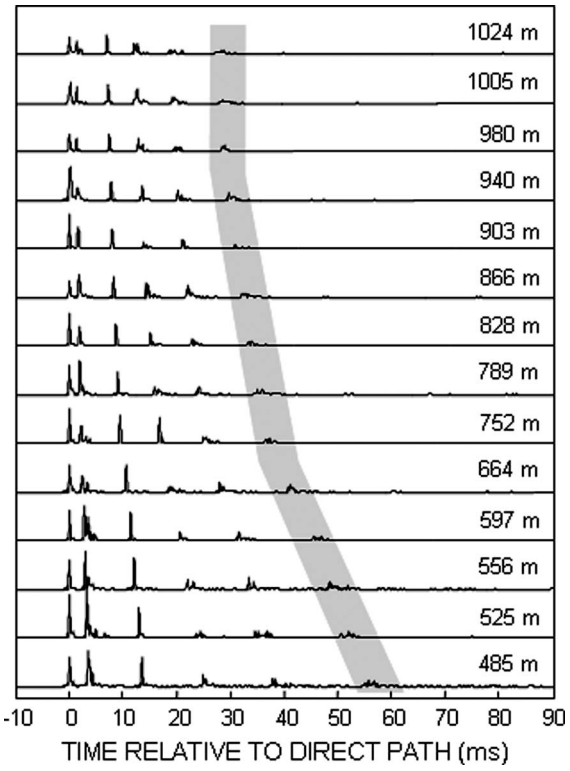


FIG. 12. The amplitude of the matched filter envelope response as a function of range for a single transmission of a FM pulse, frequency range 8–16 kHz. Timing for the arrivals are referenced to that of the direct path for each case (with source-receiver range noted on right) and the shaded gray area denotes the sixth arrival (S-B-S path). Measurements were taken on 6 June beginning at 00:55 UTC and lasting about 2 h during which the wind speed and rms waveheight were 1–2 m/s and ~0.3 m, respectively.

that the B path arrives before the S path, the modeling suggests that time spreading associated with near-surface bubbles will set the background floor.

ACKNOWLEDGEMENTS

This work was funded by the Office of Naval Research Code 321 Ocean Acoustics.

[1]D. E. Weston and P. B. Rowlands, "Guided acoustic waves in the ocean," Rep. Prog. Phys. **42**, 347–387 (1979).
[2]P. H. Dahl, R. Zhang, J. H. Miller, L. R. Bartek, Z. Peng, S. R. Ramp, J.-X. Zhou, C.-S. Chiu, J. F. Lynch, J. A. Simmen, and R. C. Spindel, "Overview of results from the Asian Seas International Acoustic Experiment in the East China Sea," IEEE J. Ocean. Eng. **29**, 920–928 (2004).
[3]M. Stojanovic, "Retrofocusing techniques for high rate acoustic communications," J. Acoust. Soc. Am. **117**, 1173–1185 (2005).
[4]D. B. Kilfoyle and A. B. Baggeroer, "The state of the art in underwater acoustic telemetry," IEEE J. Ocean. Eng. **25**, 4–27 (2000).
[5]C. Bjerrum-Niese and R. Lützen, "Stochastic simulation of acoustic communication in turbulent shallow water," IEEE J. Ocean. Eng. **25**, 523–532 (2000).
[6]D. Rouseff, D. R. Jackson, W. L. J. Fox, C. D. Jones, J. A. Ritcey, and D. R. Dowling, "Underwater acoustic communication by passive-phase conjugation: theory and experimental results," IEEE J. Ocean. Eng. **26**, 821–831 (2001).
[7]W. J. Higley, P. Roux, W. A. Kuperman, W. S. Hodgkiss, H. C. Song, and T. Akal, "Synthetic aperture time-reversal communications in shallow water: experimental demonstration at sea," J. Acoust. Soc. Am. **118**, 2365–2372 (2005).
[8]J.-P. Hermand, "Broad-band geoacoustic inversion in shallow water from waveguide impulse response measurements on a single hydrophone: theory and experimental results," IEEE J. Ocean. Eng. **24**, 41–66 (1999).
[9]P. H. Dahl, "Forward scattering from the sea surface and the van Cittert-Zernike theorem," J. Acoust. Soc. Am. **115**, 589–599 (2004).
[10]J. W. Choi and P. H. Dahl, "Mid-to-high-frequency bottom loss in the East China Sea," IEEE J. Ocean. Eng. **29**, 980–987 (2004).
[11]D. Tang, "Fine-scale measurements of sediment roughness and subbottom variability," IEEE J. Ocean. Eng. **29**, 929–939 (2004).
[12]J. K. Lewis, J. Rudzinsky, S. Rajan, P. J. Stein, A. Vandiver, and The KauaiEx Group, "Model-oriented ocean tomography using higher frequency, bottom-mounted hydrophones," J. Acoust. Soc. Am. **117**, 3539–3554 (2005).
[13]N. C. Makris, "The effect of saturated transmission scintillation on ocean acoustic intensity measurements," J. Acoust. Soc. Am. **100**, 769–783 (1996).
[14]P. Ratilal, Y. Lai, D. T. Symonds, L. A. Ruhlmann, J. R. Preston, E. K. Scheer, M. T. Garr, C. W. Holland, J. A. Goff, and N. C. Makris, "Long range acoustic imaging of the continental shelf environment: The Acoustic Clutter Reconnaissance Experiment 2001," J. Acoust. Soc. Am. **117**, 1977–1998 (2005).
[15]P. H. Dahl, "High-frequency forward scattering from the sea surface: The characteristic scales of time and angle spreading," IEEE J. Ocean. Eng. **26**, 141–151 (2001).
[16]R. L. Culver and D. L. Bradley, "On the relationship between signal bandwidth and frequency correlation for ocean surface forward scattered signals," J. Acoust. Soc. Am. **118**, 129–138 (2005).
[17]K. L. Williams and D. R. Jackson, "Bistatic bottom scattering: model, experiments, and model/data comparison," J. Acoust. Soc. Am. **103**, 169–181 (1998).
[18]P. H. Dahl, "On bistatic sea surface scattering: Field measurements and modeling," J. Acoust. Soc. Am. **105**, 2155–2169 (1999).
[19]D. R. Jackson, *A model for Bistatic Bottom Scattering in the Frequency Range 10–100 kHz*, APL-UW TR 9305, Applied Physics Laboratory, University of Washington (August 1993).
[20]P. H. Dahl and W. L. J. Fox, "Measurement and interpretation of angular spreading from multiple boundary interactions in a shallow water channel," in *High Frequency Acoustics in Shallow Water*, edited by N. G. Pace, E. Pouliquen, O. Bergem, and A. P. Lyons (NATO SACLANT Undersea Research Centre, La Spezia, Italy, 1997), pp.107–114.
[21]P. H. Dahl, "The sea surface bounce channel: bubble-mediated energy loss and time/angle spreading," in *High Frequency Ocean Acoustics*, edited by M. B. Porter, M. Siderius, and W. A. Kuperman (AIP, New York, 2004).
[22]C. W. Holland, "Coupled scattering and reflection measurements in shallow water," IEEE J. Ocean. Eng. **27**, 454–470 (2002).
[23]P. H. Dahl and J. W. Choi, "The East China Sea as an underwater acoustic communication channel: measurements of the channel impulse response (U)," U.S. Navy J. Underwater Acoustics, submitted.

Detailed study of a traveling wave thermoacoustic refrigerator driven by a traveling wave thermoacoustic engine

Wei Dai and Ercang Luo
Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing, China, 100080

Yong Zhang and Hong Ling
Graduate School of Chinese Academy of Sciences, and Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing, China, 100080

(Received 23 June 2005; revised 28 October 2005; accepted 13 February 2006)

Thermoacoustic systems have very attractive features and possible wide applications in many areas, especially for cooling purposes. Inventions of traveling wave thermoacoustic engines and traveling wave thermoacoustic refrigerators with work recovery capability have greatly improved the thermodynamic efficiencies of the thermoacoustic systems. To fully utilize the advantages of traveling wave systems, a traveling wave work-recoverable thermoacoustic refrigerator has been designed and built that is driven by a traveling wave thermoacoustic engine, which is aimed at domestic refrigeration purposes. So far, the lowest temperature of −64.4 °C and 250 W cooling power at −22.1 °C are obtained by the system with 3.0 MPa helium gas and 57.7 Hz working frequency. Heat input into the system is 2.2 kW. Simulations based on linear thermoacoustic theory have also been done for comparison with experimental results, which shows reasonable agreement within a certain pressure wave amplitude range and cold end temperature range. The comparisons meanwhile imply some serious loss mechanism inside the traveling wave thermoacoustic refrigerator. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184267]

PACS number(s): 43.35.Ud, 43.25.−x [RR] Pages: 2686–2692

I. INTRODUCTION

Thermoacoustic systems use high-intensity sound waves to realize the conversion between heat and acoustic power. According to the direction of the conversion, they can be classified into two kinds of systems: thermoacoustic engines and thermoacoustic refrigerators (or coolers). The invention of traveling wave thermoacoustic engines[1] (TWTEs) and traveling wave thermoacoustic refrigerators with work recovery capability[2] (TWTRs) has greatly improved the thermodynamic efficiencies over previous standing-wave-based thermoacoustic systems. The main reason for the improvement is that traveling wave acoustic fields with high local impedance can be established inside the regenerators of these systems. Moreover, in the TWTRs, acoustic work coming from the cold end is recovered which further improves its efficiency especially within the domestic refrigeration temperature range. Naturally, combining the TWTE with the TWTR will lead to a highly efficient thermoacoustic refrigerator, which has other attractive features such as a heat-driven mechanism, no moving parts, environmental friendliness, etc. This system may find important applications in domestic refrigeration since present CFC- or HCFC-based technologies present serious threats to the environment. On the basis of our previous work on a TWTR driven by a standing wave thermoacoustic engine[3] and an improved design of the TWTE,[4] we have built an experimental setup of the TWTR driven by the TWTE and tested the cooling performance. In the second section, we introduce the details of the experiments. The third section introduces the simulation model and gives comparison between experiments and calculation results. Finally, conclusions and discussions are made.

II. EXPERIMENT DETAILS

A. Experimental setup

According to the computation results of our simulation program (to be introduced later), we have built an experimental setup which is illustrated in Fig. 1. It includes two subsystems: the TWTE and the TWTR, which is coupled at the second tee.

The TWTE loop mainly includes a feedback tube, a main water-cooled heat exchanger, a regenerator, a heater block, a thermal buffer tube, a secondary water-cooled heat exchanger, and a resonance tube. Details of the part dimensions are listed in Table I. The TWTR loop mainly includes a feedback tube, a water-cooled heat exchanger, a regenerator, a cold end heat exchanger, a thermal buffer tube, and a flow straightener. Details of the part dimensions are listed in Table II. Core parts of all the heat exchangers use similar design.[5]

As is known in thermoacoustic theory, Gedeon streaming[6] imposes strong influence on the thermal performance of a thermoacoustic system where there exists a loop. For simplicity, we insert a piece of elastic membrane above the upper water-cooled heat exchangers inside both subsystems, where the amplitudes of volume flow rate are relatively small. The benefit of using membranes is that it enables us to fairly compare the system's experimental performance with numeric calculations without the interference of the Gedeon streaming which is very difficult to mea-

 0001-4966/2006/119(5)/2686/7/$22.50

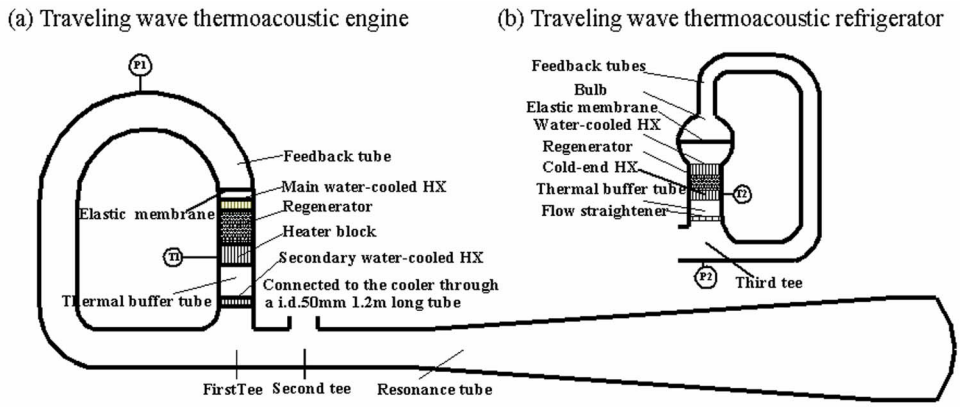


FIG. 1. Illustration of the traveling wave thermoacoustic refrigerator driven by the traveling wave thermoacoustic engine.

sure if it exists. Meanwhile, we use straight thermal buffer tubes in both the TWTE and TWTR without consideration of suppressing the Rayleigh streaming inside,[7] which could also be a big source of thermal loss.

One interesting feature of the system is that the cooler is not directly coupled to the engine at the second tee position but through a 50-mm i.d. 0.2-m-long globe valve and a 50-mm i.d. 1.0-m-long stainless steel tube connected in sequence. The purpose of this tube is to increase the pressure ratio available to the cooler, which is the result of our analysis and also has been verified through experiments. The globe valve is just for easy operation of the system.

For the TWTE, electric heaters are used for heat input into the heater block. Ceramic fiber bricks are used to thermally insulate the hot parts of the system from the surroundings. In the experiments, we found that the material is not so satisfying due to its softness and brittleness. *K*-type sheathed thermocouple T_1 is used to measure the temperature of a solid point on the outside shell of the heater block. For the TWTR, electric heaters are also used for cooling power measurements. Plastic foam is used to thermally insulate the cold parts from the surroundings. A Pt-100 resistance thermometer T_2 is used to measure the cold end temperature. Two pressure sensors P_1 and P_2 measure the pressure waves at the loop top of the TWTE and the inlet of the TWTR, respectively.

B. Experiment results

Generally, when heat is input through the TWTE's heater block, spontaneous oscillation will start once the temperature inside exceeds a critical value. The generated pressure wave will drive the cooler to cool down. With the setup charged with helium gas, we have done a series of experiments with different average pressures and heating powers. Maximum heating power is limited due to the material limitation of the heater block, where the temperature is always controlled around 650 °C by closely observing the temperature sensor T_1. Under all the working conditions, the system frequency slightly changes around 57.5 Hz with variations smaller than 0.5 Hz.

Figures 2(a)–2(c) gives the cooling power versus cold end temperature with different average pressures and heating powers. Generally speaking, a higher average pressure makes it possible for a larger heat input without breaking the temperature limit set by the heater block materials. This increases the acoustic power density within the system and thus leads to a larger maximum cooling power and a lower no-load cold end temperature within the above-mentioned high temperature limit. So far, a lowest temperature of −64.4 °C and a cooling power of 250 W at −22.1 °C are obtained with 2.2 kW heating power and 3.0 MPa average

TABLE I. Details of the traveling wave thermoacoustic engine.

Parts name	Dimension	Other details
Feedback tube	80-mm i.d., length 1.6 m	
Water-cooled heat exchanger (HX)	45 mm long with about 0.32 m^2 heat exchange area with gas	Mainly made from copper, housed in 80-mm-i.d. stainless steel tube
Regenerator	80-mm i.d., length 80 mm	120 mesh stainless screens
Heater block	58 mm long with about 0.24 m^2 heat exchange area with gas	Mainly made from copper, housed in an 80-mm-i.d. stainless steel shell
Thermal buffer tube	80-mm i.d., length 240 mm	Straight
Secondary water-cooled heat exchanger	22 mm long with about 0.16 m^2 heat exchange area with gas	Made from copper, housed in 80-mm-i.d. stainless steel tube
Resonance tube	1.5-m 80-mm-i.d. +3.0-m tapered tube (80- to 200-mm i.d.) +0.5-m 200-mm-i.d. tube	

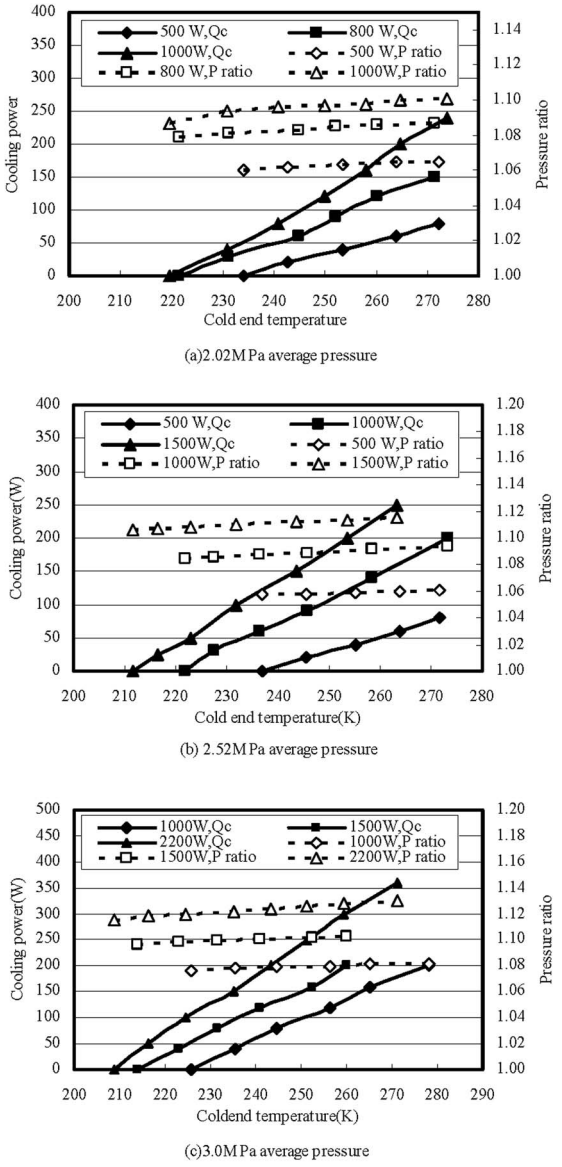


FIG. 2. Measured TWTR cooling power and inlet pressure ratio versus cold end temperature with different heating powers and average pressures: (a) 2.02 MPa, (b) 2.52 MPa, and (c) 3.0 MPa.

pressure as shown in Fig. 2(c). The pressure ratio at the inlet of the cooler is about 1.14, which is the maximum value among all the experiment data.

Furthermore, Fig. 3 gives the comparison between the data with different average pressures but with the same heating power of 1.0 kW. It can be seen that lower average pressure leads to a lower values of lowest temperature and a larger cooling power at a certain temperature. But this does not mean that the relative thermal efficiency (the ratio between the practical thermal efficiency and the Carnot efficiency) is also better, since a lower average pressure gives a higher heater block temperature which leads to a higher value of ideal Carnot efficiency.

With these experimental data, we start the numeric simulation and compare the calculated results with experiments.

III. SIMULATION DETAILS

A. Linear thermoacoustic model

The computational model is based on linear thermoacoustic theory, the details of which have been given in Refs. 8 and 9 Minor losses and turbulence influence are not included in the model. For conciseness, the main control equations in complex form are

$$\frac{d\hat{U}}{dx} + R_1\hat{p} - R_3\hat{U} = 0, \tag{1}$$

$$\frac{d\hat{p}}{dx} + R_2\hat{U} = 0, \tag{2}$$

$$\frac{dH}{dx} = Q, \quad H = c_1 + c_2\frac{dT_x}{dx}, \tag{3}$$

where $\hat{p}, \hat{U}$ are complex pressure wave and volume flow rate, T_x is the mean temperature of the gas, H is the total energy including integrated enthalpy flow over the fluid passage cross section and static heat conduction, and Q is the heat input to the gas from the outside of the system (e.g., heat input at the heater block, heat brought away by the water passing the water heat exchanger). R_1, R_2, R_3, c_1, c_2 are complex coefficients, which are functions of the fluid passage geometry, frequency, average pressure, volume flow rate, temperature distribution, etc.

With these equations, simulation can be done with given boundary conditions. Different from DeltaE,[10] we use a mixed algorithm combining the advantages of both analytical and numeric methods. First, the whole system is divided into a certain number of cells. Analytically solving Eqs. (1) and (2) gives the relationship of two neighboring cells in the form of a transfer matrix:

$$\begin{bmatrix} \hat{p}(n+1) \\ \hat{U}(n+1) \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \hat{p}(n) \\ \hat{U}(n) \end{bmatrix}, \tag{4}$$

where n is the sequence number of the cell. Each element in the 2×2 transfer matrix is a function of R_1, R_2, R_3. For Eq. (3), it can be written with a first-order accuracy on most computational cells in a discrete form,

$$B_1T_x(n-1) + B_2T_x(n) + B_3T_x(n+1) = B_4, \tag{5}$$

except for some special places (e.g., the tee) where a few more cells get involved. Here, $B_1 - B_4$ are coefficients related to the fluid passage geometry, frequency, average pressure, volume flow rate, temperature distribution, etc. Combining this form of equation of all the computational cells leads to a sparse matrix which can be easily solved using linear algebra algorithms.

B. Solving process

For thermoacoustic systems consisting of thermoacoustic engines, two parameters need to be guessed at the beginning of the computation: the frequency and the pressure amplitude at the starting cell. With the guessed values and assumed temperature distribution, $\hat{p}, \hat{U}$ of all cells can be

TABLE II. Details of the traveling wave thermoacoustic refrigerator.

Parts Name	Dimension	Other details
Feedback tube	32 mm i.d., 0.7 m long	
Water-cooled heat exchanger (HX)	35 mm long with about 0.09 m^2 heat exchange area with gas	Made from copper, housed in 50-mm-i.d. stainless steel tube[3]
Regenerator	50 mm i.d., length 35 mm	120 mesh stainless screens
Cold end heat exchanger	30 mm long with about 0.077 m^2 heat exchange area with gas	Made from copper
Thermal buffer tube	50 mm i.d., length 50 mm	Straight
Flow straightener	Several pieces of small mesh number stainless steel screens	

calculated from Eq. (4). Then Eq. (5) is used to calculate the new temperature values. According to the results, the temperature distribution of the system is modified. Since elements of the matrix in Eq. (4) are functions of temperature, $\hat{p}, \hat{U}$ needs to be recalculated. This iteration process should be continued until certain criteria are met. By comparing the calculated volume flow rate with the physical boundary conditions of the real system, we can judge whether the guessed frequency is reasonable. If not, we modify the frequency and do the calculations of $\hat{p}, \hat{U}$ again. Then the distribution of acoustic work flow can be evaluated to see if the guessed pressure amplitude at the starting cell is reasonable. If not, we do the calculations all over again with a new guessed pressure amplitude value. After these iteration steps, the final distributions of $\hat{p}, \hat{U}$ and T_x are obtained, as well as the acoustic work flow, total energy flow, and heat flow. Generally, the calculation time of an operating point is less than 5 min with relatively good initial guessed values.

In the model, helium gas is assumed as being ideal and has temperature-dependent heat conductivity and viscosity. The regenerator materials also have temperature-dependent heat conductivity and specific heat.

The definition of length axis according to Fig. 1 is given below. The origin is set at the first tee and the axis passes through the TWTE loop clockwise. Then it passes through a short 80-mm i.d. connecting tube, arrives at the second tee, and from there it passes through the 50-mm i.d. 1.2-m-long section and enters the TWTR loop counter-clockwise. When it returns to the inlet of the TWTR, the axis jumps to the position where the resonance tube is connected with the second tee and moves to the right end of the resonance tube. This jumping causes the rightmost abrupt changes in the curves of Fig. 4 and the acoustic power curve in Fig. 5.

C. Simulation results and comparison with experiment results

Due to the limitation of linear thermoacoustic theory when the pressure ratio is high, our comparisons focus on the operating conditions which have pressure ratios lower than 1.1 and TWTR cold end temperature around 270 K, thus influences of turbulence and Rayleigh streaming in the thermal buffer tube will not be serious.

After initializing the geometric parameters, boundary conditions, average pressure, and heat input, the calculation can start. It should be addressed here that due to the uncertainty of heat leakage through the insulation bricks at the TWTE's heater block, we input into the program with the measured heating power multiplied by a factor around 0.8–0.9, which depends on the heater block temperature.

Before discussing the calculation results, an important thermoacoustic concept of acoustic work W is defined as

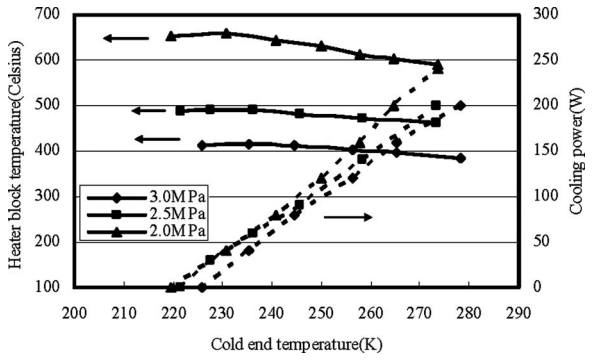


FIG. 3. Measured TWTE heater block temperature and TWTR cooling power versus TWTR cold end temperature, with heating power 1 kW and different average pressures.

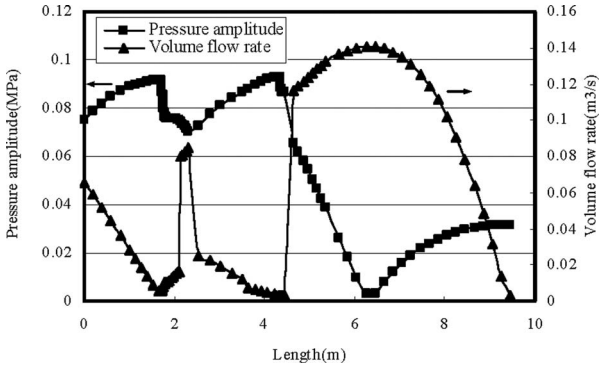


FIG. 4. Calculated distributions of pressure amplitudes and volume flow rates inside the system: average pressure 2.52 MPa, 58 Hz, heating power 500 W and TWTR cold end temperature 271.7 K.

TABLE III. Comparison between experimental results and calculated results with working conditions being 2.52 MPa average pressure, 500 W heating power, and 271.7 K TWTR cold end temperature.

	Frequency (Hz)	Pressure ratio at the TWTR inlet	Temperature at heater block (K)[a]	Cooling power of the TWTR(W)
Experiment	57.4	1.069	597	80
Calculation	58.0	1.071	515	90.3

[a]The measured temperature is on a solid point near the outer surface of the heater block, while the calculated temperature gives the highest gas temperature inside the heater block, which may cause big errors

$$W = \tfrac{1}{2}\mathrm{Re}(\hat{P}\hat{U}^{*}) \tag{6}$$

where Re means taking real part and * means complex conjugation. In analogy with the work done by a sinusoidally moving mechanical piston, this concept denotes how much work the virtual gas pistons at any x position can do. Most importantly, the magnitude of the acoustic work at the cold end of TWTR's thermal buffer tube roughly determines how much gross cooling power can be obtained.

First, we calculate an operating point with average pressure 2.52 MPa, the TWTR's cold end temperature 271.7 K, and heating power 500 W. Table III lists the comparison between experimental values and calculated values. The calculated results agree surprisingly well with experimental results, especially for frequency. With this good agreement, we believe that all the other calculated parameters will not deviate much from the reality. So, a deep look into the details of different parameters inside the system will help us to evaluate the system performance and find possible problems.

Figure 4 gives the distribution of the amplitudes of the oscillating pressures and volume flow rates. Except for the rightmost one, the other abrupt changes in volume flow rate curve are simply due to the bifurcation in the three tees shown in Fig. 1. It is clear that the pressure node and volume flow rate anti-node appear inside the resonance tube near the joint between the 1.5-m-long straight tube and the 3-m-long tapered tube. According to our analysis,[4] this is not a good operating condition for the resonance tube, which makes the tapered part much less useful. Figure 5 gives the distribution of temperature and acoustic work flow. Similarly, except for the rightmost one, the other abrupt changes in the acoustic power curve are simply due to the bifurcations at the three tees shown in Fig. 1. It can be seen in the figure that about 340 W of acoustic work comes out of the TWTE's heater block. At the first tee of the TWTE where the acoustic work gets dissipated to 336 W, most of the acoustic work ($\approx$272 W) is fed back through the feedback tube, flows through the regenerator, and is amplified to about 346 W. The remaining acoustic work ($\approx$64 W) flows rightward from the first tee, gets dissipated slightly, and arrives at the second tee where about 31 W goes to drive the cooler and 30 W gets dissipated inside the resonance tube. Inside the TWTR loop, 102 W acoustic work flows out of the cold end heat exchanger and, together with the 27 W (31 W minus 4 W due to the dissipation in the 50-mm i.d. tube) flows back to the ambient end of TWTR's regenerator. Then about 20 W of acoustic work is consumed to pump the heat from the cold end heat exchanger to the water-cooled heat exchanger. Furthermore, Fig. 6 shows the details of phase difference between pressure wave and volume flow rate inside the regenerators of both the TWTE and TWTR. It can be seen that the phase differences are generally larger than zero, especially for the regenerator of the TWTR. This is not good since we want to achieve an in-phase relationship in the middle of the regenerators for high efficiencies.

It should be noted that the above-mentioned case is only one of the few cases in which the calculations match well with the experiments. Keeping the same average pressure and heating power, Fig. 7 gives the comparison between calculation and experimental results of cooling powers and pressure wave amplitudes at the inlet of the TWTR when the cold end temperatures are lower. It can be seen that for the pressure wave amplitude, the differences between calculations and experiments are within 7%. In terms of acoustic

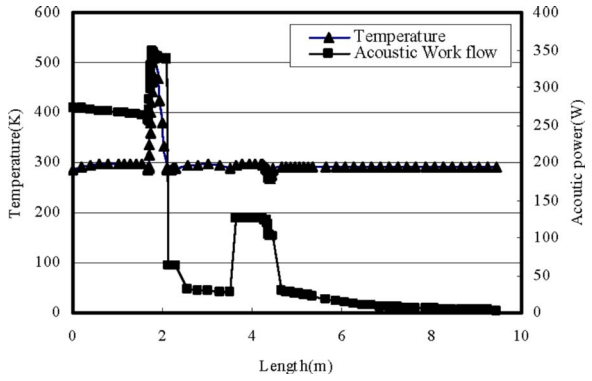


FIG. 5. Calculated distributions of temperatures and acoustic works inside the system: average pressure 2.52 MP, 58 Hz, heating power 500 W, and TWTR cold end temperature 271.7 K.

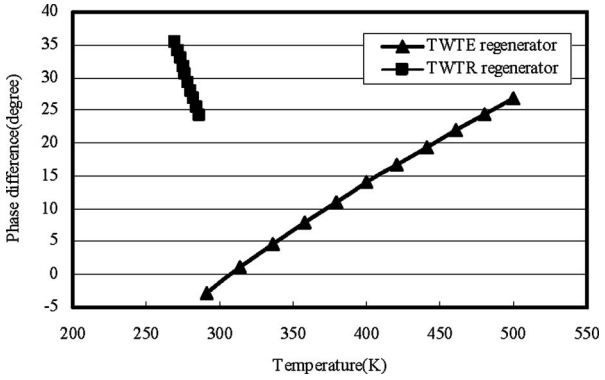


FIG. 6. Calculated distribution of phase relationship between pressure waves and volume flow rates inside the two regenerators. For clarity, the x axis uses temperatures which correspond to the positions inside the regenerators. Average pressure 2.52 MPa, 58 Hz, heating power 500 W, and TWTR cold end temperature 271.7 K.

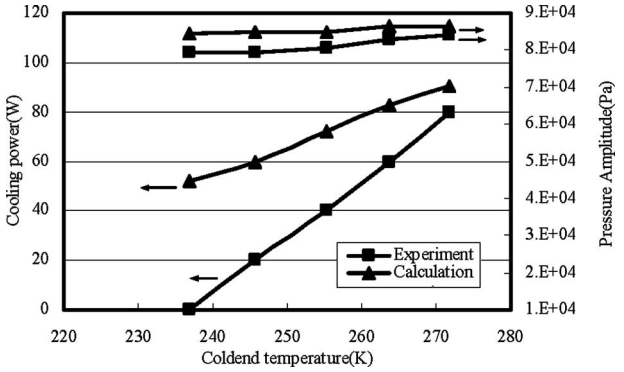


FIG. 7. Comparison between calculated and experimental results on TWTR's cooling power and inlet pressure amplitude versus TWTR cold end temperature, average pressure 2.52 MPa, and heating power 500 W.

power delivered to the cooler (proportional to the squared pressure wave amplitude), this means only about 15% difference, which cannot explain the large discrepancy between calculated and experimental cooling powers at temperatures below 265 K. The reason can only be found in the TWTR subsystem itself. Indeed, it is shown in Fig. 7 that the curve slope of the calculated cooling power is apparently smaller than that of the experiments, which meanwhile implies that the TWTR should be able to generate much lower temperature than that of experiments. Since pressure amplitude at the inlet of the TWTR varies only slightly with the cold end temperature in both experiments and calculations and the two curves gets rather close at around 270 K, this may indicate that some serious loss mechanism related to the cold end temperature exists in the experimental setup. Similar comparisons done with other operating conditions show the same phenomenon.

On the other hand, as we try to simulate other working conditions with larger pressure ratios, the deviation of calculated pressure amplitudes from experimental values also becomes more apparent, which, in addition to the possible losses addressed above, leads to a larger discrepancy of calculated cooling powers from experimental ones. Table IV gives comparisons of measured and calculated TWTR inlet pressure amplitudes under several typical operating conditions. The relative differences are not so big, but in terms of acoustic power, which is proportional to the squared pressure amplitude, the power difference can be as high as 54% for the first data point. This large discrepancy is not surprising since we have neglected the influence of nonlinear behavior (especially the turbulence) in our linear model.

IV. CONCLUSIONS

We have designed and built up an experimental setup of a traveling wave thermoacoustic refrigerator driven by a traveling wave thermoacoustic engine. Many experiments are done to test the performance with different average pressures and heating powers. So far, lowest temperature of −64.4 °C and highest cooling power of 250 W at −22.1 °C are obtained with 2.2 kW heating power, 3.0 MPa average pressure, and pressure ratio of 1.14 at the inlet of the cooler. The results show good application potential for the system in the field of domestic refrigeration, especially when the cooling power is required at around −20 °C or even lower.

For comparison, we have used a simulation program with a different algorithm from DeltaE to study the system with the corresponding experimental parameters. The working frequency can always be accurately predicted. For the others, the calculations agree with the experimental results rather well within a certain pressure wave amplitude range and cold end temperature range. A deep look into parameter distributions inside the system shows that there is a good potential of further improving the system's performance.

Smaller slopes of the calculated curves of TWTR cooling power versus cold end temperature compared with the measured curves may indicate that some loss mechanisms strongly related to the TWTR's cold end temperature are seriously underestimated. One of the most possible reasons is the Rayleigh streamings inside the TWTR's thermal buffer tube since we have not taken any special measures to suppress them. Besides, the dimension of the thermal buffer tube and arrangement of the flow straightener should also be carefully checked. The flow resistance and heat transfer inside the regenerators may need to be more reasonably evaluated. On the other hand, study of stand-alone TWTE and TWTR are necessary for a thorough understanding of the sources of the difference between experiments and calculations, especially when highest pressure ratios exceed 1.1.

With the experimental results, we will further modify and improve the program (mainly to include the turbulence effects and minor losses) for a better match and use it for future optimizations.

TABLE IV. Comparison between calculations and experiments of pressure amplitudes at the TWTR inlet with different operating conditions.

No.	Average pressure (MPa)	Heating power (kW)	Cold end temperature (K)	Calculated P_2 amplitude (MPa)	Measured P_2 amplitude (MPa)	Relative difference (Cal-Meas)/Meas
1	3.0	2.2	208.8	0.228	0.184	23.9%
2	3.0	1.0	225.9	0.139	0.122	13.9%
3	2.52	1.5	211.5	0.166	0.142	16.9%
4	2.52	0.5	236.9	0.084	0.079	6.3%
5	2.02	1.0	219.6	0.115	0.094	22.3%
6	2.02	0.5	234.2	0.077	0.065	18.5%

ACKNOWLEDGMENT

This work is financially supported by the Natural Science Foundation of China with Grant No. 50536040.

[1]S. Backhaus and G. W. Swift, "A thermoacoustic Stirling heat engine," Nature (London) **399**, 335–338 (1999).
[2]G. W. Swift, D. L. Gardner, and S. Backhaus, "Acoustic recovery of lost power in pulse tube refrigerators," J. Acoust. Soc. Am. **105**, 711–724 (1999).
[3]Y. Huang, E. Luo, W. Dai, Y. Zhang, and Z. Wu, "A high performance traveling-wave thermoacoustic refrigerator operating in room-temperature range," Chin. Sci. Bull. **50**(2), 192–194 (2005) (in Chinese).
[4]W. Dai, E. Luo, J. Hu, and H. Ling, "A heat-driven thermoacoustic cooler capable of reaching liquid nitrogen temperature," Appl. Phys. Lett. **86**, 224103 (2005).
[5]W. Dai, E. Luo, Y. Zhou, J. Wu, and W. Zhu, "Building a high-efficiency and compact-sized thermoacoustically-driven pulse tube cooler," Cryocoolers **13**, 195–200 (2005).
[6]D. Gedeon, "DC gas flows in Stirling and pulse-tube cryocoolers," Cryocoolers **9**, 385–392 (1997).
[7]J. R. Olson and G. W. Swift, "Acoustic Streaming in pulse tube refrigerators: Tapered pulse tubes," Cryogenics **37**, 769–776 (1997).
[8]G. W. Swift, "Thermoacoustic engines," J. Acoust. Soc. Am. **84**, 1145-1180 (1998); *Thermoacoustics: A Unifying Perspective for Some Engines and Refrigerator* (Acoustical Society of America, Sewickley, PA, 2002).
[9]J. H. Xiao, "Thermoacoustic heat transportation and energy transformation Part I: Formulation of the problem," Cryogenics **35**, 15–19 (1995); J. H. Xiao, "Thermoacoustic heat transportation and energy transformation Part II: Isothermal wall thermoacoustic effects," *ibid.* **35**, 21–26 (1995); J. H. Xiao, "Thermoacoustic heat transportation and energy transformation Part III: Adiabatic wall thermoacoustic effects," *ibid.* **35**, 27–29 (1995).
[10]W. C. Ward and G. W. Swift, "Design environment for low amplitude thermoacoustic engines (DeltaE)," J. Acoust. Soc. Am. **95**, 3671–3672 (1994). Software and user's manual also available from the Los Alamos National Laboratory web site http://www.lanl.gov/thermoacoustics

Coupling mechanism of electromagnetic acoustical transducers for ultrasonic generation

X. Jian, S. Dixon, R. S. Edwards, and J. Reed
Department of Physics, Ultrasonic Group, University of Warwick, Coventry CV4 7AL, United Kingdom

(Received 16 March 2005; revised 14 February 2006; accepted 14 February 2006)

Electromagnetic acoustic transducers (EMATs) generate ultrasonic waves in metals through an electromagnetic coupling mechanism. A concept for EMAT generation, using a coil alone without a permanent magnet, but with a pulse generator and a sample, is introduced. A simplified equivalent coil circuit is given and has been validated by experimental measurements. Such an equivalent circuit can be used for variations in excitation current calculations, which have often been neglected in previous publications in this area but have proved to be of great importance in considering the efficiency and frequency characteristics of ultrasonic generation. The equivalent coil inductance is dependent on the distance between the coil and the sample, due to coil interactions with eddy currents and the Ampere current of the magnetization mechanism. Analytical solution for an annular coil above metal samples is given, and the influences of the lift-off, coil radius, material magnetic permeability and electrical conductivity on the equivalent coil inductance are discussed. Experimental measurements agree well with predictions. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2184288]

PACS number(s): 43.35.Yb, 43.38.Dv, 43.35.Rw, 43.38.Ar [AJZ] Pages: 2693–2701

I. INTRODUCTION

Electromagnetic acoustic transducers (EMATs) are one of several types of noncontact ultrasonic transducers.[1–16] They are noncontact, work at elevated temperatures, and are fairly insensitive to misalignment, and hence are suitable for measurements on-line and are less subject to coupling problems than traditional piezoelectric transducers (PTs).

Within a traditional PT there is an active piezoelectric element or array of elements. Its thickness and vibration mode determine the ultrasonic center frequency. The internal electrical elements (resistance, capacitance, and inductance) together with the backing layers determine the frequency bandwidth. If a pulse generator has bandwidth wider than that of the transducer, the ultrasonic waves generated by the transducer driven by the pulse generator are independent of the internal parameters of the pulse generator used.[6]

At present, the term EMAT is almost always taken to denote a coil with or without the presence of a permanent magnet. Ultrasonic waves are generated through the Lorentz force and/or magnetic mechanisms only when the coil is held close to a metal sample. The important point to bear in mind is that the frequency characteristics and generation efficiency of the ultrasound are dependent on both the sample and the pulse generator.[1–3,5] When a coil is designed, the pulse generator and sample to be measured must be taken into account in order to achieve the required frequency characteristics and optimize generation efficiency. It is well known that the lift-off, which is the distance between a coil and a sample, has an influence on eddy current and ultrasound generation,[1–3,7–16] however, the mechanism has not been addressed thoroughly and it is necessary to address this problem in detail.

The mechanism of ultrasonic generation by EMATs has been studied previously,[7–16] however, the influence on the excitation current through a coil from the rest of the system is typically ignored. The contribution of Ampere current of magnetization mechanism is ignored,[7] therefore the model cannot be used for ferromagnetic materials such as steel, iron, and ferrite. The contribution of eddy current is not fully considered, therefore it is only suitable for very low frequency. In previous studies using the finite element method (FEM), where attention has been focused on the optimal spatial design of the coil and the influence of lift-off,[13–16] the excitation current is usually assumed constant. However, the excitation current changes with lift-off or the load, not only in amplitude but also in frequency. Once the excitation current is obtained, we can calculate the eddy current, Lorentz force, and the ultrasonic generation analytically or numerically.[1–3,7–16]

In this paper, we first give an equivalent coil circuit and equivalent EMAT ultrasonic generation circuit, which can be used to calculate the excitation current, where equivalent inductance and equivalent capacitance are introduced. However, the equivalent inductance of a coil is dependent on the lift-off or load. The physical insight into the EMAT performance is given, and analytical expressions of equivalent inductance and the temporal function of Lorentz force for an annular coil are given that have taken both the eddy current and Ampere current into account. With the temporal function of Lorentz force, we can clearly see how excitation current is related to the frequency characteristics of the generated ultrasonic wave. In Sec. II, experimental measurements are reported which validate our equivalent circuit and predictions of the equivalent coil inductance with respect to frequency and lift-off. Section III gives the conclusions and a discussion of the results.

II. THEORY

EMAT ultrasonic generation and eddy current inductance have been investigated analytically for a few specific

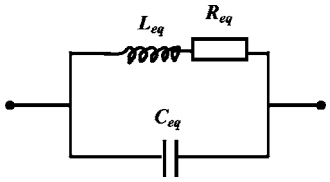


FIG. 1. Simplified equivalent coil circuit.

designs, such as annular coil, and numerically by finite element method (FEM), where the excitation current is assumed as continuous or transient. However, the excitation current changes with the load (close to electrical conductor) in both frequency characteristics and amplitude. EMAT generates ultrasonic waves in the electric conductor by means of Lorentz force due to the interaction of the induced eddy current with external static magnetic and dynamic magnetic field which is induced by the excitation current. The temporal function of the eddy current and dynamic magnetic field, as well as the Lorentz force, is dependent on that of the excitation current through the coil.

How the excitation current changes is of interest for us to investigate the characteristics of the generated ultrasonic waves. We should take the whole excitation system including the pulse generator, coil, and electrical conductor into account. First, we propose an equivalent circuit for the excitation system, based on which the excitation current can be calculated. An analytical expression for the equivalent coil inductance is given. The equivalent coil inductance is reduced by the eddy current induced in the electrical conductor, which tends to decrease the rate of change of the magnetic flux through the excitation coil and is increased for a ferrite where the relative magnetic permeability is a number much larger than 1, due to the magnetization apparently strengthening the magnetic flux through the excitation current so as to raise the rate of change of the magnetic flux through the coil. Finally, the temporal function of Lorentz force is given, which is the convolution of excitation current.

A. Coil equivalent circuit

Generally, an EMAT coil has many turns for optimal ultrasonic generation or receiving.[1–3] There is capacitance between the turns and between the coil and the metal sample, the value being dependent on coil configuration and lift-off. A coil also possesses a real resistive component to its impedance.

The impedance of a coil increases with the frequency of the excitation current flowing through the coil. The impedance is very small for direct current. The measured coil impedance has finite bandwidth around the frequency of the resonant impedance. When the bulk resistance is in series with the coil inductance, the quality factor of the coil circuit can be controlled by the bulk resistance. Therefore, the capacitance should be parallel to the coil inductance and the bulk resistance should be in series with the coil inductance. Such a circuit gives peak impedance around a particular frequency. A simplified equivalent circuit may be introduced to predict the behavior of a coil, as shown in Fig. 1, where L_{eq} is the equivalent inductance, R_{eq} is the internal bulk resistance of the coil, and C_{eq} is the equivalent capacitance representing all potential sources.

However, it is very difficult to give a circuit to simulate the behavior of a real coil exactly because the inductance process and capacitance process are not completely separate. The whole coil may be divided into many circuit elements connecting in series. Each circuit element is of similar structure as shown in Fig. 1. It would be very difficult to assess all the parameters and evaluate the behavior of a coil based on the parameters, so we choose the simple equivalent circuit to predict the behavior of a coil and have validated it by fitting experimental data to our predictions.

The equivalent impedance and equivalent phase of the equivalent coil circuit can be given by

$$Z_{\mathrm{eq}} = \frac{R_{\mathrm{eq}} + j((1-\omega^2 L_{\mathrm{eq}} C_{\mathrm{eq}})\omega L_{\mathrm{eq}} - \omega C_{\mathrm{eq}} R_{\mathrm{eq}}^2)}{(1-\omega^2 L_{\mathrm{eq}} C_{\mathrm{eq}}) - (\omega C_{\mathrm{eq}} R_{\mathrm{eq}})^2}, \tag{1}$$

$$\theta_{\mathrm{eq}} = \tan^{-1}\left(\frac{(1-\omega^2 L_{\mathrm{eq}} C_{\mathrm{eq}})\omega L_{\mathrm{eq}} - \omega C_{\mathrm{eq}} R_{\mathrm{eq}}^2}{R_{\mathrm{eq}}}\right). \tag{2}$$

We can measure the equivalent coil impedance and phase with an HP Vector Impedance Meter:

$$\omega_{\mathrm{reson}} = \frac{1}{\sqrt{L_{\mathrm{eq}} C_{\mathrm{eq}}}}. \tag{3}$$

When $\omega < \omega_{\mathrm{reson}}$, the phase angle is positive; when $\omega > \omega_{\mathrm{reson}}$, the phase angle is negative. From Eqs. (1) and (2), we can calculate the equivalent resistance (or in-phase voltage component) of the equivalent coil circuit, the real part of the equivalent impedance, Z_{eq}, by

$$\mathrm{real}(Z_{\mathrm{eq}}) = \frac{R_{\mathrm{eq}}}{(1-\omega^2 L_{\mathrm{eq}} C_{\mathrm{eq}}) - (\omega C_{\mathrm{eq}} R_{\mathrm{eq}})^2} \approx \frac{R_{\mathrm{eq}}}{1-\omega^2 L_{\mathrm{eq}} C_{\mathrm{eq}}}. \tag{4}$$

When $\omega < \omega_{\mathrm{reson}}$, and the equivalent inductance of the coil circuit decreases for some reason such as the coil approaching a metal sample, the real part of the equivalent coil impedance, real(Z_{eq}), decreases. But when $\omega > \omega_{\mathrm{reson}}$, a decrease in the equivalent inductance L_{eq} will make the equivalent resistance, real(Z_{eq}), increase. Note that the dc bulk resistance of the coil does not change and the ac bulk resistance of the coil increases with increasing frequency.

The discharge circuit on connecting a coil to a pulse generator is shown in Fig. 2, where C_0 and R_0 are the internal capacitance and electrical resistance of the pulse generator. Based on the discharge circuit, the excitation current through the coil can be calculated.[2,3] We first set

$$r = \sqrt{\frac{R^2 C_0}{4 L_{\mathrm{eq}}}}, \tag{5}$$

where the total resistance is given by $R = R_0 + R_{\mathrm{eq}}$. When $r = 1$ for a critically damped circuit, the current through the coil is given by[3]

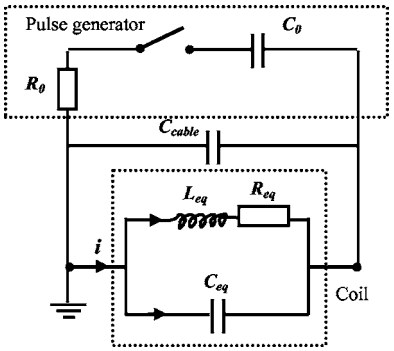


FIG. 2. Discharge circuit for a pulse EMAT.

$$i(t) \approx -I_0(r\omega_0 t)e^{-r\omega_0 t}, \tag{6}$$

where $I_0=\omega_0 C_0 U_{\max}$, $U_{\max}$ is the maximum voltage to which the capacitor C_0 is charged, and $\omega_0=1/\sqrt{L_{\text{eq}}C_0}$ is the natural resonant angular frequency of the circuit. When $r<1$ for the underdamped case, the current is given by[3]

$$i(t) = -\omega_0 C_0 U_{\max}\left(\frac{\sin(\omega_0 t\sqrt{1-r^2})}{\sqrt{1-r^2}}\right)e^{-r\omega_0 t}. \tag{7}$$

When $r>1$ for the overdamped case, the current is given by[3]

$$i(t) = -\omega_0 C_0 U_{\max}\left(\frac{e^{\omega_0 t\sqrt{r^2-1}} - e^{-\omega_0 t\sqrt{r^2-1}}}{2\sqrt{r^2-1}}\right)e^{-r\omega_0 t}. \tag{8}$$

B. Coil inductance

When an excitation current flows through a transmitting coil above an electrical conductor, an eddy current is induced within the skin depth. The eddy current is mainly in the opposite direction to the excitation current and reduces the rate of change of the magnetic flux through the excitation coil. Surface and volume Ampere currents are also to be generated due to the magnetization mechanism. The direction of the Ampere current is dependent on the magnetic characteristics of the sample and the external magnetic field. For paramagnetic and diamagnetic materials, the Ampere current is very small, and therefore the contribution to the magnetic flux through the excitation coil can be ignored. For ferromagnetic materials, the Ampere current is strong and is in the same direction as the excitation current provided that the frequency of the current is not so high that the magnetic domains cannot move at the same rate. The Ampere current increases the rate of change of the magnetic flux through the excitation coil and thus increases the inductance of the coil.

The total rate change of the magnetic flux through the excitation coil is expressed by

$$\frac{d\Phi}{dt} = \frac{d\Phi_{\text{air}}}{dt} - \frac{d\Phi_{\text{eddy}}}{dt} + \frac{d\Phi_{\text{Ampere}}}{dt}, \tag{9}$$

where

$$\frac{d\Phi}{dt} \equiv L_{\text{eq}}\frac{di}{dt}, \tag{10}$$

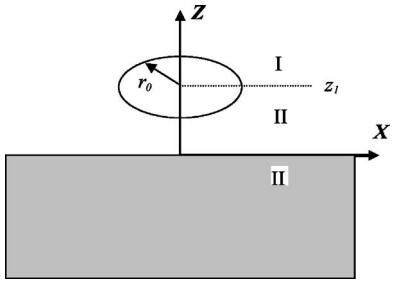


FIG. 3. A coordinate system for the analysis of a coil coupling mechanism. An annular coil is parallel to an electrical conductor with lift-off z_1. Partial differential equations at three regions are solved with the consideration of boundary conditions between each two connecting regions. The annular coil is of radius r_0 with its axis overlapping with the z axis.

$$\frac{d\Phi_{\text{air}}}{dt} \equiv L_{\text{air}}\frac{di}{dt}, \tag{11}$$

$$\frac{d\Phi_{\text{eddy}}}{dt} \equiv L_{\text{eddy}}\frac{di}{dt}, \tag{12}$$

$$\frac{d\Phi_{\text{Ampere}}}{dt} \equiv L_{\text{Ampere}}\frac{di}{dt}. \tag{13}$$

Equation (9) can be rewritten as

$$L_{\text{eq}} = L_{\text{air}} - L_{\text{eddy}} + L_{\text{Ampere}}. \tag{14}$$

L_{eq} is the equivalent inductance of the coil. L_{air}, L_{eddy}, and L_{Ampere} are the coil inductance in the absence of metal, the inductance contribution from eddy currents, and the inductance contribution from the Ampere current of magnetization, respectively. From Eq. (14), an induced eddy current tends to reduce the equivalent coil inductance, while the Ampere current tends to increase the equivalent coil inductance. It is noted that the excitation current, the induced eddy current, and the Ampere current are related to each other by the governing Maxwell's equations. The analytical expression for the equivalent coil inductance can be deduced only for a very few cases. We consider an annular coil above an electrical conductor as shown in Fig. 3, with a sinusoidal current of $\mathbf{i}(t)=\mathbf{i}_0 e^{j\omega t}$ flowing in the coil.

The differential equation for the magnetic vector potential, $\boldsymbol{A}$, in an isotropic, linear, and homogeneous medium due to an applied current is given by[2,3,7–16]

$$\nabla^2\boldsymbol{A} = -\mu\boldsymbol{i} + \mu\sigma\frac{\partial\boldsymbol{A}}{\partial t} + \mu\varepsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2}. \tag{15}$$

Equation (15) is rewritten into the polar coordinate system,

$$\frac{\partial^2\boldsymbol{A}}{\partial r^2} + \frac{1}{r}\frac{\partial\boldsymbol{A}}{\partial r} + \frac{\partial^2\boldsymbol{A}}{\partial z^2} - \frac{\boldsymbol{A}}{r^2} = -\mu\boldsymbol{i} + \mu\sigma\frac{\partial\boldsymbol{A}}{\partial t} + \mu\epsilon\frac{\partial^2\boldsymbol{A}}{\partial t^2}. \tag{16}$$

With axial symmetry, there is only a tangential component of current $i(t)$ and therefore of vector potential $\boldsymbol{A}$. Vector potential $\boldsymbol{A}$ in Eqs. (15) and (16) will be replaced with the tangential component, A.

The vector potentials in the areas I, II, and III as shown in Fig. 3 can be obtained by solving the partial differential

equations with the boundary conditions, where the N turns of an annular coil are assumed to occupy same space.[7] Canceling out the term $e^{j\omega t}$, the vector potentials are given by

$$A_I(r,z,\omega) = \frac{\mu_1 r_0 N i_0}{2} \int_0^\infty J_1(\alpha_1 r_0) J_1(\alpha_1 r) \times \left[e^{\alpha_1 z_1} - \frac{\alpha_r - \mu_r}{\alpha_r + \mu_r} e^{-\alpha_1 z_1} \right] \frac{\alpha}{\alpha_1} e^{-\alpha_1 z} \, d\alpha, \quad (17)$$

$$A_{II}(r,z,\omega) = \frac{\mu_1 r_0 N i_0}{2} \int_0^\infty J_1(\alpha r_0) J_1(\alpha r) e^{-\alpha_1 z_1} \frac{\alpha}{\alpha_1} \times \left[e^{+\alpha_1 z} - \frac{\alpha_r - \mu_r}{\alpha_r + \mu_r} e^{-\alpha_1 z} \right] d\alpha, \quad (18)$$

$$A_{III}(r,z,\omega) = \mu_1 r_0 N i_0 \int_0^\infty J_1(\alpha r_0) J_1(\alpha r) \frac{\mu_r}{\alpha_r + \mu_r} \frac{\alpha}{\alpha_1} \times e^{\alpha_2 z - \alpha_1 z_1} \, d\alpha, \quad (19)$$

where

$$\alpha_1 = \sqrt{\alpha^2 - \omega^2 \mu_1 \varepsilon_1}, \quad (20)$$

$$\alpha_2 = \sqrt{\alpha^2 + j\omega \mu_2 \sigma_2}, \quad (21)$$

$$\mu_r = \frac{\mu_2}{\mu_1}, \quad (22)$$

$$\alpha_r = \frac{\alpha_2}{\alpha_1}, \quad (23)$$

where μ_1 and ε_1 are the magnetic permeability and dielectric permittivity in the air, and μ_2 and σ_2 are the permeability and conductivity of the conductor. The voltage over a single loop of wire of radius r_0 at $z=z_1$ due to magnetic coupling (the voltage due to current over bulk resistance is ignored) is given by

$$V(\omega) = \oint_{r_0} \mathbf{E} \, d\boldsymbol{l} = \oint_{r_0} -\frac{\partial \boldsymbol{A}_{II}}{\partial t} \, d\boldsymbol{l} = -j\omega 2\pi r_0 A_{II}(r_0, z_1). \quad (24)$$

So the total voltage over the coil of N turns is

$$V(\omega) = -jN\omega 2\pi r_0 A_{II}(r_0, z_1). \quad (25)$$

The impedance of a coil is defined by

$$Z(\omega) \equiv \frac{V(\omega)}{i_0}. \quad (26)$$

The equivalent inductance of a coil is defined by

$$L_{\text{eq}}(\omega) \equiv \frac{Z(\omega)}{-j\omega} = \frac{V(\omega)}{-j\omega i_0}. \quad (27)$$

Substituting Eq. (24) into Eq. (27), the inductance of a coil is obtained by

$$L_{\text{eq}}(\omega) = \pi \mu_1 r_0^2 N^2 \int_0^\infty J_1(\alpha r_0)^2 \frac{\alpha}{\alpha_1} \left[1 - \frac{\alpha_r - \mu_r}{\alpha_r + \mu_r} e^{-2\alpha_1 z_1} \right] d\alpha \quad (28)$$

When z_1/r_0 is very large, Eq. (28) can be rewritten as

$$L_{\text{eq}}(\omega) \approx L_{\text{air}}(\omega) = \pi \mu_1 r_0^2 N^2 \int_0^\infty J_1(\alpha r_0)^2 \frac{\alpha}{\alpha_1} \, d\alpha, \quad (29)$$

where $L_{\text{air}}(\omega)$ is the self-inductance of a coil in air, which is proportional to the radius of the coil and the square of the coil turn number, but independent of frequency, and the impedance of the coil increases linearly with angular frequency.

$$L_{\text{eq}}(\omega) = L_{\text{air}}(\omega) - (L_{\text{eddy}}(\omega) - L_{\text{Ampere}}(\omega)), \quad (30)$$

$$L_{\text{eddy}}(\omega) - L_{\text{Ampere}}(\omega) = \pi \mu_1 r_0^2 N^2 \int_0^\infty J_1(\alpha r_0)^2 \frac{\alpha_r - \mu_r}{\alpha_r + \mu_r} \frac{\alpha}{\alpha_1} e^{-2\alpha_1 z_1} \, d\alpha. \quad (31)$$

When z_1/r_0 is close to zero, Eq. (28) can be rewritten as

$$L_{\text{eq}}(\omega) \approx L_0(\omega) = \pi \mu_1 r_0^2 N^2 \int_0^\infty J_1(\alpha r_0)^2 \frac{2\mu_r}{\alpha_r + \mu_r} \frac{\alpha}{\alpha_1} \, d\alpha. \quad (32)$$

$L_0(\omega)$ denotes the equivalent inductance of a coil when the coil contacts the metal directly. In case that $\mu_r \gg \alpha_r$ for a ferrite material, Eq. (32) is rewritten as

$$L_0(\omega) \approx 2L_{\text{air}}(\omega). \quad (33)$$

The increase of the coil inductance is due to the strengthening of the magnetic flux through the coil. In the case that $\mu_r \approx 1$ for a nonferromagnetic material, such as aluminum and copper, Eq. (32) is rewritten as

$$L_0(\omega) = \pi \mu_1 r_0^2 N^2 \int_0^\infty J_1(\alpha r_0)^2 \frac{2}{\alpha_r + 1} \frac{\alpha}{\alpha_1} \, d\alpha. \quad (34)$$

The decrease of the coil inductance is due to the contribution of the induced eddy current in the electrical conductor, which reduces the rate of change of the magnetic flux through the coil.

Taking the derivative of the coil inductance in Eq. (28) over the lift-off z_1, we get

$$\frac{\partial L_{\text{eq}}(\omega)}{\partial z_1} = 2\pi \mu_1 r_0^2 N^2 \int_0^\infty J_1(\alpha r_0)^2 \frac{\alpha_r - \mu_r}{\alpha_r + \mu_r} e^{-2\alpha_1 z_1} \alpha \, d\alpha. \quad (35)$$

It is found that the ratio of the lift-off to the coil radius, z_1/r_0, determines the coil inductance sensitivity to the lift-off.

C. Temporal function of Lorentz force

Approximately, the Lorentz force can be expressed by[2,3]

$$\boldsymbol{F}(r,z,t) = \boldsymbol{F}_\delta(r,z) F_t(t), \quad (36)$$

where $\boldsymbol{F}_\delta(r,z)$ and $\boldsymbol{F}_t(t)$ are the spatial distribution function and temporal function of the Lorentz force, respectively. An

ultrasonic wave generated by a Lorentz force can be expressed by[2,3]

$$\boldsymbol{u}(r,z,t) = \boldsymbol{u}_{\delta}(r,z,t) \otimes F_t(t), \tag{37}$$

where $\boldsymbol{u}_{\delta}(r,z,t)$ is the ultrasonic wave generated by a Lorentz force of the spatial distribution described by $\boldsymbol{F}_{\delta}(r,z)$ of the delta temporal function.

The frequency characteristics of the generated ultrasonic wave are dependent on the temporal function and the spatial distribution functions of the corresponding Lorentz forces. From Maxwell's equation, the eddy current density in the electrical conductor directly under the excitation coil is given by

$$J_{\text{eddy}}(r_0,z,\omega) = -\sigma_2 \frac{\partial A_{III}(r_0,z,\omega)}{\partial t}. \tag{38}$$

Considering an excitation current pulse of temporal function $i(t)$, its frequency spectrum is described by

$$I(\omega) = \text{FFT}(i(t)), \tag{39}$$

where FFT is Fast Fourier Transform operator. Replacing i_0 with $I(\omega)$ in Eqs. (17)–(19) and (38), the eddy current density due to the current pulse is obtained by performing inverse FFT operation,

$$J_{\text{eddy}}(r_0,z,t) = \text{FFT}^{-1}(J_{\text{eddy}}(r_0,z,\omega)). \tag{40}$$

The magnetic flux density directly under the excitation current has mainly radial components given by

$$B_d(r_0,z,\omega) = \frac{\partial A_{III}(r_0,z,\omega)}{\partial z}. \tag{41}$$

The dynamic magnetic flux densities in time domain can be given by performing an inverse FFT operation,

$$B_d(r_0,z,t) = i(t) \otimes B_{d,\delta}(r_0,z,t), \tag{42}$$

where

$$B_{d,\delta}(r_0,z,t) = \text{FFT}^{-1}\left(r_0 N \int_0^{\infty} J_1(\alpha r_0)^2 \frac{\alpha \alpha_r \mu_2}{\alpha_r + \mu_r} e^{\alpha_2 z - \alpha_1 z_1}\, d\alpha \right). \tag{43}$$

The Lorentz force due to the eddy current interacting with the dynamic magnetic field induced has mainly component in negative Z-axis direction and is given by

$$F_d(r_0,z,t) = J_{\text{eddy}}(r_0,z,t) B_d(r_0,z,t). \tag{44}$$

Similarly, the Lorentz force due to the eddy current interacting with an external static magnetic field is dependent on the static magnetic field and is given by

$$\boldsymbol{F}_s(r_0,z,t) = \boldsymbol{J}_{\text{eddy}}(r_0,z,t) \times \boldsymbol{B}_s. \tag{45}$$

When eddy current $J_{\text{eddy}}(r_0,z,t)$ and dynamic magnetic field $B_d(r_0,z,t)$ have similar frequency characteristics, an enhanced Lorentz force can be obtained to generate ultrasonic waves more efficiently.[2,3] The coil is optimized for ultrasonic wave generation with specific frequency characteristics.

TABLE I. Sizes of annular coils used in this paper.

Coil label	AN1	AN2	AN3
Size (mm)[a]	11	11	11
Turns	3	6	9

[a]Internal diameter of coils.

III. EXPERIMENTAL MEASUREMENTS

Measurements were done using annular coils. All of the coils were made from copper wire of diameter 0.22 mm. The structures of the coils are described in Table I. Aluminum, copper, and steel samples were used. Aluminum is paramagnetic and copper is diamagnetic, and both have magnetic susceptibility χ_m of the order of 10^{-5}. The steel sample contains 0.5% chromium, molybdenum, and vanadium in total, and has been treated at temperature 980 °C for 1 h. The steel is ferromagnetic and of high magnetic susceptibility χ_m. The electrical conductivities and relative magnetic permeability of the materials are shown in Table II and will be used in later calculations. Please note that the relative permeability of steel and ferrite given in Table II are for direct current. The relative permeability of an alternating current is dependent on the frequency, but such information is not available in the literature. In this paper, the relative permeability of steel for frequencies above 30 kHz is given by $\mu_r = \mu_{r0} e^{-(f-30\,000)/800}$, and the relative permeability of ferrite for frequencies above 30 kHz is given by $\mu_r = \mu_{r0} e^{-(f-30\,000)/300}$, in order to achieve best agreement between the calculation and the measurement, where μ_{r0} is the relative permeability of dc given in Table II.

The experimental measurement setup is shown in Fig. 4. The pulse generator used in this work has an internal resistance of 2Ω and a maximum charge voltage of 850 V on the capacitor (1 μF). The capacitances of two cables of length 500 and 1150 mm used to connect a coil and the pulse generator are 12 and 53 pF, respectively. For an EMAT working at a frequency ranging from 200 kHz to 1 MHz, the influence of the cable capacitances can be neglected.

In order to accurately control the lift-off, a micrometer is used. The coil is attached to a nonmetallic block in parallel and the block is put on a platform controlled using a micrometer. The coil together with the fixed nonmetallic block moves with the micrometer platform. In this paper, coil impedance is measured with an HP Vector Impedance Meter for frequency ranging from 0.4 to 110 MHz, and by an Impedance Card installed in a computer for frequency ranging from 1 kHz to 1 MHz. The excitation current through the coil is

TABLE II. Material properties.

	Relative permeability[a]	Electrical conductivity
Aluminum	1.000 023	37 710 000
Copper	0.999 902	59 610 000
Steel (annealed)	10	6 831 000
Ferrite	200	50 000

[a]At 20° for dc.

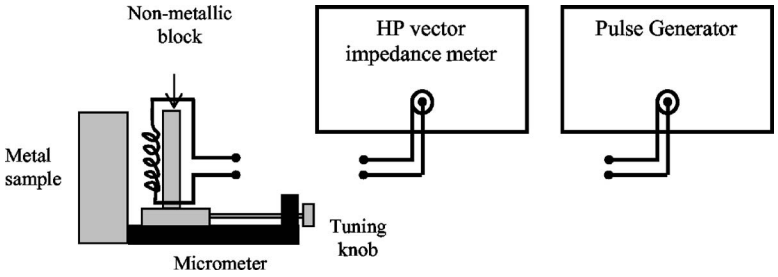


FIG. 4. Measurement setup for impedance and excitation current measurements.

measured using a LeCroy oscilloscope with a digitalization rate of 100 MHz and an analog bandwidth of 400 MHz.

A. Validation of equivalent coil circuit

We now consider the capacitance in the simplified equivalent coil circuit shown in Fig. 1. The resonant frequency of such an LCR circuit is given by[3]

$$f_R \approx \frac{1}{2\pi\sqrt{L_{\mathrm{eq}}C_{\mathrm{eq}}}}. \tag{46}$$

At the resonant frequency, the input current approaches zero. The coil impedance, which is inversely proportional to the current, achieves a maximum around frequency f_R.

Figure 5 shows the measured and calculated impedance for the annular coil labeled AN3 based on the simplified coil equivalent LCR circuit and the LR circuit for frequencies ranging from 400 kHz to 100 MHz for annular and linear coils. We choose values for capacitance, resistance, and inductance for the coil equivalent circuit such that the difference between the measured and simulated impedance with respect to frequency is close to the minimum. The simulation based on the simplified coil equivalent circuit agrees well with the measurements, while the calculation based on an LR circuit does not agree with the measurements. This confirms the existence of a significant capacitance within the coil. The effective inductance, capacitance, and resistance of coil AN3 in the presence of air, aluminum, copper, and steel are calculated based on the equivalent circuit and are shown in Table III.

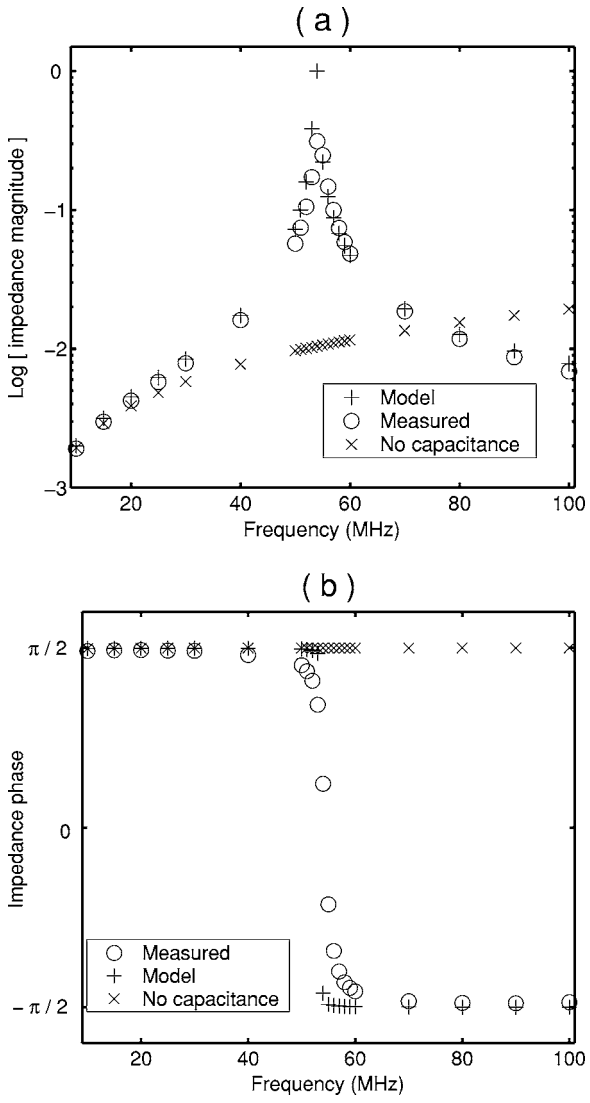


FIG. 5. Magnitude (a, plotted on a logarithm scale) and phase (b, plotted on a linear scale) of the impedance of coil AN3 with respect to frequency. The impedance magnitude in (a) is divided by the maximum magnitude of the calculated impedance peak. Circle is measured, plus sign is calculated based on the model of this paper, and cross is calculated where capacitance is not considered.

B. Eddy current and magnetization

Aluminum is paramagnetic while copper is diamagnetic, and both have a very small magnetic susceptibility χ_m. Therefore the magnetic field generated by magnetization can be ignored for these materials. All coils described in Table I were used with these samples, and the impedance was taken for each coil in three situations: without a sample present (surrounded by air), contacting the aluminum block, and contacting the copper sample. No peak impedance is found in the frequency range 0–100 MHz for the presence of both aluminum and copper because the equivalent inductance is too small. The maximum impedance is predicted to be around 100–110 MHz. The peak maximum impedance and phase reversal frequency appear at a higher frequency than

TABLE III. Effective inductance, capacitance, and resistance of annular coils with the presence of varied samples.

Coil label	Sample	L_{eq} (μH)	R_c (Ω)	Z_χ(Ω)	C_{eq} (pF)	f_{R} (MHz)
AN3	Air	1.38	0.6	0	6.36	53.6
	Aluminum	0.356	0.6	0	7.22	99.2
	Copper	0.355	0.6	0	6.90	102
	Steel	0.62	0.6	2.5	5.1	89.5

[a]Z_χ (Ω) denotes the impedance at frequency 500 kHz introduced in the equivalent circuit due to magnetization.

TABLE IV. Effective equivalent coil inductance (μH).

Coil label	AN1	AN2	AN3
Air	0.26	0.72	1.38
Aluminum	0.12	0.32	0.36
Steel	0.20	0.53	0.62

that in air (over 100 MHz, as opposed to 53 MHz). Table IV shows the equivalent coil inductances with and without steel and copper samples. These results prove that the eddy current tends to decrease the equivalent coil inductance.

The steel sample used in these measurements is ferromagnetic. Based on the simplified equivalent coil circuit, the equivalent inductance and capacitance of coil AN3 are calculated and are given in Table III. Through comparison with those of air and copper samples, it is found that the equivalent inductance of steel is smaller than that of air but bigger than that of copper. Same conclusion can be drawn from the results of coils AN1 and AN2 shown in Table IV.

Figure 6 shows the measured and calculated impedance of coil AN3 with the presence of air with no nearby magnetic or electrically conducting materials and in close proximity to ferrite, steel, and copper samples. The impedance measured with the ferrite present, where only the Ampere current of magnetization contributes to increase the coil inductance, is always higher than that of air. The impedance in the presence of copper, where only the eddy current contributes to reduce the coil inductance, is always lower than that of air. For steel, both the eddy current and Ampere current of magnetization contribute to coil inductance. The influence of eddy current increases faster with the frequency of the excitation current than the contribution from the Ampere current. At low frequencies, the coil inductance in the presence of steel is higher than that in air. When the frequency of the excitation current is above about 350 kHz, the coil inductance in the presence of steel becomes lower than that in air. This is related to the mobility of magnetic domains in the material.

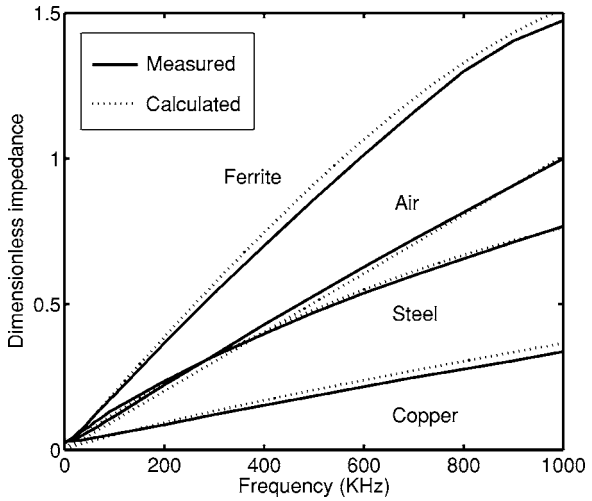


FIG. 6. Measured (solid lines) and calculated (dotted lines) impedances of the coil AN3 with respect to frequency at low frequency. For better comparison, the measured and calculated impedances are divided by the measured impedance of the coil at a frequency of 1 MHz in the presence of air.

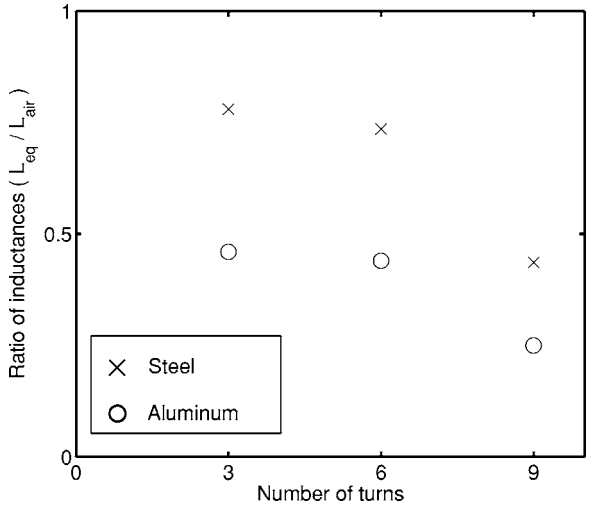


FIG. 7. Ratio of the measured inductances of coils AN1 (three turns), AN2 (six turns), and AN3 (nine turns) in the presence of aluminum (circle) or steel (cross) to that in the absence of a metal. Here L_{eq} or L_{air} denotes the impedances of a coil in the presence of a metal or air.

C. Lift-off sensitivity

Figure 7 shows the experimentally measured ratio of the equivalent coil inductance with aluminum and steel samples to that without any nearby sample (air), $L_{eq\text{-}metal}/L_{eq\text{-}air}$, for annular coils at a frequency of 500 kHz. The equivalent coil inductance with the steel sample is higher than that with the aluminum sample. The coil inductance of coil AN3 of nine turns is more influenced by the eddy current or eddy current and Ampere current than coil AN1 of three turns.

From Eqs. (28), (29), and (32), the ratio of coil inductance in the presence of metal to that in the absence of metal is not influenced by the number of coil turns. Figure 8 shows the ratio of the coil inductance in the presence of metal to that in the absence of metal calculated based on Eqs. (32) and (29), respectively, at frequency 300 kHz. The coil induc-

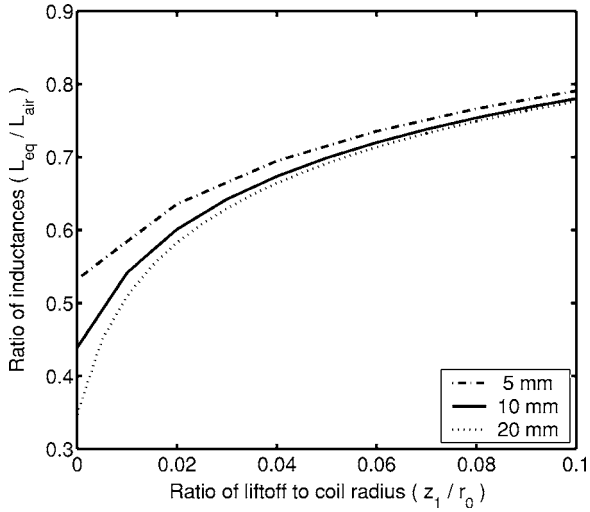


FIG. 8. Ratio of calculated inductance of annular coils in the presence of aluminum to that in the absence of aluminum with respect to the ratio of the lift-off to the coil radius. Dotted line, solid line, and dashed-dotted line are for coil radius of 20, 10, and 5 mm, respectively. Here L_{eq} or L_{air} denotes the impedances of a coil in the presence of a metal or air; z_1 and r_0 are the lift-off and the radius of the annular coil, respectively.

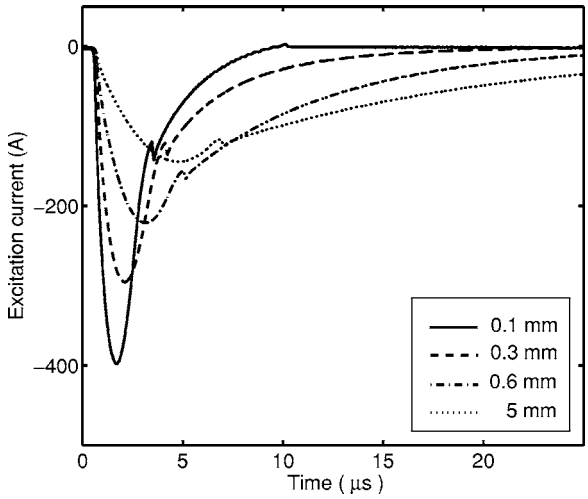


FIG. 9. Measured transient currents through an excitation coil at various lift-offs of 0.1 mm (solid line), 0.3 mm (dashed line), 0.6 mm (dashed-dotted line), and 5 mm (dotted line).

tance changes with respect to the ratio of lift-off to coil radius. For practical measurements, shown in Fig. 7, the lift-off may be assumed to be the same for coils AN1, AN2, and AN3. The internal radii of the three coils are the same, but their average radii are different. A coil of more turns has a larger overall width. These two factors explain the measured differences of the coils of varied number of turns.

With equivalent coil inductance, we can calculate coil excitation current based on the equivalent circuit as shown in Fig. 2. We give measured results in Fig. 9 to demonstrate the change of excitation current in amplitude and shape with respect to lift-off, where the excitation currents flowing in coil AN3 at lift-offs of 0.1, 0.3, 0.6, and 5 mm are measured. Rayleigh waves were observed with a reasonably high ratio of signal to noise at all of these current levels except that of 140 A at lift-off 5 mm.[2,3] The amplitude of the excitation current decreases dramatically with respect to lift-off, and the temporal width of the current pulse increases with lift-off, showing that the frequency bandwidth of the generation source drops significantly. Lift-off changes the frequency bandwidth of the excitation current, and, accordingly, the temporal function of the Lorentz force, which determines the frequency characteristics of the generated ultrasonic waves.[3]

IV. CONCLUSIONS AND DISCUSSION

In this paper, we have demonstrated why an EMAT must be considered from the point of view of a whole system, including a pulse generator, a transmitting coil, and a metal sample. The advantage of this concept is that the ultrasonic frequency characteristics and generation efficiency are considered systematically.

A coil and a metal sample are coupled by means of an electromagnetic coupling mechanism. An equivalent coil circuit is given, which includes the internal electrical resistance, self-inductance, and capacitance. The former two factors are well known. The capacitance between the coil and the metal sample and the capacitance of interwinding has a significant effect on the EMAT behavior. The influence of the coil capacitance becomes more significant for high-frequency wideband ultrasonic generation up to 30 MHz.[2,3] Due to the eddy current and magnetization mechanisms, the equivalent inductance is dependent on the magnetic susceptibility, the electrical conductivity of the metal sample, and the configuration of the coil. When a transmitting coil approaches a metal sample, a dynamic magnetic field and an eddy current are induced in the metal. Transient eddy currents produce dynamic magnetic fluxes passing through the transmitting coil, which oppose the change caused by the transient excitation current flowing in the coil. This results in the reduction of the equivalent coil inductance, making the excitation current more temporally sharp and enhancing the bandwidth. Magnetization mechanisms tend to increase or decrease the magnetic flux depending on the magnetic susceptibility of the metal sample, also resulting in an increase or decrease in the equivalent inductance. This is obvious for ferromagnetic materials.

An analytical expression of for equivalent coil inductance of an annular coil is given, which states the relative contribution from the eddy current and Ampere current on the equivalent inductance. The calculated impedances of an annular coil in air away from any nearby magnetic or electrically conducting samples, and then in close proximity to steel, ferrite, and copper, agree well with the measurements. The temporal function of Lorentz force is also given, which shows the relationship between the excitation current and can be used to predict the frequency characteristics of the ultrasonic wave that is generated by the annular coil. Experimental and calculated data can be found in our two references (Refs. 2 and 3) demonstrating ultrasonic waves of reasonably high ratio of signal to noise.

ACKNOWLEDGMENTS

This work is supported partly by the EPSRC funded UK Research Centre for Nondestructive Evaluation (RCNDE) and partly by Corus. We thank the NMR group of the Physics Department, Warwick University for allowing us to use the HP Vector Impedance Meter, especially Dr. Andrew Howes for his useful advice. Our thanks are due to Adrian Lovejoy of the electronic workshop of the department for his advice and help with measurements. We also thank Dr. Mark Potter for his invaluable comments on the paper.

[1]X. Jian, S. Dixon, and S. B. Palmer, "In-plane and out-of-plane particle velocity measurement using electromagnetic acoustical transducers," *Proceeding of the IEEE Ultrasonic Symposium 2005*, pp. 1276–1279.
[2]X. Jian, S. Dixon, and R. S. Edwards, "Modelling ultrasonic generation for Lorentz force EMATs," Insight **46**(11), 671–673 (2004).
[3]X. Jian, S. Dixon, and R. S. Edwards, "Optimising ultrasonic wideband Rayleigh wave generation by pulsed electromagnetic coils," Nondestr. Test. Eval. **20**, 42–62 (2005).
[4]R. J. Dewhurst, D. A. Hutchins, S. B. Palmer, and C. B. Scruby, "Quantitative measurements of laser-generated acoustic waveforms," J. Appl. Phys. **53**, 4064–4071 (1982).
[5]X. Jian, S. Dixon, and R. S. Edwards, "Ultrasonic generation and optimization for EMAT," Rev. Prog. Quant. Nondestr. Eval. **24A**, 1041–1046 (2005).
[6]X. Jian, J. P. Weight, and K. T. V. Grattan, "Miniature wideband ultrasonic transducers to measure compression and shear waves in solid," Sens. Actuators, A **127**, 13–23 (2006).
[7]C. V. Dodd and W. E. Deeds, "Analytical solutions to eddy-current probe-coil problems," J. Appl. Phys. **39**, 2829–2838 (1968).
[8]E. R. Dobbs, "Electromagnetic generation of ultrasonic waves," in *Physi-*

cal Acoustics Vol. 10, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1973).
[9]R. B. Thompson, "Mechanisms of electromagnetic generation and detection of ultrasonic Lamb waves in iron-nickel alloy polycrystals," J. Appl. Phys. **48**, 4942–4950 (1977).
[10]R. B. Thompson, "Physical principles of measurements with EMAT transducers," Phys. Acoust. **19**, 157–200 (1990).
[11]K. Kawashima, "Theory and numerical calculation of the acoustic field produced in metal by an electromagnetic ultrasonic transducer," J. Acoust. Soc. Am. **60**, 1089–1099 (1976).
[12]K. Kawashima, "Quantitative calculation and measurement of longitudinal and transverse ultrasonic wave pulses in solid," IEEE Trans. Sonics Ultrason. **31**, 83–93 (1984).
[13]H. Ogi, M. Hirao, and T. Ohtani, "Line-focusing electromagnetic acoustic transducers for the detection of slit defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 341–346 (1999).
[14]R. Ludwig and X. W. Dai, "Numerical simulation of electromagnetic acoustic transducer in the time domain," J. Appl. Phys. **69**, 89–98 (1991).
[15]R. Ludwig and X. W. Dai, "Numerical and analytical modeling of pulsed eddy currents in a conducting half-space," IEEE Trans. Magn. **26**, 299–307 (1990).
[16]R. Ludwig, Z. You, and R. Palanisamy, "Numerical simulation of an electromagnetic acoustic transducer-receiver system for NDT applications," IEEE Trans. Magn. **29**, 2081–2089 (1993).

An active headrest for personal audio

Stephen J. Elliott and Matthew Jones[a)]
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

(Received 21 September 2005; revised 7 February 2006; accepted 20 February 2006)

There is an increasing need for personal audio systems, which generate sounds that are clearly audible to one listener but are not audible to other listeners nearby. Of particular interest in this paper are listeners sitting in adjacent seats in aircraft or land vehicles. Although personal audio could then be achieved with headsets, it would be safer and more comfortable if loudspeakers in the seat headrests could be actively controlled to generate an acceptable level of acoustic isolation. In this paper a number of approaches to this problem are investigated, but the most successful involves a pair of loudspeakers on one side of the headrest, driven together to reproduce an audio signal for a listener in that seat and also to attenuate the pressures in the adjacent seat. The performance of this technique is investigated using simple analytic models and also with a practical implementation, tested in an anechoic chamber and a small room. It is found that significant attenuations, of between 5 and 25 dB, can be achieved in the crosstalk between the seats for frequencies up to about 2 kHz. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188814]

PACS number(s): 43.38.Hz, 43.38.Md, 43.38.Ja, 43.50.Ki, 43.60.Fg [AJZ] Pages: 2702–2709

I. INTRODUCTION

When two people are listening to different audio channels in adjacent seats in an aircraft or road vehicle, each person wants to hear their own channel faithfully reproduced, with a minimum of interference from the other person's channel. Figure 1 shows a diagram of the problem. Although this could be achieved if each person wore a headset, there may be comfort or safety issues with such devices and it would be more convenient if the audio channels could be reproduced through loudspeakers in the headrests of the seats. The safety implications of using headsets instead of headrest loudspeakers are concerned with the wiring to the headset and the fact that headsets can restrict communication with the wearer.

This paper investigates various approaches to actively improving the acoustic isolation between adjacent seats with headrest loudspeakers, so that the two people experience personal audio. Two strategies are investigated, initially using a simple free-field acoustic model. The first configuration is where the loudspeakers in one seat are used to control the sound from the loudspeaker in an adjacent seat. Such an arrangement has already been investigated using a finite element method by Hashimoto *et al.*[1] The second method, in general, employs a closely spaced array of acoustic sources to maintain a large pressure in one region of its near field while minimizing the pressure in another region. In this initial investigation, however, the loudspeaker array consists of just a pair of sources located in one side of the headrest in which the levels are maintained. Arrays of loudspeakers are commonly used to generate desired far-field directivity patterns using various beamforming techniques[2,3] and such a far-field array has also been investigated with a similar aim to the current paper by Druyvesteyn and Garas,[4] who emphasized that conventional active control systems could only be effectively used at low frequencies.

Another possible technology for the generation of personal audio may be the "audio spotlight"[5] in which the nonlinear interaction between two ultrasonic signals is used to generate a highly directional audio source in a similar way to an underwater parametric array.[6] These devices may, however, have unacceptable distortion properties for audio applications, because of their nonlinear generation mechanism, and there are questions about the safety of the high levels of ultrasound of such devices,[7] particularly in the near field if they were used in headrests.

Both of the control strategies using loudspeakers in the headrests could be implemented using feedforward control techniques,[8] since the audio channel provides a reference signal for a feedforward filter. An initial experimental investigation of the system is reported which uses an array of two closely spaced loudspeakers in the headrest of one of a pair of aircraft seats. Measurements of the frequency response from the loudspeakers to an array of microphones were first taken in an anechoic chamber and used to predict the performance in this experimental arrangement. A real-time implementation is then described, together with some measurements in a small room, which confirm that results very similar to those predicted from the anechoic measurements can be achieved in practice. Preliminary versions of this work were reported in Refs. 9–12.

II. SIMULATION STUDY

The simplest method of controlling the sound from a loudspeaker in one seat, in an adjacent seat, would be to use the loudspeakers already present in the seat as secondary sources. This is illustrated in Fig. 2 in which s is the audio signal used to drive the loudspeaker in the right-hand seat,

a)Electronic mail: mj@isvr.soton.ac.uk

 0001-4966/2006/119(5)/2702/8/$22.50

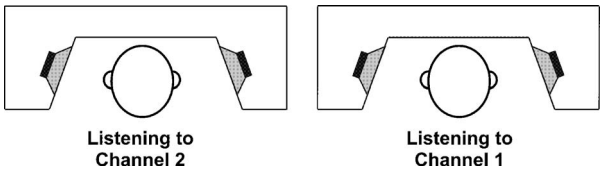


FIG. 1. Plan view of the physical arrangement in which listeners in adjacent seats are listening to different audio channels radiated by loudspeakers in the seat headrests.

with volume velocity q_p, which is also fed via filters H_1 and H_2 to the two loudspeakers on the left-hand seat which then have volume velocities q_{s1} and q_{s2}.

In order to quantify the potential acoustic behavior of such a system, a performance criterion that can be optimized needs to be defined. A convenient criterion would be to minimize the sum of the squared pressures at a number of locations in the left-hand headrest while keeping the sum of the squared pressures at locations in the right-hand headrest constant. This is equivalent to maximizing the ratio of the sum of squared pressures in the right-hand headrest to the sum of squared pressures in the left-hand headrest, which Choi and Kim[13] refer to as the acoustic contrast. The complex pressures at the locations that we are required to maintain at a high sound level, which may be termed the bright zone, p_1-p_4 in Fig. 2, can be written, for single frequency excitation, as elements of the vector

$$\mathbf{p}_B = \mathbf{Z}_B\mathbf{q}, \tag{1}$$

where $\mathbf{q}$ is a vector containing all the source strengths, which are q_p, q_{s1}, and q_{s2} in this case, and $\mathbf{Z}_B$ is a matrix of acoustic impedances from each source to every point at which the pressure is sampled in the bright zone. Similarly we can write the vector of complex pressures in the region where we want to reduce the levels, which may be termed the dark zone, p_5-p_8 in Fig. 2, as

$$\mathbf{p}_D = \mathbf{Z}_D\mathbf{q}. \tag{2}$$

The acoustic contrast, C, which is the ratio of the sum of squared pressures in the bright and dark zones, may thus be defined as

$$C = \frac{\mathbf{p}_B^H\mathbf{p}_B}{\mathbf{p}_D^H\mathbf{p}_D} \tag{3}$$

where the superscript H denotes Hermitian, complex conjugate, transpose. The acoustic contrast is a scalar quantity that we seek to maximize by adjusting q_{s1} and q_{s2} with respect to

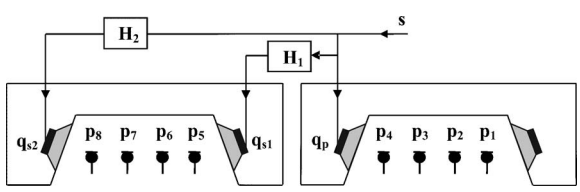


FIG. 2. Feedforward control strategy in which the loudspeakers in the left-hand headrest are used to control the sound in that seat due to a loudspeaker in the right-hand seat being driven by audio channel 1, shown here as the signal s.

q_p. Using Eqs. (1) and (2) the acoustic contrast may also be written as

$$C = \frac{\mathbf{q}^H\mathbf{Z}_B^H\mathbf{Z}_B\mathbf{q}}{\mathbf{q}^H\mathbf{Z}_D^H\mathbf{Z}_D\mathbf{q}}, \tag{4}$$

which Choi and Kim[13] show is maximized if $\mathbf{q}$ is the eigenvector of the matrix $[\mathbf{Z}_D^H\mathbf{Z}_D]^{-1}\mathbf{Z}_B^H\mathbf{Z}_B$ that corresponds to its largest eigenvalue. Once this optimum value of the vector $\mathbf{q}$ is calculated at a number of frequencies, the ratio of q_{s1} and q_{s2} to q_p can be calculated to determine the frequency responses of the filters H_1 and H_2 in Fig. 2.

An initial simulation has been performed for this geometry assuming that each of the acoustic sources behave as monopoles in free space. The distance between the two sources in one headrest, q_{s1} and q_{s2}, taken to be 27 cm, the distance between the closest sources in adjacent headrests, q_{s1} and q_p, was taken to be 22.4 cm, and four microphones were used to define the bright and dark zones each 4 and 7 cm from their nearest source. (The 4 cm distance is composed of a physical separation of about 1 cm and a correction for the finite size of the loudspeaker of one quarter of its diameter, as discussed in the Appendix.) These dimensions approximate those of the practical arrangement described in Sec. III. The sources and microphones are thus arranged in a line, although arrays that are not linear have also been investigated with similar results.

The results of this simulation are a pressure distribution for the optimized source strengths, whose contour plots in a two-dimensional plane running through the sources and microphones are shown in Fig. 3 for various frequencies of excitation. Although the performance is good at low frequencies, the sound field becomes more complicated and localized at high frequencies so that small changes in listening position cause large changes in acoustic pressure.

The reason why only local attenuations in pressure are achieved at high frequencies is that the secondary sources are a significant distance from the primary compared with the acoustic wavelength.[13] In order to be able to control the pressure over a larger region of the dark zone at high frequencies, a secondary source must be used which is close to the primary. One such arrangement is illustrated in Fig. 4. In this arrangement a single secondary source, q_s, located in the same headrest as q_p is fed via a filter H_1 from the primary source, and the frequency response of H_1 that maximizes the acoustic contrast can again be calculated using the abovementioned theory, except that the vector $\mathbf{q}$ now only contains q_p and q_s.

The results of monopole simulations with this geometry, where the primary and secondary are assumed to be 5 cm apart, are shown as contour plots in Fig. 5 for various frequencies of excitation. It can be seen that the region over which the pressure has been reduced by 20 dB compared with the pressure in the bright zone is considerably larger at high frequencies than it was in Fig. 3. The far-field directivity of this pair of monopoles is approximately cardioid below about 4 kHz, indicating that useful reductions in the sound from the right-hand headrest have been produced at all points to the left of it up to this frequency.

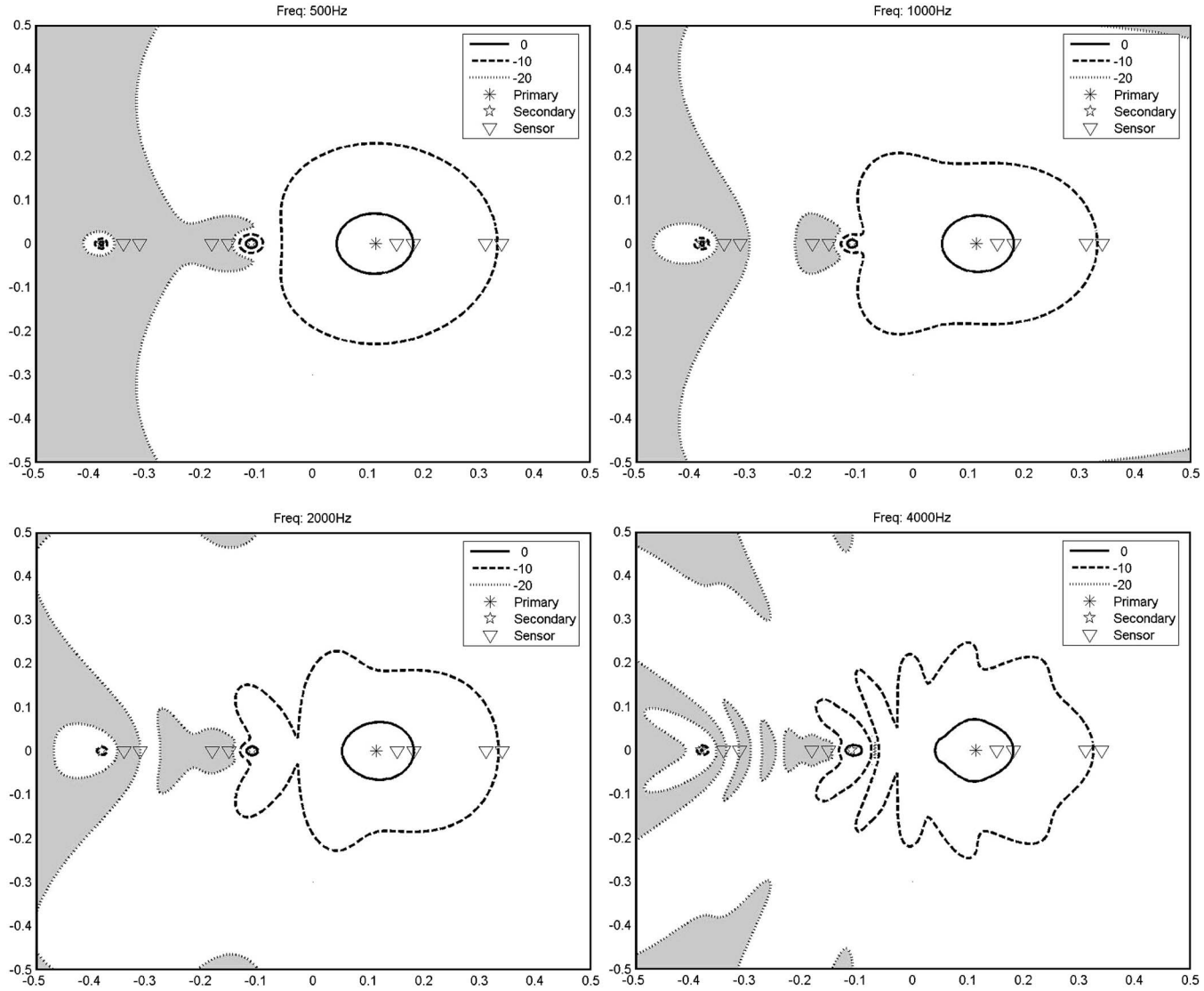


FIG. 3. Contour plots of the pressure levels at excitation frequencies of 500 Hz, 1, 2, and 4 kHz in a free-field acoustic model. Two secondary sources q_{s1} and q_{s2} (at $x=-0.11$ and -0.38 m) are used, remote from q_p (at $x=0.11$ m), to maximize the acoustic contrast between the pressures in the two seat positions. This maximization is done with respect to the microphones denoted by triangles. The shaded area indicates a reduction of 20 dB compared with the pressure at a point in the right-hand seat.

More subtle behavior can be revealed if the pressure distribution along the axis of the sources and microphones is plotted, and compared with that generated when q_s is adjusted to cancel the pressure in the far-field. These levels are plotted, at 500 Hz, as the solid and dashed lines in Fig. 6 and the former can be considered as a horizontal slice through the corresponding contour plot in Fig. 5. When q_s is adjusted to cancel the far-field pressure, the pressure field to the left of the sources falls off monotonically as $1/r^2$, where r is the distance from the sources. When q_s is adjusted to maximize the acoustic contrast between the two seats, the pressure field goes to zero at about the position of the microphone measuring P6 in Fig. 4 but then rises before falling off as $1/r$ at larger distances than are shown in this figure. The maximum level to the left of this dip is still about 40 dB down on the level in the bright zone, however, so although the far-field attenuation is not as great as when the far-field pressure is canceled, it is still probably sufficient for audio purposes.

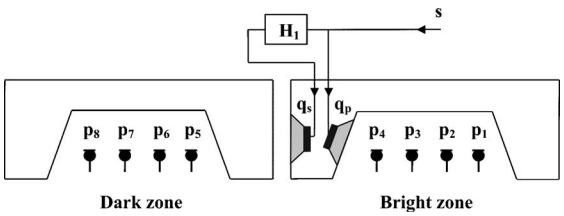


FIG. 4. Feedforward control strategy in which the outer loudspeaker of a pair mounted on one side of the headrest in the right-hand seat is used to control the sound in the left-hand seat due to the inner loudspeaker, which is driven by the audio signal s.

An important implication of the level curve in Fig. 6, which is confirmed by separate simulations, is that with just two sources, the optimal result for contrast maximization can be closely approximated by placing a null at a point in the dark zone. With such a simple geometry the system is unable to produce a complicated zone of silence, and just a single null can be created whose center point may be adjusted to maximize the contrast. An almost identical pressure distribu-

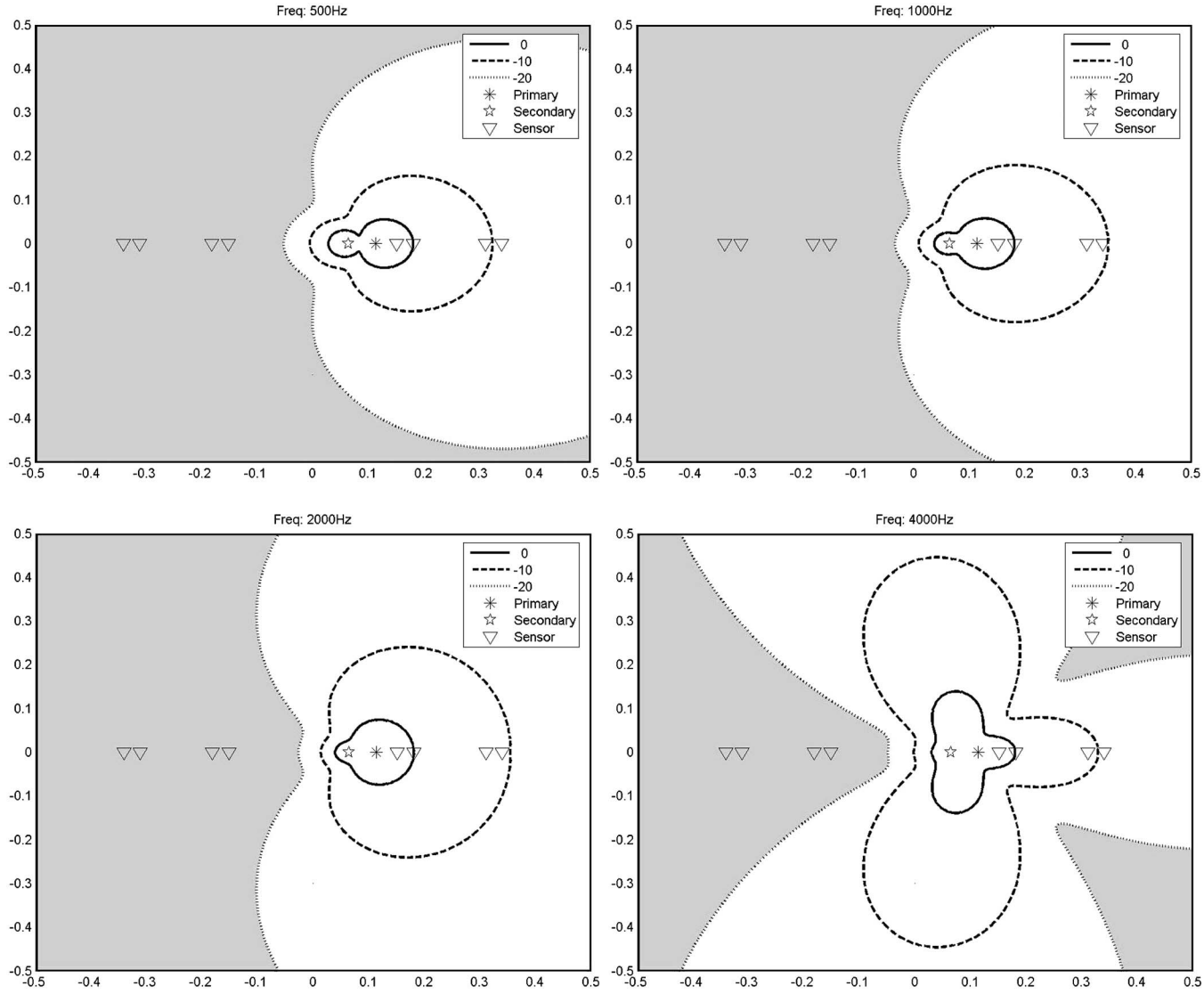


FIG. 5. Contour plot of the pressure levels at 500 Hz, 1, 2, and 4 kHz in a free-field acoustic model simulating the use of the secondary acoustic source q_s (at $x=0.06$ m) to maximize, close to q_p (at $x=0.11$ m), the acoustic contrast between the pressures in the two seat positions. The microphones over which the contrast is maximized are denoted here as triangles. The shaded area indicates a reduction of 20 dB compared with the pressure at a point in the right-hand seat.

tion could be achieved by adjusting q_s to simply cancel the pressure at a single point in the dark zone. This is a control strategy that would be considerably simpler to implement in a practical arrangement than contrast maximization, and is investigated further in the following section.

III. EXPERIMENTAL IMPLEMENTATION

Figure 7 shows a photograph of the experimental arrangement in which a pair of aircraft seats is fitted with prototype headrests containing loudspeakers. Dummy heads with microphones at the ear positions were arranged in the two seats to measure the pressures. In the inner headrest of the furthermost seat a pair of loudspeakers is fitted, mounted together facing each other, as shown in the right-hand side of Fig. 7. The loudspeakers had a diameter of 11.5 cm and were selected because their thickness was only 4.2 cm, so that a pair only had a depth of 8.4 cm when fitted into the headrest.

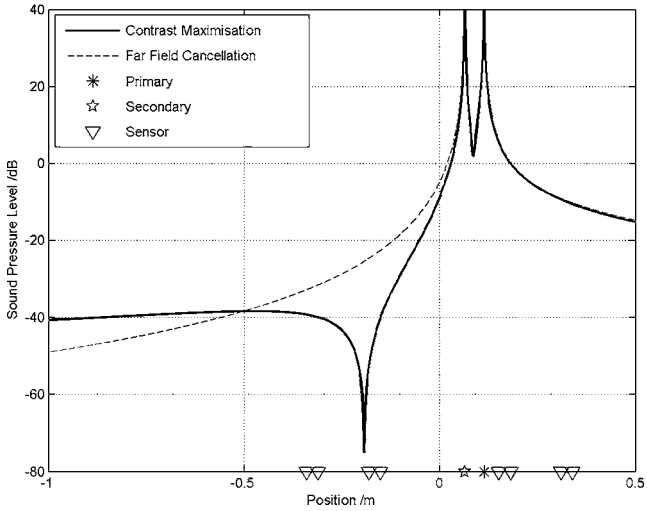


FIG. 6. Pressure level along the line joining the sources and microphone positions in Fig. 5 when the secondary source is adjusted to maximize the acoustic contrast, solid line, and when it is adjusted to cancel the far-field pressure, dashed line.

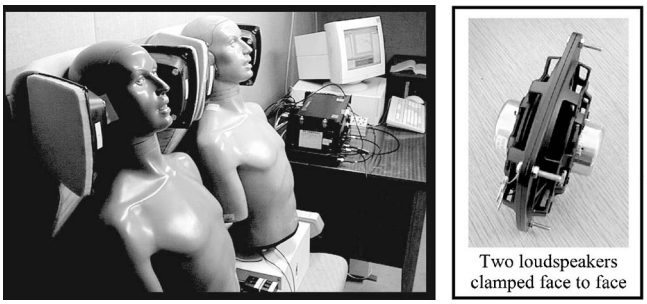


FIG. 7. Practical headrest arrangement with dummy heads in the two listening positions, left, and the two face-to-face loudspeakers to be mounted in the left-hand headrest of the right-hand seat, right.

The remaining headrests were only fitted with a single loudspeaker, of a similar type, in this preliminary setup.

Initially experiments were conducted in an anechoic chamber, in which the frequency response was measured from each loudspeaker to the microphones in each of the dummy heads, and also to microphones attached to the sides of the headrests, about 1 cm from the facing loudspeaker. The sum of the squared signals from the microphones both in the dummy head and on the surface of the headrest were used to define the overall level in the bright zone in the right-hand seat, and in the dark zone, in the left-hand seat. By manipulating the set of measured frequency responses, the optimum drive signals to the loudspeaker pair required to maximize the acoustic contrast was calculated, frequency by frequency, using the theoretical formulation discussed in Sec. II. The results in terms of this acoustic contrast are shown in Fig. 8, where the contrast without control, passive, and the contrast with control, active, are shown. Even when the outer loudspeaker is not being driven electrically, its cone moves

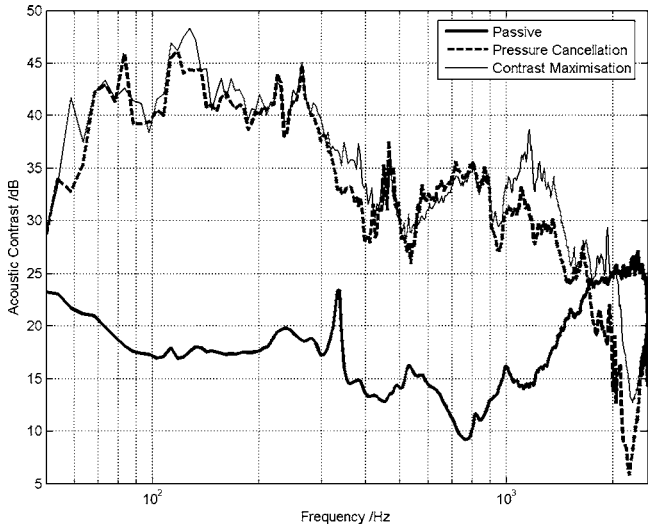


FIG. 8. Predictions of the acoustic contrast calculated as a function of frequency from the individual frequency response measurements between the loudspeakers and microphones in the anechoic chamber. The solid line is when only the single loudspeaker in the closed-back enclosure is driven, passive, the dashed line is when the outer loudspeaker in the open-backed pair is driven to cancel the pressure at a point in the adjacent seat, and the feint line is when this outer loudspeaker is driven to maximize the acoustic contrast. These results are almost identical if the control filters are constrained to be causal and truncated to 320 coefficients, which is the length used in the real-time implementation.

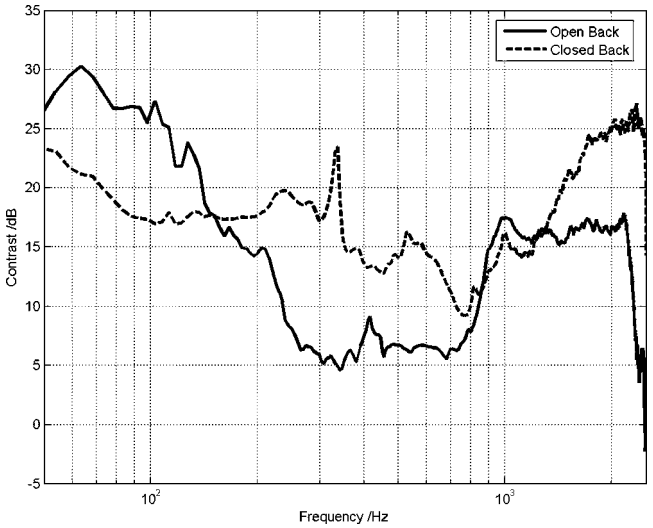


FIG. 9. Predictions of the acoustic contrast calculated from the individual frequency responses from a single loudspeaker and the microphones in the anechoic chamber. The solid line is when driving the inner loudspeaker of the loudspeaker pair mounted in an open-back enclosure, and the dashed line is when driving a single loudspeaker mounted in a closed-back enclosure in the same position.

due to acoustic coupling with the inner loudspeaker. In order to compare the actively controlled results with a passive system that approximated a single source, the open-backed loudspeaker pair was replaced by a single loudspeaker in a sealed enclosure, and all the pressure measurements repeated, in order to calculate the "passive" response with no control.

The acoustic contrast predicted using the simple monopole model for this geometry with no control is about 14 dB, independent of frequency. The measured acoustic contrast of the passive system is clearly worse than this from about 400 to 1 kHz, which is thought to be due to the presence of reflecting objects and local acoustic effects like resonant cavities. The passive acoustic contrast gets even worse in this frequency range when the back of the loudspeaker is removed and the second loudspeaker introduced, so that although the secondary loudspeaker in the pair is not driven electrically in the passive condition, it is acoustically excited by the primary loudspeaker so that the loudspeaker pair shows significant backward radiation. Figure 9 shows the results comparing the contrast without control using the loudspeaker pair, which is mounted in an open-backed enclosure, with that using only the single loudspeaker in a closed-backed enclosure. Apart from the peak at about 300 Hz, the calculated contrast with the single loudspeaker in the closed backed enclosure is more uniform than that of the loudspeaker pair in the open backed enclosure, which is particularly poor from about 200 to 800 Hz.

Of more significance than the absolute level of the contrast before control, however, is the improvement in acoustic contrast when the second loudspeaker is optimally driven. The improvement in acoustic contrast predicted by the monopole model when using a pair of monopoles rather than a single one is about 25–30 dB in this frequency range up to 2 kHz. The contrast calculated from the measured data, when the contrast is maximized, is shown as the feint line in Fig. 8.

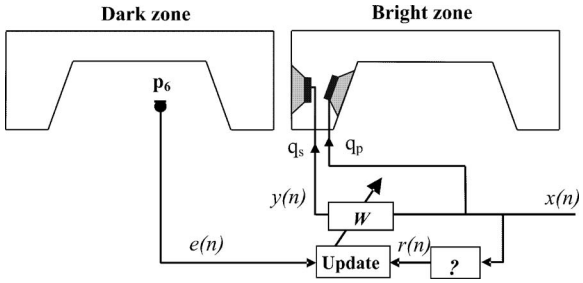


FIG. 10. Practical arrangement for the real-time implementation of an adaptive filter, **W**, driving the outer loudspeaker from the reference signal driving the inner loudspeaker, x, using the filtered reference LMS algorithm. The filter $\widehat{G}$ is the estimate of the plant response. Data converters and smoothing filters have not been included for clarity.

The improvement in contrast over the single loudspeaker result is about 20 dB below 300 Hz, falling to about 15 dB up to about 1 kHz. The frequency range was limited to 2 kHz in these experiments, but measurements on a similar arrangement by Chung[10] indicated that the predicted improvements in contrast became small above 2 kHz. The simple monopole model thus provides useful guidance on the performance of such systems up to 2 kHz, even though accurate prediction is not possible because the physical presence of the headrests and the heads of the listeners complicate the details of the sound field in the practical case.

Also shown by the dashed line in Fig. 8 is the acoustic contrast predicted from the anechoic measurements when the pressure at the ear of the right-hand dummy head closest to the adjacent seat was set to zero. This control strategy is inspired by the zero in the predicted pressure field in Fig. 6, and results in contrast values which are almost as high as those predicted under optimal conditions over the whole of the frequency range. The results from contrast maximization and pressure cancellation become even more similar when the control filter driving the second loudspeaker is constrained to be causal and of finite length. These results suggest that for this geometry near-optimal performance can be obtained using a rather simple control strategy.

Motivated by this similarity, a real time digital control system was implemented, using a dSPACE hardware/software system,[14] whose block diagram is shown in Fig. 10. The original audio signal driving the inner loudspeaker of the pair, $x(n)$, is passed through the adaptive filter W before being used to drive the outer loudspeaker. The filter is adapted using the filtered reference LMS algorithm[8] to minimize the pressure at one ear of the dummy head in the adjacent seat. The adaptive FIR filter used in this implementation has 320 coefficients and the sample rate used is 5 kHz. Once adapted, the coefficients of W are fixed, and the performance measured using a spectrum analyzer to estimate the combined frequency response of both sources to all the microphone positions. For these results the pair of seats is located in a small room approximately 6 m $\times$ 5 m $\times$ 2 m in size which had a reverberation time, T_{60}, of between 420 and 550 ms in the third octaves from 200 Hz to 2.5 kHz.

Figure 11 shows the acoustic contrast measured in the small room when driving only a single closed-back loudspeaker, and also the contrast when the pair of loudspeakers

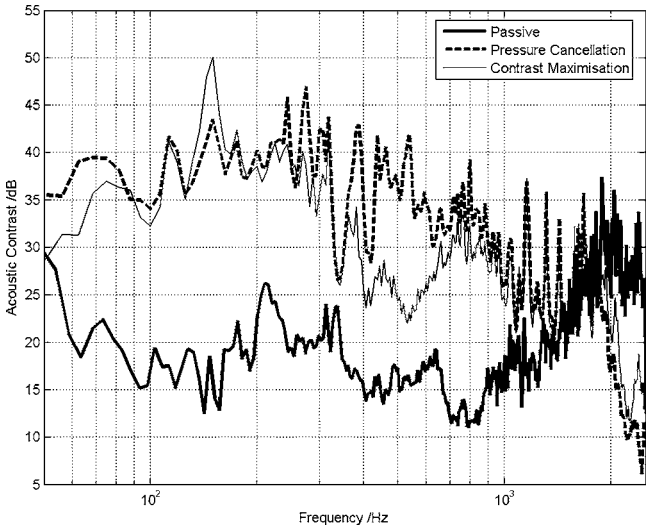


FIG. 11. The acoustic contrast as a function of frequency measured in a small room when driving only a single loudspeaker in a closed-back enclosure, solid, when the feedforward filter is implemented with a real-time system adapted to cancel the pressure at a point in the adjacent headrest (using the loudspeaker pair mounted in the open-back enclosure), dashed, and when the individual frequency responses are used to calculate the optimum performance for contrast maximization, feint.

mounted in the open-backed loudspeaker are driven, once the adaptive filter has fully converged. It is worth noting how similar these curves, measured with a real-time implementation in the small room, are to the corresponding curves in Fig. 8, calculated from the measurements made in the anechoic chamber. The passive contrast with the closed-back loudspeaker is greater in the room compared to that predicted in the anechoic chamber, which is probably due to increased levels in the bright zone due to reflections from nearby surfaces. There are also some additional dips and peaks that occur both before and after control in the small room, Fig. 11, due to individual modes in the room. It can be seen, however, that the general enhancement in contrast achieved using active control, and the variation with frequency, is similar for the two acoustic conditions. This is because the individual frequency responses between the loudspeakers and microphones are dominated by the direct response of the loudspeakers rather than the reflections off the walls, and are thus similar in the anechoic chamber and the small room. Most importantly, the levels of contrast using direct cancellation, implemented as shown in Fig. 10, are similar to those predicted from the anechoic chamber results and still show a significant improvement over the passive results. The feint line in Fig. 11 shows the predicted performance calculated from the individual frequency responses measured in the small room when the estimated contrast is maximized offline. It is interesting to note that at some frequencies the predicted optimal performance is worse than the performance measured after adaptive cancellation. This is thought to be due to the accumulation of errors in the calculation of the optimal performance from the measured frequency responses, which inevitably are not all measured simultaneously, and the ill conditioning of the matrix $\mathbf{Z}_D^H\mathbf{Z}_D$, which is inverted in the calculation of the filter that maximizes the

contrast in Eq. (4). The adaptive system is able to compensate for these small errors and thus achieve a better performance in this frequency range.

Informal listening tests indicate that the reduction in acoustic cross-talk between the seats is perceived as being very striking. The subjective impression when the active control system was switched on was that instead of the interfering sound appearing to come directly from the adjacent seat, it became more diffuse and noninvasive, and seemed to originate from many locations in the room. This is consistent with the control system operating to cancel the direct sound transmission, but not being able to cancel the reflected sound transmission, which will vary as people move about in the room. If the person in the left-hand seat was listening to a separate music channel or there was a reasonable level of background noise in the room, the audibility of the sound from the adjacent seat, when played at normal listening levels for a listener in that seat, would be significantly reduced when the control system was switched on.

Most importantly, however, was that good subjective impressions were maintained even if the listener moved their head side-to-side by the 5 cm or so distance allowed by the presence of the headrest, and up to about 15 cm forward and back or up and down. This suggests that the initial predictions from the monopole model, Fig. 5, that the sound is reduced over a significant region of space, also hold true in a practical implementation on a real headrest.

IV. CONCLUSIONS AND DISCUSSION

A number of arrangements of loudspeakers in the headrests of adjacent seats have been investigated to improve the acoustic isolation of reproduced sound between the seats. These were initially investigated using a simple free field model. An arrangement in which the loudspeakers in one seat were used to control the sound from an adjacent seat was predicted to give rather small zones of quiet so that the attenuation would be significantly affected by movements of the listener's head. A second arrangement looked more promising, in which a closely spaced pair of loudspeakers was used in one side of the headrest to both reproduce the sound in the seat to which it is attached and to cancel the sound radiated to the adjacent seat. This was predicted to produce much larger zones of quiet at higher frequencies, because the sound fields from the two loudspeakers would be much more closely matched if they were spaced close together compared with an acoustic wavelength. With a spacing of 5 cm the sources are within a tenth of a wavelength of each other, which is the spacing required for 10 dB of global control,[15] up to about 700 Hz.

These predictions were then investigated experimentally using a pair of thin loudspeakers mounted face to face in one side of the headrest. Measurements in the anechoic chamber confirmed that improvements in isolation of the order of those predicted should be possible, and also that this could be closely approximated by the practical strategy of adjusting a filter driving the outer loudspeaker from the signal driving the inner one to cancel the pressure at a point in the adjacent seat. An adaptive digital filter was then used to adjust such a filter in a real-time implementation. When the acoustic isolation was measured in a small room, the results were similar to those predicted from the anechoic chamber measurements, indicating that it is the direct field of the interfering sound which is dominant, even in a small room. Improvements in acoustic isolation of between 5 and 25 dB were measured from about 50 Hz to 2 kHz. These values are defined by comparing the contrast when the system is switched on with the passive contrast when only a single closed-back loudspeaker is driven.

Informal listening tests indicated that the reduction in crosstalk between the seats was impressive, that the residual sound was less intrusive, and that the subjective impression was maintained even for significant head movements.

The frequency response of the inner loudspeaker at the original listening position will be affected by such a cancellation system. However, it would always be possible to compensate for these effects at a single point by using a second filter to act on the signal $x(n)$ in Fig. 10 before it is fed to the inner loudspeaker. More formal listening tests are required to quantify the effects of such a system on the sound heard by two listeners in the adjacent seats.

It is not yet certain whether the control filter needs to be adapted for different environments, or whether a filter with a single fixed frequency response will give acceptable performance. If a fixed filter were acceptable, a simple modular system could be incorporated into each side of the headrest that only required a single input signal.

ACKNOWLEDGMENTS

The work of M.J. is supported by EPSRC Grant No. GR./S63915. We are also grateful to Ian Stothers of Ultra Controls for initial discussions and the supply of the loudspeakers.

APPENDIX: EQUIVALENT MONOPOLE SOURCE POSITION FOR NEAR FIELD OF PISTON

The complex pressure a distance r along on the axis of a baffled piston of complex volume velocity q and radius a is given, assuming an $e^{j\omega t}$ time dependence, by[16]

$$p(r) = \frac{\rho_0 c_0 q_0}{\pi a^2}\left[e^{-jkr} - e^{-jk\sqrt{r^2+a^2}}\right]. \tag{A1}$$

Note that the piston velocity U_0 in Ref. 16 has been replaced by the volume velocity q_0 using the relation $U_0 = q_0/\pi a^2$, where πa^2 is the piston surface area.

If the pressure is measured very close to the piston, so that $r \ll a$, and the frequency is low enough that $kr \ll 1$, so that e^{-jkr} is almost unity, then

$$p \approx \frac{\rho_0 c_0 q_0}{\pi a^2}\left[1 - e^{-jka}\right]. \tag{A2}$$

Expanding the exponential into three terms of its infinite series, assuming $(ka)^3$ is negligible, Eq. (A2) can be written as

$$p \approx \frac{\rho_0 c_0 q_0}{\pi a^2}\left[1 - 1 + jka + \frac{(ka)^2}{2}\right], \tag{A3}$$

so that

$$p \approx \frac{j\omega\rho_0 q_0}{\pi a}\left[1 - j\frac{ka}{2}\right] \approx \frac{j\omega\rho_0 q}{\pi a}e^{-jkal/2}. \tag{A4}$$

This expression may be compared with the pressure a distance of d from a monopole source of strength q in an infinite baffle given by[16]

$$p = \frac{j\omega\rho_0 q_0}{2\pi d}e^{-jkd}. \tag{A5}$$

The magnitude and phase of the pressure in the near field of the piston are thus the same as that of an equivalent monopole source at a distance of

$$d = \frac{a}{2}, \tag{A6}$$

i.e., the distance from the equivalent monopole is half the piston radius, or one quarter of its diameter.

[1]H. Hashimoto, K. Terai, M. Kiba, and Y. Nakama, "Active noise control for the seat audio system," Proceedings of ACTIVE95, pp. 1279–1290.

[2]D. L. Smith, "Discrete-element line arrays—Their modelling and optimisation," J. Audio Eng. Soc. **45**, 949–964 (1997).

[3]Special Issue on Loudspeakers Arrays. J. Audio Eng. Soc. **38** (1990).

[4]W. F. Druyvesteyn and J. Garas, "Personal sound," J. Audio Eng. Soc. **45**, 685–701 (1997).

[5]M. Yoneyama *et al.*, "The audio spotlight," J. Acoust. Soc. Am. **73**, 1532–1536 (1983).

[6]P. J. Westervelt, "Parametric acoustic array," J. Acoust. Soc. Am. **35**, 535–537 (1963).

[7]C. H. Hansen, "Current and future industrial applications of active noise control," Proceedings of ACTIVE04, International Symposium on Active Sound and Vibration, Williamsburg, VA, 2004, Paper 53.

[8]S. J. Elliott, *Signal Processing for Active Control* (Academic, New York, 2001).

[9]S. J. Elliott, M. Abulula, and T. Bravo, "Active control and the design of nearfield arrays," Proceedings of ICSV10, Tenth International Congress in Sound and Vibration, Stockholm, 2003, pp. 243–250.

[10]C. Chung, *Personal Audio Seat*, M.Sc. thesis, University of Southampton, 2005.

[11]M. Jones and S. J. Elliott, "The implementation of an active headrest for personal audio," Proceedings of ICSV12, 12th International Congress on Sound and Vibration, Lisbon, 2005, Paper 93.

[12]M. Jones and S. J. Elliott, "An active headrest for personal audio," Proceedings of IWAENC, 2005 International Workshop on Acoustic Echo and Noise Control, Eindhoven, 2005, Paper 60.

[13]J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," J. Acoust. Soc. Am. **111**(4), 1695–1700 (2002).

[14]http://www.dspaceinc.com/

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

[16]L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd ed. (Wiley, New York, 2000).

Sources of excess noise in silicon piezoresistive microphones[a)]

Robert Dieme, Gijs Bosman, and Toshikazu Nishida[b)]
Department of Electrical and Computer Engineering, Interdisciplinary Microsystems Group, University of Florida, Gainesville, Florida 32611-6200

Mark Sheplak
Department of Mechanical and Aerospace Engineering, Interdisciplinary Microsystems Group, University of Florida, Gainesville, Florida 32611-6250

(Received 17 November 2005; revised 24 February 2006; accepted 24 February 2006)

This paper provides evidence for electrical $1/f$ noise as the dominant source of excess noise in piezoresistive microelectromechanical systems (MEMS) microphones. In piezoresistors, the fundamental noise sources may be divided into frequency independent thermal noise and frequency dependent $1/f$ excess noise dominating at low frequencies. Noise power spectra are presented for both commercial and research-prototype MEMS piezoresistive microphones as a function of applied voltage bias for both free and blocked membranes. The contributions of various mechanical and electrical noise sources are compared using a lumped noise equivalent circuit of the piezoresistive microphone. The bias dependence and membrane independence of the output noise indicate that the primary source of the excess noise is electrical in origin. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2188367]

PACS number(s): 43.38.Kb, 43.38.Gy, 43.38.Ar [AJZ] Pages: 2710–2720

I. INTRODUCTION

The reduction of acoustic microphone size using microelectromechanical systems (MEMS) technology enables increased spatial resolution. However, whether the small size can be effectively utilized depends on the bandwidth, linearity, and the minimum detectable signal (MDS) of the microphone. The MDS is proportional to the total root-mean-square (rms) output noise due to all (acoustic, mechanical, and electrical) noise sources and inversely proportional to the overall sensitivity which depends on the transduction method and geometry.

The focus of this paper is on noise sources in piezoresistive MEMS microphones.[1–6] Piezoresistive transduction exhibits favorable scaling for small size compared to capacitive techniques.[7] Internal noise sources in piezoresistive microphones are divided into two categories. The first consists of noise sources present in all diaphragm-based microphones, pressure fluctuation noise due to the acoustic vent resistance, and mechanical thermal noise resulting from molecular agitation and structural damping of the diaphragm.[8] The second comprises noise sources specific to piezoresistive transduction, electrical thermal noise,[9,10] and electrical excess (or low frequency or $1/f$) noise in the piezoresistor.[11,12] Recently, there have been suggestions of a mechanical $1/f$ noise source that is correlated with the mechanical diaphragm damping resistance in condenser, piezoelectric, and piezoresistive microphones.[13,14] However, in contrast, electrical $1/f$ noise has been observed to be the dominant noise source in small piezoresistive cantilevers used in atomic force microscopy,[15,16] piezoresistive accelerometers,[17] and piezoresistive microphones.[3]

This present study investigates the dependence of output noise power spectra on bias voltage and membrane structure to elucidate the primary contribution to the measured low frequency noise in piezoresistive microphones. Both commercial and research prototype piezoresistive microphones are examined. The organization of this paper is as follows. First, noise mechanisms in piezoresistors are identified and the piezoresistive microphone structure and lumped noise equivalent circuit are reviewed. Next, the electrical noise measurement setup and method for extracting sensor noise from the amplifier setup noise are described. Experimental results using a metal film resistor with known noise spectra are presented and compared to theory. Then, output noise power spectra for commercial and research prototype piezoresistive MEMS microphones are measured as a function of bias voltage for free and fixed membrane structures. The paper concludes with a discussion of the dominant noise sources in piezoresistive microphones.

II. NOISE

In general, random fluctuations in a sensor output that do not originate in the input signal compose noise. In the electrical domain, nonrandom fluctuations (deterministic interference) caused by capacitively or inductively coupled electromagnetic interference may be present in the output signal, most commonly at 50 or 60 Hz and its harmonics due to interference from external power lines. These externally sourced deterministic fluctuations may be reduced through adequate shielding and proper grounding. However, intrinsic random fluctuations that originate from the sensor itself or the downstream amplifier circuit cannot be avoided by

[a)]Portions of this work were presented in "Sources of Excess Noise in Silicon Piezoresistive Microphones" at the 148th Meeting of the Acoustical Society of America, San Diego, CA, November 2004.
[b)]Electronic mail: nishida@ufl.edu

 0001-4966/2006/119(5)/2710/11/$22.50

shielding. These interfering random signals can be large enough to mask the signal of interest. Intrinsic noise sources are classified next in terms of the thermodynamic state and energy domain.

A. Noise sources

Noise in MEMS sensors can be classified by the absence (equilibrium) or presence (non-equilibrium) of external stimuli (voltage, light, thermal gradient, and strain). The magnitude of the noise depends on the material properties and the sensor geometry. Contributions of external amplifiers to the measured output noise will be discussed in Sec. IV on the noise experimental setup.

1. Electrical noise

Equilibrium thermal noise. Random scattering of charge carriers in thermodynamic equilibrium with the lattice give rise to voltage fluctuations at the terminals of a conductor or semiconductor piezoresistor. This equilibrium electrical noise was modeled by Nyquist[9] via a thermal noise voltage power spectral density (PSD),

$$S_{th} = 4k_B RT\,[V^2/\mathrm{Hz}], \tag{1}$$

where k_B is the Boltzmann constant, R is the resistance, and T is the temperature in Kelvin and was experimentally verified by Johnson.[10] Electrical thermal noise is independent of bias voltage since the agitation of the charge carriers by thermal lattice vibrations is present regardless of bias voltage. However, an increase in the temperature induces more agitation of the charge carriers, hence the Johnson (thermal) noise PSD is temperature dependent. Furthermore, Johnson noise is frequency independent because lattice vibrations are random, and thus are not related to any time constants.

Nonequilibrium $1/f$ *noise.* When an external voltage or current bias is applied to imperfect electrical conductors such as silicon with interfacial or bulk defects, an excess noise above the thermal equilibrium noise floor is observed that exhibits an inverse frequency $(1/f)$ dependence. The mechanism that generates electrical $1/f$ noise is still an active area of research. The two widely accepted mechanisms of electrical $1/f$ noise are the fluctuation in the bulk mobility $(\Delta\mu)$ described by Hooge[11] and the fluctuation in the number of carriers (Δn) developed by McWhorter.[12]

Hooge originally conducted noise experiments on homogeneous metal and semiconductor samples at low frequency.[11] The noise PSD in semiconductor samples exhibited an inverse frequency $(1/f)$ dependence, which he suggested was due to mobility fluctuations $(\Delta\mu)$ (Ref. 18) in the bulk semiconductor. He gave an empirical formula for the noise PSD of $1/f$ noise

$$S_V = \frac{\alpha V_{\mathrm{bias}}^2}{Nf}[V^2/\mathrm{Hz}], \tag{2}$$

where α is the Hooge parameter (an empirical material parameter varying from 1×10^{-6} to 1×10^{-3}), V_{bias} is the bias voltage, and N is the number of carriers. Since α is process dependent, α includes the effect of bulk defects. It is important to note that this nonequilibrium noise source is only observable when a voltage is applied across the resistor. In addition, the noise PSD is inversely proportional to the number of carriers N. Thus, the low frequency noise depends on the piezoresistor doping concentration N_{doping} and piezoresistor volume.

McWhorter[12] postulated that $1/f$ noise is caused by fluctuations of the number of surface charge carriers, due to trapping and detrapping of charge carriers at these traps. At the semiconductor surfaces and interfaces, physical defects give rise to electronic traps that capture and emit charge. Both Hooge's and McWhorter's $1/f$ noise models are actively used for modeling low frequency noise in electronic devices. The excess noise defined within the Hooge model can be expressed in terms of an effective α parameter which is a convenient figure of merit for device noise comparison between devices.[19] Thus, Hooge's model will be used to analyze the measured low frequency noise PSD of the MEMS piezoresistive microphone.

2. Mechanical noise

Equilibrium mechanical thermal noise. Mechanical thermal noise is the mechanical analog of electrical thermal noise. By the fluctuation-dissipation theorem, any dissipative mechanism that results in mechanical damping must be balanced by a fluctuation force to maintain macroscopic energy balance.[8] In analogy with Eq. (2), mechanical thermal noise is given by

$$S_{mth} = 4k_B R_m T\,[N^2/\mathrm{Hz}], \tag{3}$$

where R_m is the equivalent mechanical resistance.[8] An example of a fluctuation-dissipation mechanism in MEMS pressure sensors is thermal agitation of molecules impinging the diaphragm.[20]

Nonequilibrium mechanical $1/f$ *noise.* A mechanical $1/f$ noise correlated with the mechanical diaphragm damping resistance in condenser, piezoelectric, and piezoresistive microphones has been suggested.[13,14] In Ref. 14, the output noise of a condenser microphone was measured in an acoustic isolation vessel that maintained the microphone at constant temperature and pressure. The membrane noise was extracted by subtracting the preamplifier noise power (obtained when no external voltage was applied across the capacitor) from the total noise power of the biased capacitive condenser microphone. The resulting $1/f$ noise power spectrum was attributed to the mechanical diaphragm damping resistance. In Ref. 13, in addition to the well-known noise mechanisms present in piezoresistive microphone (mechanical thermal noise, Johnson noise, and electrical $1/f$ noise), a mechanical $1/f$ noise, correlated to the diaphragm damping resistance, was proposed.

III. PIEZORESISTIVE MICROPHONE

Before describing the noise measurement, the piezoresistive microphone structure,[3] operation, and lumped equivalent circuit are reviewed. When a pressure fluctuation impinges on the microphone, the diaphragm deflects producing a mechanical strain field. In a Wheatstone bridge circuit (Fig. 1) comprised of four silicon piezoresistors located at the diaphragm edge where the stress is maximum, the pressure-induced strain is transduced into an output voltage signal as discussed below. For a balanced bridge, the nominal unstressed resistances are equal, $R_1=R_2=R_3=R_4=R$, such that at zero pressure, the output bridge voltage is zero. Top and cross-sectional views of a circular diaphragm piezoresistive microphone (UF1) are shown, respectively, in Figs. 2(a) and

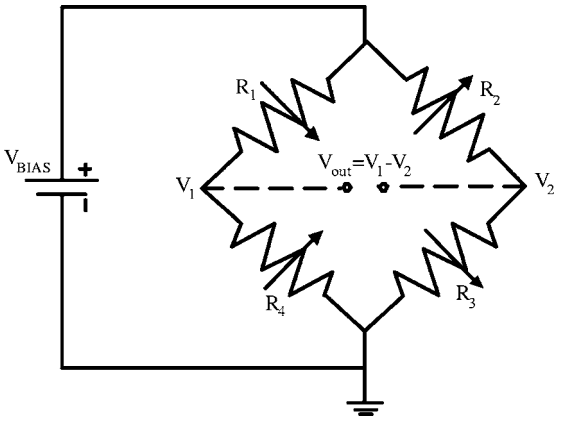


FIG. 1. Piezoresistors configured in Wheatstone bridge.

2(b).[3] Two tapered piezoresistors are positioned radially to sense the stress parallel to the current flow while the other two arc piezoresistors are placed tangentially to sense the stress perpendicular to the current flow. In a piezoresistor, the normalized resistance change $\Delta R/R$ is proportional to the longitudinal stress σ_l and transverse stress σ_t (defined with respect to the direction of current flow) through the longitudinal and transverse piezoresistance coefficients π_l and π_t, respectively,[21]

$$\frac{\Delta R}{R} = \sigma_l \pi_l + \sigma_t \pi_t. \tag{4}$$

When pressure is incident on a circular diaphragm microphone, the corresponding diaphragm pressure load p_d produces radial and tangential stresses $\sigma_r(r,\theta)$ and $\sigma_t(r,\theta)$ that are linearly related to p_d if the pressure-induced deflection is small.[22] Here, r and θ are the radius and polar angle, respectively. Noting that σ_l is longitudinal to current flow in the tapered resistor and transverse to current flow in the arc resistor, tapered and arc piezoresistors are designed with equal but opposite polarity resistance change[23]

$$\left.\frac{\Delta R(p_d)}{R}\right|_{\text{arc}} = -\left.\frac{\Delta R(p_d)}{R}\right|_{\text{tapered}}. \tag{5}$$

When the piezoresistors are arranged in a fully active Wheatstone bridge, these equal but opposite polarity resistance changes produce a differential output voltage $\Delta V_{\text{out}} = V_1 - V_2$ across the bridge that is proportional to the diaphragm pressure-induced resistance change $\Delta R(p_d)/R$ and the applied bias voltage V_{bias}

$$\Delta V_{\text{out}} = \frac{\Delta R(p_d)}{R} V_{\text{bias}}. \tag{6}$$

For use in the lumped element model of the piezoresistive microphone described below in Sec. III A, we define the mechanical voltage sensitivity

$$S_{\text{me}} = \left.\frac{\Delta R(p_d)}{R}\right|_{\text{arc}} \frac{V_{\text{bias}}}{p_d}. \tag{7}$$

A. Lumped element model

The dynamic input-output behavior of the piezoresistive microphone is modeled using an equivalent lumped two-port network circuit as shown in Fig. 3(a). Here, the distributed kinetic and potential energy in each energy domain of the microphone is lumped as an effective inertance and compliance[24] which is valid as long as the length scale of the microphone is much smaller than the acoustic wavelength. The lumped impedances are identified using two subscripts, the first refers to the energy domain, indicated by the subscript "a", for the acoustic domain, and the second refers to the specific component of the microphone. Given an input pressure $p_{\text{in}}(t)$ referenced to the microphone back cavity, there are two volumetric flow paths, (1) through the vent channel q_v and (2) through the deformation of the compliant diaphragm q_d. In the vent channel, dissipation via friction is modeled as a vent resistance R_{av}. The deformation of the compliant lossy diaphragm, induced by the volumetric flow,

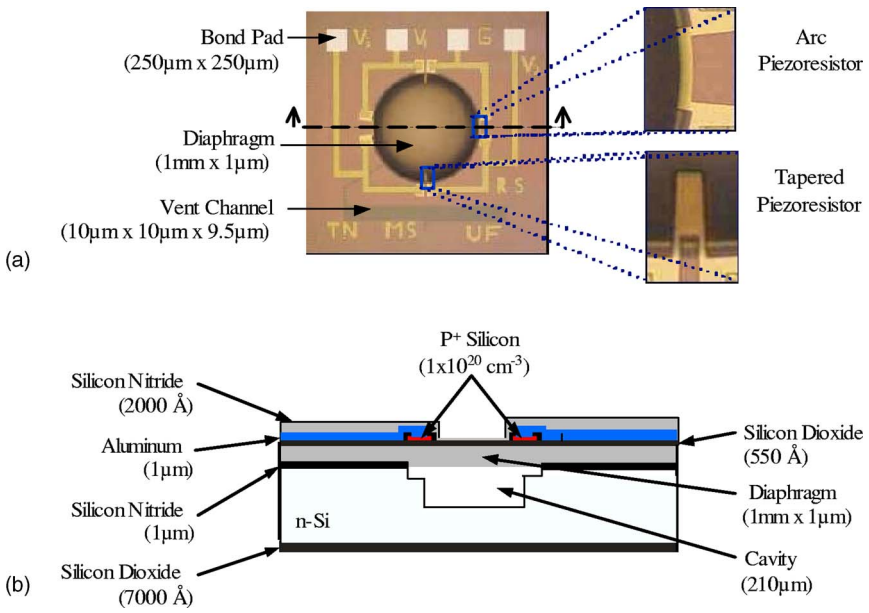


FIG. 2. (Color online) (a) Top view and (b) cross section of a piezoresistive microphone (Ref. 3).

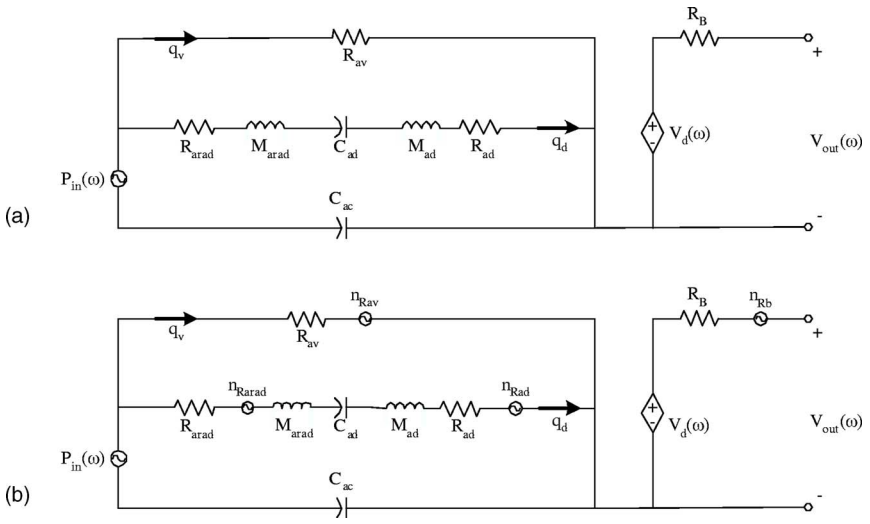


FIG. 3. (a) Lumped element model of a piezoresistive microphone, (b) noise equivalent circuit of a piezoresistive microphone.

is lumped in terms of an effective lumped mass of the diaphragm M_{ad}, damping resistance R_{ad}, and compliance of the diaphragm C_{ad}. In addition, the effect of the ambient fluid interaction with the membrane is modeled through the radiation mass of the diaphragm M_{arad} and the radiation resistance R_{arad}. The transduction of pressure to an output voltage is modeled as a dependent voltage source $V_d(\omega)$ in series with the bridge output resistance R_B. The dependent source shown in Fig. 3(a) is a function of the volumetric flow q_d instead of the bias voltage[25] since q_d is the flow variable in the acoustic domain part of the lumped equivalent circuit.

$$V_d(\omega) = \left(\frac{q_d}{j\omega C_{\mathrm{ad}}}\right) S_{\mathrm{me}}. \tag{8}$$

However, it is clear from the bias voltage dependence of S_{me} in Eq. (7) that there is no coupling between the acoustic input and the piezoresistive microphone electrical output if there is no applied bias voltage. Calculated values of the lumped parameters for the MEMS piezoresistive microphone shown in Fig. 2 are listed in Table I.[23,26]

B. Noise equivalent circuit

The noise equivalent circuit shown in Fig. 3(b) is obtained from the lumped element circuit [Fig. 3(a)] by shorting the input pressure, i.e., setting the acoustic pressure input to zero and replacing each dissipation element by an ideal noiseless dissipation element in series with its noise source.[27] Therefore, $n_{R_{\mathrm{av}}}(t)$ is the noise associated with R_{av}, $n_{R_{\mathrm{ad}}}(t)$ is the noise associated with R_{ad}, $n_{R_{\mathrm{arad}}}(t)$ is the noise associated with R_{arad}, and $n_{R_B}(t)$ is the noise associated with R_B. Also, M_{eff} is defined as $M_{\mathrm{av}}+M_{\mathrm{arad}}$. This noise equivalent circuit will be used to evaluate the relative contributions of the different noise sources in Sec. V.

IV. EXPERIMENTAL NOISE SETUP

Since the noise disturbance that limits the minimum detectable signal (MDS) of the microphone is small, special attention must be placed on the experimental noise setup, particularly to distinguish the noise in the device under test (DUT) from the noise present in the measuring equipment

TABLE I. Lumped element parameters of piezoresistive microphone (Ref. 26) computed for UF microphone (Ref. 3).

Definition	Acoustic impedance	Numerical values
Effective compliance of the cavity	$C_{\mathrm{ac}}=V_{\mathrm{cav}}/\rho_{\mathrm{air}}c^2$	$1.5619\times10^{-15}[\mathrm{m^5/N}]$
Resistance of the vent	$R_{\mathrm{av}}=128\mu L_{\mathrm{eff}}/\pi D_{\mathrm{vent}}^4$	$2.2402\times10^{+14}[\mathrm{N\,s/m^5}]$
Mass of the vent	$M_{\mathrm{av}}=16\rho_{\mathrm{air}}L_{\mathrm{eff}}/3\pi D_{\mathrm{vent}}^2$	$1.0715\times10^{-13}[\mathrm{N\,s^2/m^5}]$
Damping resistance	$R_{\mathrm{ad}}=2\pi f_0 M_{\mathrm{eff}}/Q$	$1.3277\times10^{+10}[\mathrm{N\,s/m^5}]$
Compliance of the diaphragm	$C_{\mathrm{ad}}=\pi a^4/2\sigma_0 h$	$1.1550\times10^{-15}[\mathrm{m^5/N}]$
Mass of the diaphragm	$M_{\mathrm{ad}}=\rho h/3\pi a^2$	$1.3581\times10^{+03}[\mathrm{kg/m^4}]$
Radiation resistance at 1 kHz (function of w)	$R_{\mathrm{arad}}=\rho c\pi a^2(ka)^2/4S^2$	$1.1365\times10^{+04}[\mathrm{N\,s/m^5}]$
Radiation mass of the diaphragm	$M_{\mathrm{arad}}=8\rho_{\mathrm{air}}/3\pi^2 a$	$6.485\times10^{+02}[\mathrm{N\,s^2/m^5}]$

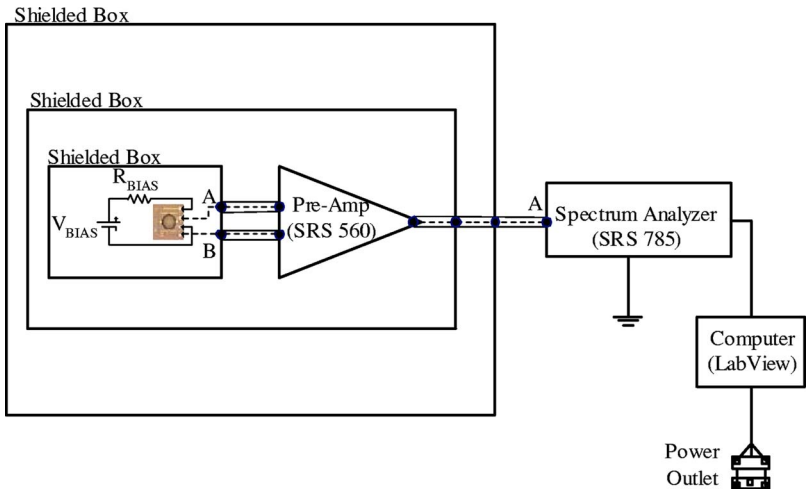


FIG. 4. (Color online) Experimental setup of the noise measurement.

and inadvertently picked up in the experimental setup. The setup noise PSD must be subtracted from the total measured noise PSD to determine the output noise of the DUT. The setup noise PSD is the contribution to the measured noise PSD by electromagnetic and acoustic interference as well as the noise present in the low-noise preamplifier (SRS560), the spectrum analyzer (SRS 785), and the power source. Electromagnetic interference is minimized by using three nested Faraday cages made up aluminum (3.48 mm thick), iron (1.46 mm thick), and high permeability magnetic foil (Less EMF Inc. - A276), single-point ground, and use of a lead acid battery (Panasonic - LC-R127R2P1) for biasing the MEMS microphones and the low noise preamplifier. The current and voltage noise PSD of the setup are measured, respectively, via open and short circuit of the preamplifier inputs. The preamplifier is ac coupled to block the dc bias. Differential input is employed to increase dynamic range through improved common mode noise rejection. When the SRS 560 is used during measurement, the filter is set as a bandpass filter with lower and higher limits set at 0.03 Hz and 300 kHz with ±6 dB/octave roll off. The gain is set to 1000 ensuring that the output signal of the preamplifier is higher than the noise floor of the spectrum analyzer ($<$10 nV_{rms} above 200 Hz) allowing accurate noise measurement. The spectrum analyzer is ac coupled with Hanning window selected to provide good selectivity and reduced power spectral density leakage. Overlapping frequency spans of 800 fast Fourier transform (FFT) lines are chosen in accordance to the degree of resolution desired. For instance, at low frequencies, a span of 12.5 Hz gives a frequency resolution of 15.625 mHz. At low frequencies, the binwidth is small to ensure good frequency resolution and therefore accurate measurements. At high frequencies, a larger frequency span of 12 800 Hz is used resulting in 16 Hz frequency resolution. Overall, the PSD is obtained by overlapping approximately the first 10% of each measurement span to minimize error due to leakage at the lowest frequencies since the digital window filter Hanning is not infinitely sharp. LABVIEW is used to automate the measurements. The experimental dc setup of the noise measurement is illustrated in Fig. 4.

The verification of the noise measurement setup is accomplished by using a 1 kΩ metal film resistor R_{dut} as the DUT since $1/f$ noise is minimal in metals at low current densities, thus its theoretical noise PSD is determined solely by the well-known thermal noise for a 1 kΩ resistance [4 $nV_{rms}/\sqrt{Hz}$ from Eq. (1)]. A small signal representation of the noise measurement setup is shown in Fig. 5. In this figure, S_{VRdut} is the thermal noise PSD associated with R_{dut}, and the input referred voltage and current noise sources of the preamplifier are represented by S_{va} and S_{ia}, respectively. The voltage noise source S_{va} of the preamplifier is obtained by shorting the input of the preamplifier, which allows the cancellation of S_{ia}. To measure the current noise power spectral density of the setup, the input of the preamplifier is open circuited. This permits cancellation of the voltage noise source S_{va}, allowing measurement of the current power spectral density S_{ia}. The voltage noise PSD of the DUT R_{dut} is extracted by subtracting the noise PSD contribution of the setup from the raw measured voltage PSD. The plot of the noise voltage PSD of the setup and raw PSD of a 1 kΩ metal film resistor without subtraction of the equipment setup noise

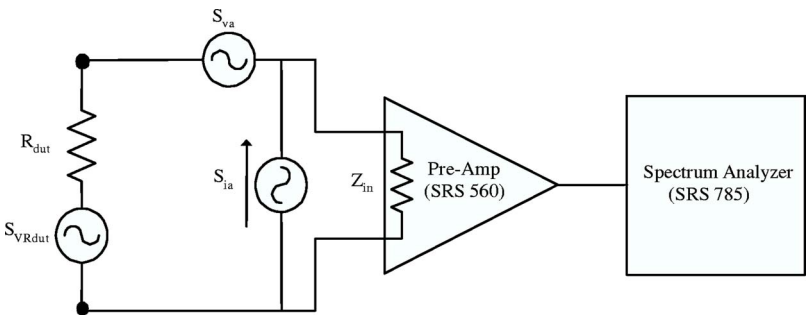


FIG. 5. Small signal representation of the setup with a 1 kΩ metal film resistor.

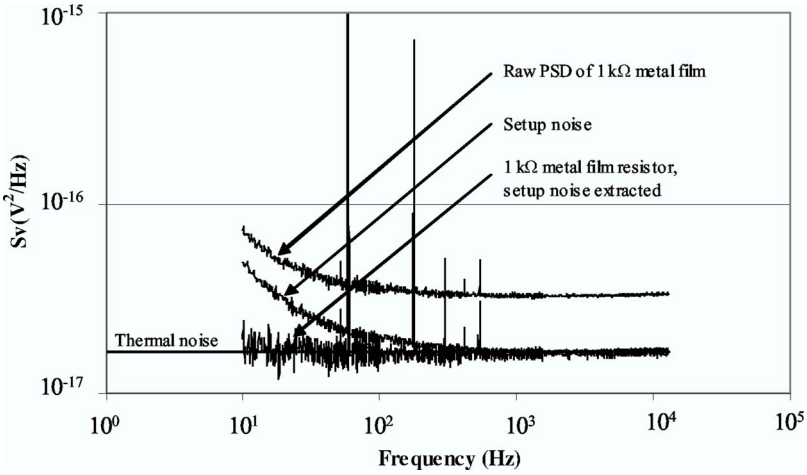


FIG. 6. Voltage noise PSD of a 1-kΩ metal film resistor without and with the subtraction of the equipment setup noise.

PSD is shown in Fig. 6. Once the setup noise PSD is subtracted from the raw noise PSD, the noise PSD of the 1 kΩ metal film resistor DUT is obtained. The noise PSD of the 1 kΩ metal film resistor, after the setup noise has been subtracted, also shown in Fig. 6, agrees well with the theoretical thermal noise PSD of a 1 kΩ metal film resistor, verifying the experiment noise setup and small signal model.

The preamplifier noise can be a limiting factor to a dc bridge noise measurement. If the preamplifier noise is larger than the DUT noise, the DUT noise PSD cannot be extracted from the raw measured noise PSD. To overcome this impediment, the microphone may be biased using an ac bias voltage to modulate the microphone low frequency noise to a frequency higher than the corner frequency of the preamplifier where it can be detected and measured. Figure 7 shows the setup for an ac bridge measurement.[28] Here, an ac modulation with carrier frequency f_c results in noise PSD at frequencies f_c-f_n and f_c+f_n, where f_n is the frequency component of the noise PSD. The ac bridge noise measurement method is useful when the measured low frequency noise PSD is dominated by the preamplifier noise at low frequencies. However, the measured ac modulated low frequency noise PSD must be multiplied by a factor of four to obtain the correct magnitude,[28] and there is a discontinuity of the PSD at the corner frequency since the frequency independent thermal noise cannot be ac modulated. Hence, the dc bridge noise measurement is preferred when the low frequency noise of the DUT is larger than the preamplifier noise so that a smooth, continuous noise PSD can be obtained.

V. MICROPHONE NOISE MEASUREMENT

In order to investigate the origin of the primary low frequency noise source in MEMS piezoresistive microphones, the noise PSD of four piezoresistive microphones was measured under equilibrium and varying nonequilibrium conditions. The four DUT consist of two commercial piezoresistive microphones [Kulite MIC-093 (Ref. 6) and Endevco 8510B-1 (Ref. 5)] and two research prototypes [piezoresistive microphone UF1) (Ref. 3) and piezoresistive

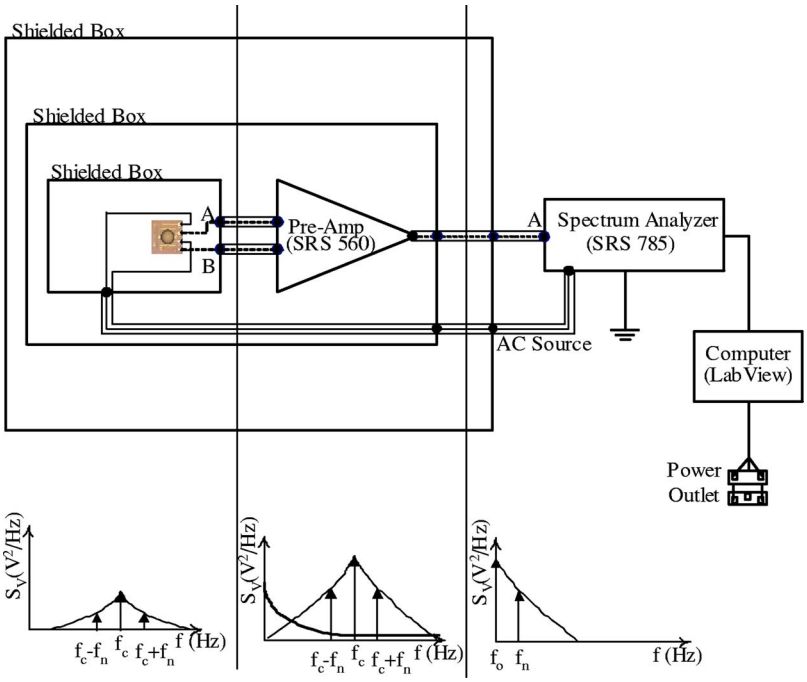


FIG. 7. (Color online) Experimental setup for an ac bridge measurement.

TABLE II. Measured input and output impedances and overall sensitivity of UF1 piezoresistive microphone, UF2 piezoresistive proximity sensor, and Kulite and Endevco piezoresistive microphones.

Microphones	Rin (Ω)	Rout (Ω)	Normalized overall sensitivity (μV/Pa V) at 1 kHz
UF1 piezoresistive microphone[a]	578	579	0.56
UF2 piezoresistive proximity sensor[b]	9623	9634	0.71
Kulite MIC-093[c] (with temperature compensation module)	3102	1148	0.42
Endevco 8510B-1[d]	2057	1820	2.69

[a]Reference 3.
[b]Reference 4.
[c]Reference 6.
[d]Reference 5.

proximity sensor UF2) (Ref. 4)]. The measured input and output impedances and overall normalized sensitivities (μV/Pa V) of the microphones are listed in Table II. Note that the Kulite microphone includes a temperature compensation module. The noise PSD of the piezoresistive microphones are evaluated using the experimental noise setup and procedures developed in Sec. IV. However, for the Kulite piezoresistive microphone, instead of a dc bias voltage, an ac bias voltage is used to extract the noise PSD since the low frequency noise of the Kulite microphone is less than the setup noise at low frequency.

A. Electrical bias dependence of 1/*f* noise

1. Excess noise under zero bias condition

Under zero bias, the excess noise coincides with the thermal noise as shown in Fig. 8(a) for research prototype UF1. In this case, the excess noise corresponds to the excess noise of the low noise preamplifier (SRS 560). The extracted noise PSD for UF1 at equilibrium after the setup noise PSD is subtracted, Fig. 8(a), shows good agreement with the theoretical thermal noise corresponding to the microphone output resistance. The UF1 noise PSD at zero bias is shown starting around 90 Hz since at lower frequencies the preamplifier noise dominates the UF1 noise PSD, thus it is impossible to proceed to the subtraction of the UF1 noise PSD from the raw measured noise PSD. The other three microphones show similar equilibrium noise PSD characteristics.

2. Excess noise under bias condition

Piezoresistive microphone (UF1). The voltage noise PSD of UF1 at different bias voltages is shown in Fig. 8(b). One can observe from Fig. 8(b) that as the bias voltage increases, the magnitude of the excess noise at low frequencies increases. At zero bias, there is no excess noise, and the noise PSD coincides with the thermal noise predicted by the device resistance (Table II). However, at high frequencies regardless of the bias voltage, the noise PSD asymptotically coincides with the Johnson noise of the DUT. The corner frequency between $1/f$ and thermal noise is approximately 1×10^4 Hz for the UF1 microphone. The Hooge parameter is extracted from the measured PSD at a low frequency (where $1/f$ noise is dominant) as a function of external bias using Eq. (2). This may be accomplished when the geometry of the piezoresistor and the doping concentration profile are known which is the case for the two research prototype MEMS microphones (UF1 and UF2). The average Hooge parameter extracted at 12 Hz using the three external bias voltages shown in Fig. 8(b) is $(3.2\pm0.3)\times10^{-3}$ for UF1.

Piezoresistive proximity sensor (UF2). The PSD of the piezoresistive proximity sensor (UF2) is measured with the same equipments and settings as the piezoresistive microphone (UF1). However, a bias is applied to the substrate since the piezoresistors in UF2 are junction isolated while

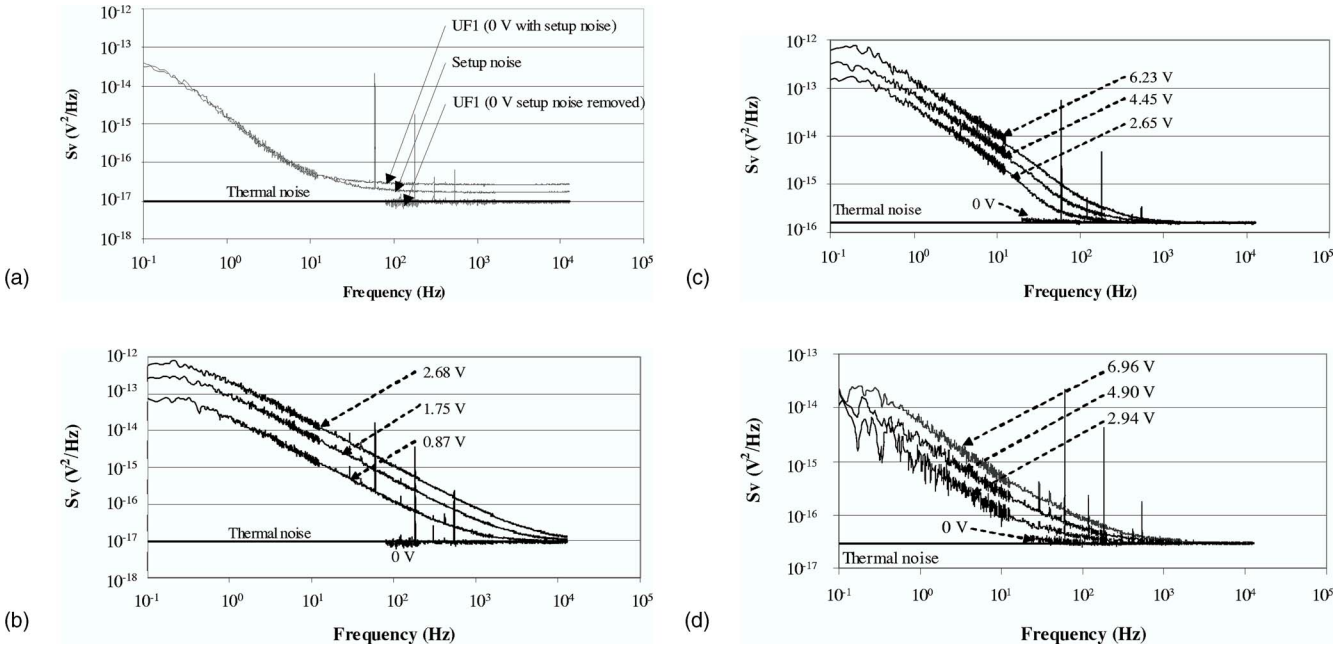


FIG. 8. (a) Power spectral density of UF piezoresistive microphone at zero bias voltage, (b) power spectral density of UF piezoresistive microphone at different bias voltages, (c) power spectral density of UF proximity sensor at different bias voltage (constant reverse bias of −0.5 V), and (d) power spectral density of Endevco microphone (Model 8510B-1) at different bias voltage.

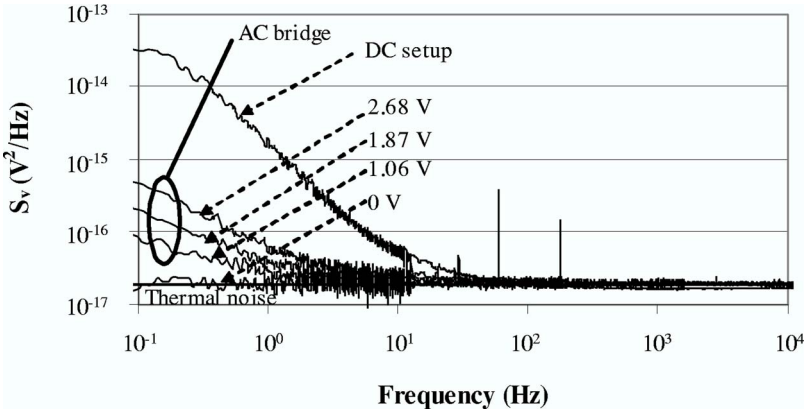


FIG. 9. Power spectral density of Kulite microphone (MIC-093) at different bias voltage compared to the dc setup noise.

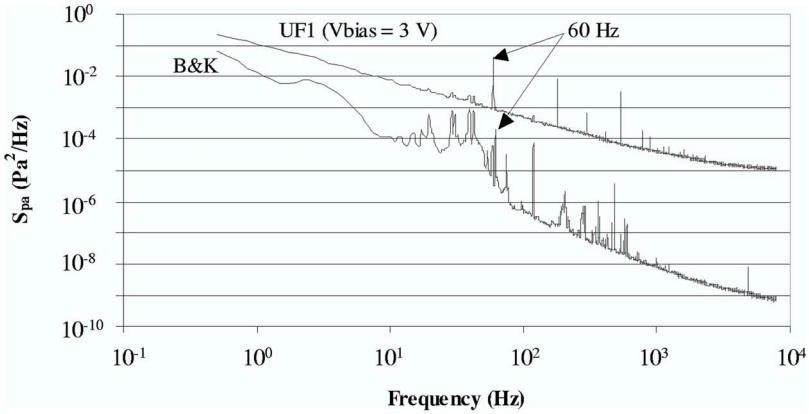
the piezoresistors in UF1 are dielectrically isolated. The substrate is biased at 6.71, 4.99, and 3.15 V when the sensor is biased, respectively, at 6.23, 4.45, and 2.65 V. This leads to a reverse bias voltage of about −0.5 V that corresponds to a relatively small isolation leakage current of 0.96 nA. Since the shot noise PSD is directly proportional to the leakage current $S_v = 2qI$, the contribution of shot noise is negligible. The voltage noise PSD curves of UF2 with different bias voltages are shown in Fig. 8(c) at a constant reverse bias of −0.5 V. A similar behavior is observed, that is, the excess noise increases as the bias voltage is increased. The corner frequency is approximately 100 Hz, and the Hooge parameter estimated using Eq. (2) and Fig. 8(c) is $(6.1 \pm 0.6) \times 10^{-5}$.

Endevco piezoresistive microphone. The noise measurements are conducted on the commercial Endevco 8510B-1 piezoresistive microphone using the same conditions as the previous microphones. The noise power spectral density obtained is shown in Fig. 8(d). The excess noise again shows a voltage bias dependence. The corner frequency is approximately 100 Hz.

Kulite piezoresistive microphone. The noise power spectral density of the Kulite microphone (MIC-093) is obtained using an ac bridge because the low frequency noise is less than the dc setup noise. The noise measurements are conducted with an ac bias frequency of 50 kHz and different peak voltage bias amplitudes of 1.06, 1.87, and 2.68 V. The noise PSD of the Kulite microphone is shown in Fig. 9. As seen in Fig. 9, the measured noise PSD at low frequency is less than the dc setup noise, hence an ac bridge is required. The corner frequency is 10 Hz.

In summary, all four piezoresistive MEMS microphones, both commercial and research prototypes, show a voltage bias dependence of the excess noise. The acoustic interference contribution to the measured electrical noise is discussed next.

(a)



(b)

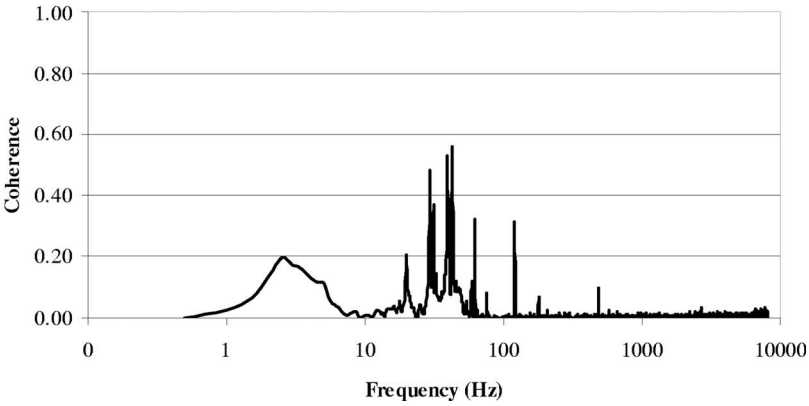


FIG. 10. (a) Power spectral density of the Bruel and Kjaer 4138 condenser microphone and of the UF piezoresistive microphone and (b) coherence function between the B&K 4138 and UF piezoresistive microphone.

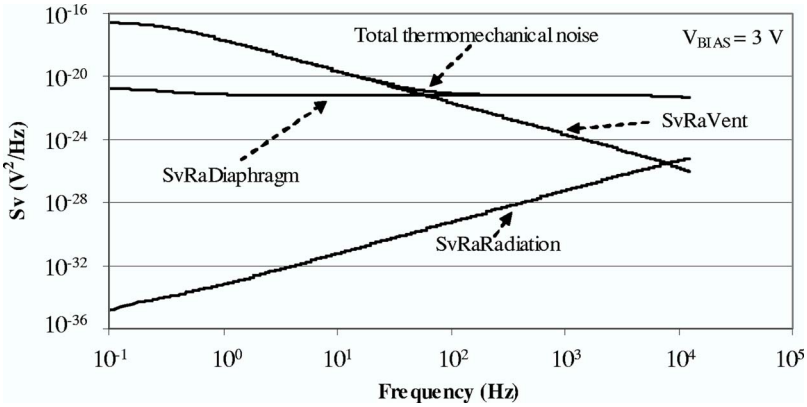


FIG. 11. Power spectral densities of noise sources due to the damping resistance, the radiation resistance, and the vent resistance.

3. Acoustic isolation test

The purpose of the acoustic isolation test is to quantify the acoustic sound pressure in the Faraday cage using a sensitive Bruel and Kjaer (B&K) microphone (model 4138) during the electrical noise measurement. The 1/8-in. Bruel and Kjaer 4138 condenser microphone at 1 kHz has a noise level of about 26 dB for a 3.16 Hz bin.[29] Figure 10(a) plots the measured acoustic pressure PSD for the Bruel and Kjaer 4138 condenser microphone and the UF1 microphone that were measured simultaneously. The measured acoustic PSD using the Bruel and Kjaer 4138 condenser microphone is approximately two orders of magnitude smaller than the measured acoustic PSD of the piezoresistive MEMS microphone UF1 in the triple nested Faraday cage. The coherence between the acoustic signal of the Bruel and Kjaer 4138 condenser microphone and the UF1 piezoresistive microphone (electrical noise in UF1 since the actual acoustic signal is two orders of magnitude smaller), shown in Fig. 10(b) is small for most frequencies. Although the acoustic isolation is not as elaborate as in Ref. 30, the acoustic isolation is sufficient such that the acoustic interference contribution to the measured electrical noise is negligible.

B. Membrane contribution to 1/*f* noise

To investigate whether the dominant source of the measured $1/f$ noise of the four piezoresistive microphones is mechanical or electrical in origin, two tests are applied. First, the output noise due to thermomechanical noise sources is computed using the noise equivalent circuit developed in Fig. 3(b) and compared to the measured output voltage noise PSD. Second, the dependence of the output noise on the membrane (fixed versus free) is measured as a function of applied voltage bias.

1. Thermomechanical contribution

The output voltage noise PSD due to thermomechanical fluctuation dissipation in the vent resistance, diaphragm damping resistance, and radiation resistance are calculated using the noise equivalent circuit shown in Fig. 3(b) and compared to the measured output voltage noise PSD. Since the detailed geometry is known for UF1, the thermomechanical contribution to the output voltage noise PSD is estimated for the piezoresistive microphone UF1, using values of the lumped element parameters given in Table I and the mechanical voltage sensitivity, S_{me} (2.9 μV/Pa at 1 kHz and 3 V dc bias) which is extracted from the measured overall sensitivity given in Table II at a specific frequency and bias voltage.

The output noise voltage PSD due to three thermomechanical noise courses (fluctuation dissipation in the vent resistance, diaphragm damping resistance, and radiation resistance) are calculated using the noise equivalent circuit in Fig. 3(b) and plotted in Fig. 11 for UF1. Among the thermomechanical noise sources, the output voltage noise PSD due to the vent resistance is seen to be dominant at low frequencies and frequency shaped by the noise equivalent circuit. The total estimated thermomechanical output voltage noise PSD is the sum of the three thermomechanical noise source contributions to the output voltage PSD. Although the ther-

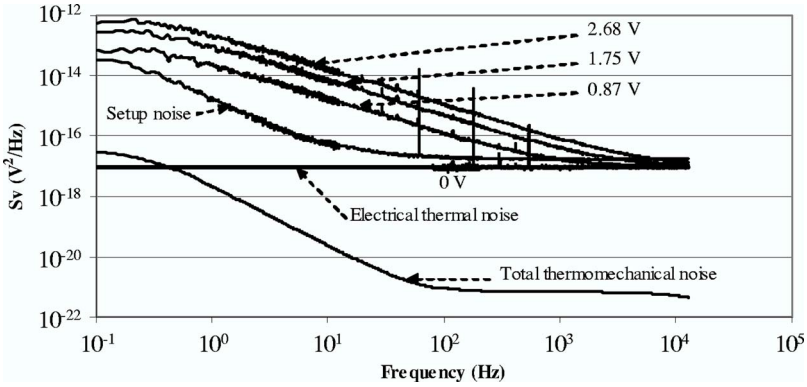


FIG. 12. Power spectral densities comparison of total thermomechanical noise to setup noise, electrical thermal noise, and electrical $1/f$ noise.

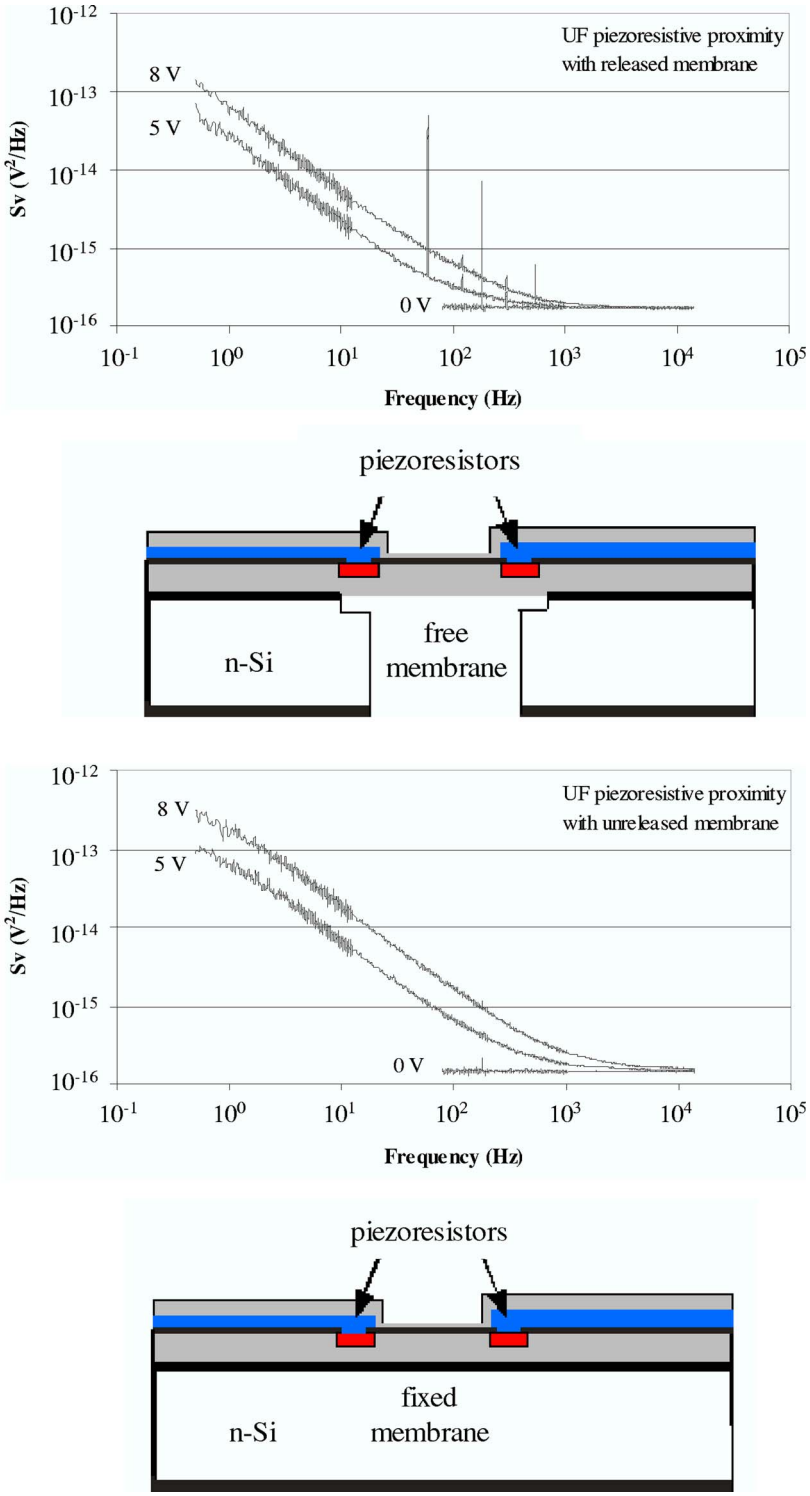


FIG. 13. (Color online) (a) Power spectral density of the UF piezoresistive proximity sensor with free diaphragm and (b) power spectral density of the UF piezoresistive proximity sensor with fixed diaphragm.

momechanical noise sources are not functions of frequency, their respective output voltage noise spectra are frequency shaped by their location in the noise equivalent circuit. When compared to the setup noise PSD, electrical thermal noise PSD, and the electrical $1/f$ noise PSD in Fig. 12, thermomechanical noise is seen to be negligible with respect to the aforementioned electrical noise sources for the biased piezoresistive MEMS microphones investigated in this study. This suggests that the dominant source of the measured $1/f$ noise of the four piezoresistive microphones is electrical in origin.

2. Membrane contribution

As a second test, the output voltage noise PSD is measured on two identical piezoresistive devices (UF2) except that one has a free membrane and the other has a fixed membrane. If the noise is primarily mechanical in origin, then a $1/f$ noise PSD characteristic must be present on the microphone with the free membrane but not on the fixed membrane, independent of electrical bias. Figure 13(a) shows the PSD of UF2 with a free diaphragm, and Fig. 13(b) shows the PSD of UF2 with a fixed diaphragm. Comparing Figs. 13(a)

and 13(b), excess noise is measured on both the mechanically free and fixed diaphragms. In addition, as the bias voltage increases, the PSD at low frequencies increases for both devices. The bias dependence and membrane independence of the output noise and negligible thermomechanical contribution indicate that the primary source of the excess noise observed in the four commercial and research-prototype piezoresistive MEMS microphones studied is electrical in origin.

VI. CONCLUSIONS

Using a validated noise setup, the noise PSD for four piezoresistive microphones has been investigated: two commercial piezoresistive microphones [Kulite (MIC-093) and Endevco (8510B-1)] and two research prototypes [piezoresistive microphone (UF1) (Ref. 3) and piezoresistive proximity sensor (UF2) (Ref. 4)]. Each of these measurements reveals a voltage bias dependence of the excess noise which is indicative of an electrical $1/f$ noise source. At zero bias, the DUT noise reduces to the thermal noise limit at equilibrium as expected. Using a noise equivalent circuit, the thermomechanical output voltage noise PSD due to the vent resistance, diaphragm damping resistance, and radiation resistance is found to be negligible compared to the setup noise PSD, electrical thermal noise PSD, and the electrical $1/f$ noise PSD. In addition, a bias-dependent low frequency noise was also observed on both fixed and free diaphragm piezoresistive microphones. The bias dependence and membrane independence of the output noise indicate that the primary source of the excess noise observed in the four commercial and research-prototype piezoresistive MEMS microphones studied is electrical in origin.

ACKNOWLEDGMENTS

The research was supported in part by NASA–Langley Research Center Grant No. NAG-1-2133, monitored by Dr. William H. Humphreys, Jr. and by the Office of Naval Research.

[1]R. Schellin and G. Hess, "A Silicon subminiature microphone based on piezoresistive polysilicon strain gauges," Sens. Actuators, A **32**, 555–559 (1992).
[2]M. Sheplak, K. S. Breuer, and M. A. Schmidt, "A wafer-bonded silicon-nitride membrane microphone with dielectrically-isolated single-crystal silicon piezoresistors," *Technical Digest, Solid-State Sensors and Actuators Workshop*, Hilton Head, SC (TRF, Cleveland Heights, Ohio, 1998), pp. 23–26.
[3]D. P. Arnold, T. Nishida, and M. Sheplak, "Piezoresistive microphone for aeroacoustic measurement," *Proceedings of 2001 ASME International Mechanical Engineering Congress and Exposition*, New York MEMS-23841 (ASME, New York, New York, 2001).
[4]V. Chandrasekaran, E. M. Chow, T. W. Kenny, T. Nishida, L. Cattafesta, B. V. Sankar, and M. Sheplak, "Thermoelastically actuated acoustic proximity sensor with integrated through-wafer interconnects," *Technical Digest, Solid-State Sensor and Actuator Workshop* (TRF, Cleveland Heights, Ohio, 2002), pp. 102–107.
[5]Endevco. Model 8510B-1, http://www.endevco.com, [cited; Available from: http://www.endevco.com (2005).
[6]Kulite. Model MIC-093 http://www.kulite.com, [cited; Available from: http://www.kulite.com(2005).
[7]R. R. Spencer, B. M. Fleischer, P. W. Barth, and J. B. Angell, "A theoretical study of transducer noise in piezoresistive and capacitive silicon pressure sensors," IEEE Trans. Electron Devices **35**(8), 1289–1297 (1988).
[8]T. B. Gabrielson, "Mechanical thermal noise in micromachined acoustic and vibration sensors," IEEE Trans. Electron Devices **40**(5), 903–909 (1993).
[9]H. Nyquist, "Thermal agitation of electric charge in conductors," Phys. Rev. **32**, 110–113 (1928).
[10]J. B. Johnson, "Thermal agitation of electricity in conductors," Phys. Rev. **32**, 97–109 (1928).
[11]F. N. Hooge, "$1/f$ Noise is no surface effect," Phys. Lett. A **29**(3), 139–140 (1969).
[12]A. McWhorter, "$1/f$ Noise and germanium surface properties," *Semiconductor Surface Physics* (University of Pennsylvania, Philadelphia, 1957), pp. 207–228.
[13]A. J. Zuckerwar, T. R. Kuhn, and R. M. Serbyn, "Background noise in piezoresistive, electret condenser, and ceramic microphones," J. Acoust. Soc. Am. **113**(6), 2947–3437 (2003).
[14]A. J. Zuckerwar and K. C. T. Ngo, "Measured $1/f$ noise in the membrane motion of condenser microphones," J. Acoust. Soc. Am. **95**(3), 1419–1425 (1994).
[15]O. Hansen and A. Boisen, "Noise in piezoresistive atomic force microscopy," Nanotechnology **10**, 51–60 (1999).
[16]J. A. Harley and T. W. Kenny, "High-sensitivity piezoresistive cantilevers under 1000 A thick," Appl. Phys. Lett. **75**(2), 289–291 (1999).
[17]A. Partridge, K. Reynolds, B. W. Chui, E. M. Chow, A. M. Fitzgerald, L. Zhang, N. I. Maluf, and T. W. Kenny, "A high-performance planar piezoresistive accelerometer," J. Microelectromech. Syst. **9**(1), 8–56 (2000).
[18]F. N. Hooge, "$1/f$ Noise," Physica B **83**, 14–23 (1976).
[19]A. Van der Ziel, "Noise in solid-state devices and circuits" (Wiley, New York, 1986).
[20]H. L. Chau and K. D. Wise, "Noise due to Brownian motion in ultrasensitive solid-state pressure sensors," IEEE Trans. Electron Devices **ED-34**(4), 859–864 (1987).
[21]O. N. Tufte and D. Long, "Recent developments in semiconductor piezoresistive devices," Solid-State Electron. **6**, 323–338 (1963).
[22]M. Sheplak and J. Dugundji, "Large deflections of clamped circular plates under initial tension and transitions to membrane behavior," J. Appl. Mech. **65**, 107–115 (1998).
[23]R. Saini, S. Bhardwaj, T. Nishida, and M. Sheplak, "Scaling relations for piezoresistive microphones," *Proceedings of ASME IMECE 2000, International Mechanical Engineering Congress and Exposition*, MEMS-Vol. 2, Orlando, FL (ASME, New York, New York, 2000), pp. 241–248.
[24]M. Rossi, "Acoustics and Electroacoustics" (Artech House, Norwood, MA, 1988).
[25]J. Bryzek, "Modeling performance of piezoresistive pressure sensors," *Technical Digest, Solid-State Sensors and Actuators* (TRF, Cleveland Heights, Ohio, 1985), pp. 168–173.
[26]R. Saini, "Design of a MEMS based piezoresistive microphone, M.S. thesis, Aerospace Engineering, Mechanics, and Engineering Science, University of Florida, Gainesville, 2001.
[27]C. D. Motchenbacher and J. A. Connelly, *Low-Noise Electronic System Design* (Wiley, New York, 1993).
[28]J. H. J. Lorteije and A. M. H. Hoppenbrouwers, "Amplitude modulation by $1/f$ noise in resistors results in 1/Defta f noise," Philips Res. Rep. **26**, 29–39 (1971).
[29]Bruel and Kjaer, Product Data, http://www.bksv.com/pdf/Bp0100.pdf
[30]K. C. T. Ngo and A. J. Zuckerwar, "Acoustic isolation vessel for measurement of the background noise in microphones," J. Acoust. Soc. Am. **93**(5), 2974–2980 (1993).

Adaptive wave field synthesis with independent radiation mode control for active sound field reproduction: Theory[a)]

Philippe-Aubert Gauthier[b)] and Alain Berry
Groupe d'Acoustique de l'Université de Sherbrooke, Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke, Québec, Canada J1K 2R1

(Received 11 July 2005; revised 16 February 2006; accepted 19 February 2006)

Sound field reproduction finds applications in music or audio reproduction and experimental acoustics. For audio applications, sound field reproduction can be used to artificially reproduce the spatial character of natural hearing. The general objective is then to reproduce a sound field in a real reproduction environment. Wave field synthesis (WFS) is a known open-loop technology which assumes that the reproduction environment is anechoic. For classical WFS, the room response thus reduces the quality of the physical sound field reproduction. In this paper, adaptive wave field synthesis (AWFS) is analytically investigated as an adaptive sound field reproduction system combining WFS and active control with a limited number of reproduction error sensors to compensate the response of the listening environment. The primary point of this paper is the definition of AWFS. Therefore, the fundamental behavior of AWFS is illustrated by analytical considerations and simple free-field simulation results. As demonstrated, AWFS is fundamentally related to WFS and "Ambisonics." The paper introduces independent adaptive control of sound field reproduction on the basis of radiation modes, via the singular value decomposition of the transfer impedance matrix. Possible practical independent control of radiation modes for AWFS is discussed. [DOI: 10.1121/1.2186514]

PACS number(s): 43.38.Md, 43.60.Tj, 43.50.Ki [NX] Pages: 2721–2737

I. INTRODUCTION

Sound field reproduction finds applications in music or audio reproduction (multisensory immersion for teaching, concerts, etc. (Woszczyk *et al.*, 2005), and experimental acoustics and psychoacoustics. For audio applications, sound field reproduction can be used to artificially reproduce the spatial character of natural hearing. Such a challenge has always fascinated both inventors and sound artists (Poulin, 1957; Malham and Myatt, 1995; Davis, 2003; AES Staff Writer, 2005). The general objective is to reproduce a sound field in a real reproduction environment. [See AES Staff Writer (2005) for a review of spatial and multichannel audio.]

Wave field synthesis (WFS) is an open-loop sound field reproduction technology which assumes that the reproduction environment is anechoic (Berkhout, 1988, 1993; Verheijen, 1997). For objective sound field reproduction, classical WFS thus does not perform well in a real reproduction space (Gauthier *et al.*, 2005b). Typical physical inaccuracies are caused by the room resonant response at low frequencies along with reflections and reverberation at higher frequencies. Adaptive wave field synthesis (AWFS) is investigated as a potential solution for room compensation with a limited number of acoustic sensors. The primary point of this paper is the theoretical analysis of AWFS. Therefore, the fundamental behavior of AWFS is illustrated by analytical considerations and simple free-field simulations. Other research has been devoted to in-room sound field reproduction or local sound reproduction using various approaches often derived from adaptive control (Kirkeby and Nelson, 1993; Nelson, 1994; Asano and Swanson, 1995; Uto and Hamada, 1995; Kirkeby *et al.*, 1996; Garas, 1999; Ise, 1999; Santillań, 2001; Corteel *et al.*, 2002; Nelson, 2002; Spors *et al.*, 2003; Epain *et al.*, 2004; Gauthier *et al.*, 2005a; Schobben and Aarts, 2005; Spors *et al.*, 2005; Fuster *et al.*, 2005).

This paper theoretically introduces AWFS which is a combination of optimal control and open-loop wave field synthesis designed for active sound field reproduction. The adaptive nature of AWFS comes from the use of reproduction error sensors that observe the system performance and adapt the input of the reproduction sources. What differentiates AWFS from classical active control is that the regularization is not included as a quadratic effort-penalty term in the cost function, but as a quadratic departure penalty from the WFS solution.

Although AWFS can be applied to any source and sensor configuration, numerical results are presented here for a small, central and compact sensor array that provides WFS compensation without occupying the reproduction space with a large and impractical sensor array. As discussed in previous reports (Gauthier *et al.*, 2005a), such a reduced sensor array is used as a practical compromise. However, the use of a limited number of reproduction sensors results in interesting properties of AWFS.

A compact sensor array is investigated here as a practical alternative to larger or circular sensors arrays that have

[a)]Portions of this work were presented in "Wave field synthesis, adaptive wave field synthesis and ambisonics using decentralized transformed control: Potential applications to sound field reproduction and active noise control," 150th ASA meeting, Minneapolis MN, October 2005.

[b)]Electronic mail: philippe_aubert_gauthier@hotmail.com

been previously proposed. They were suggested for applications such as sound field recording (Hulsebos *et al.*, 2003; Poletti, 2005). WFS room compensation (Spors *et al.*, 2003, 2005); and active sound field reproduction (Betlehem and Abhayapala, 2005).

As introduced by Spors *et al.* (2003; 2005), one approach to active compensation in WFS is to work in a transformed domain which can potentially uncouple the problem. In the present paper, it is shown that AWFS can be achieved in a transformed domain based on the source modes and pressure modes derived from the singular value decomposition (SVD) of the source-sensor transfer impedance matrix. SVD has found various applications in acoustics (Nelson, 2001; Nelson and Kahana, 2001; Nelson, 2002; Williams *et al.*, 2003; Bai and Kwuen-Yieng, 2005).

Recently, Bai and Elliott (2004) used a plant transfer matrix SVD to define an uncoupled adaptive system. This was motivated by the need to accelerate the convergence of standard LMS adaptive multichannel algorithms and the results are illustrated in the context of binaural sound field reproduction with cross-talk cancelation. The authors used two complementary methods: (1) prewhitening of the reference and (2) plant uncoupling via SVD of the transfer impedance matrix in the frequency domain and time-domain realization of the resulting uncoupled multichannel system. This proved to accelerate the LMS convergence rate at the cost of a processing delay.

Section II of the paper presents the radiation, WFS and AWFS models. Section III introduces a first interpretation of the AWFS mechanism. Numerical simulations are introduced in Sec. IV. Section V expands the theoretical behavior of AWFS on the basis of the SVD of the transfer impedance matrix including room effects. Section VI specifically shows the link between AWFS and closed-loop "Ambisonics." Practical considerations for independent radiation mode control are discussed in Sec. VII.

II. RADIATION, WFS AND CONTROL MODELS FOR AWFS

Let an active sound field reproduction system be represented by L monopole reproduction sources located at $\{\mathbf{x}_1^{(l)} \cdots \mathbf{x}_L^{(l)}\}$ and M sound pressure sensors located at $\{\mathbf{x}_1^{(m)} \cdots \mathbf{x}_M^{(m)}\}$ where $\mathbf{x}$ are position vectors expressed in rectangular coordinates (see Fig. 1 for an explanatory sketch). In the frequency domain, the sound pressure field radiated by the reproduction sources is defined as follows at radial frequency ω [rad/s]

$$\mathbf{p}^{(\mathrm{rep})}(\omega) = \mathbf{Z}(\omega)\mathbf{q}(\omega), \tag{1}$$

where $\mathbf{p}^{(\mathrm{rep})}(\omega)$ is a M coefficients column vector comprising the reproduced complex sound pressures [Pa] at sensor locations, $\mathbf{Z}(\omega)$ is a $M \times L$ transfer impedance matrix and $\mathbf{q}(\omega)$ [m^3/s] is the L coefficients reproduction source strength column vector. The frequency dependance may be omitted. All components in Eq. (1) are complex. For practical reasons, it is assumed that $M < L$ since a convenient system should not introduce a large number of sensors in the reproduction space. For free-field radiation, the

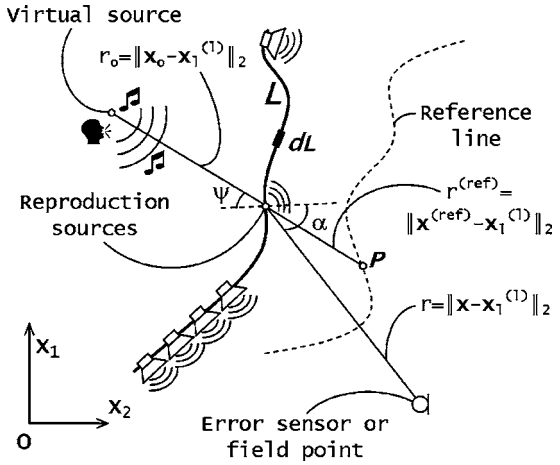


FIG. 1. Symbol convention for the WFS operators definition. The virtual source is located in $\mathbf{x}_o$. The reproduction source l is located in $\mathbf{x}_l^{(l)}$. $\mathbf{x}^{(\mathrm{ref})}$ describes points which belong to the reference line. $\mathbf{x}$ describes any field point and $\mathbf{x}_m^{(m)}$ is the m sensor position. L is the reproduction source line, the virtual source is on the left of the source line and the reproduction space is on the right of the source line. All sources and sensors are located in the x_1-x_2 plane.

transfer impedance matrix coefficients are defined by the free-field Green's function $G(\mathbf{x}|\mathbf{x}_l^{(l)})$ (Pierce, 1991; Nelson and Elliott, 1992)

$$Z_{ml} = -j\omega\rho G(\mathbf{x}_m^{(m)}|\mathbf{x}_l^{(l)}) \tag{2}$$

with

$$G(\mathbf{x}|\mathbf{x}_l^{(l)}) = \frac{e^{jk|\mathbf{x}-\mathbf{x}_l^{(l)}|}}{4\pi|\mathbf{x}-\mathbf{x}_l^{(l)}|}, \tag{3}$$

where $\mathbf{x}$ [m] is the field point, $\mathbf{x}_l^{(l)}$ [m] the l source position, $k=\omega/c$ [rad/m] the wave number (with sound speed $c=331$ m/s) and ρ [kg/m^3] the air density ($\rho=1.18$ kg/m^3). The time convention is $e^{-j\omega t}$ [$p(\mathbf{x},t)=p(\mathbf{x},\omega)e^{-j\omega t}$]. Without loss of generality, the following analytical developments can also be applied to Green's functions defined for semi-infinite or finite spaces. Free-field Green's functions are used here to illustrate general results of the fundamental AWFS behavior, through simple simulations. With these simulations, our primary objective is to clarify and illustrate the theoretical properties of AWFS.

The reproduction goal is defined as a target pressure field produced by a virtual source located at $\mathbf{x}_o$ [m]

$$p^{(\mathrm{im})}(\mathbf{x},\omega) = A(\omega)\frac{e^{jk|\mathbf{x}-\mathbf{x}_o|}}{|\mathbf{x}-\mathbf{x}_o|}, \tag{4}$$

where k is the wave number and $A(\omega)$ is the virtual source monopole amplitude, which is set to unity. Such a specific target wave field is used for simulation purposes. It could be replaced by any other target field definition (plane or directive fields, for example) and the remaining analytical developments would still be valid. Indeed, a part of the main points presented in this paper relies on analytical developments where no assumptions about the nature of $p^{(\mathrm{im})}(\mathbf{x},\omega)$ or $G(\mathbf{x}|\mathbf{x}_l^{(l)})$ are included. The virtual source position is arbitrary. A virtual source "in front" of the source array (i.e., on the same side as the listening space) is often cited

as an interesting practical possibility of sound field reproduction. In this paper, virtual sources are always located "behind" the array (with respect to the listening space).

A. WFS operators

WFS is an open-loop technology for sound field reproduction which assumes a free-field reproduction space (Berkhout, 1993; Verheijen, 1997). The WFS operators are defined from the Kirchhoff-Helmholtz integral theorem. These operators are then used to drive a reproduction source array according to the virtual source definition and to the reproduction source array geometry. In free-field conditions, using these WFS operators along with the source array would then approximately reproduce the target field produced by the virtual source. The WFS problem is usually limited to reproduction in the horizontal plane by means of a finite number of discrete sources using appropriate simplifications of the integral formulation (Verheijen, 1997; Nicol and Emerit, 1999; Spors *et al.*, 2005). A monopole array for WFS would be fed by the normal virtual sound pressure gradient evaluated along the reproduction source line. This is obtained using the monopole radiation part of the Kirchhoff-Helmholtz integral with the stationary phase approximation (Junger, 1993; Verheijen, 1997). Both the stationary phase approximation and the source line discretization introduce reproduction errors. Moreover, usage of a truncated reproduction array or piecewise linear array in practical WFS implementation causes finite aperture and corner effects which produce reproduction errors (Verheijen, 1997). The above WFS reproduction errors are readily observable in free field but they are masked in rooms by the comparatively large room effects. [This was observed by time- and frequency-domain numerical simulations in comparison with objective WFS sound field reproduction measurements in a real studio (Gauthier *et al.*, 2005b).]

On the one hand, the inherent free-field assumption of WFS reduces the objective accuracy of the sound field reproduction in real situations, thus suggesting the need for a closed-loop compensation (Spors *et al.*, 2003; Gauthier *et al.*, 2005b). On the other hand, WFS provides satisfactory results for sound localization by the human auditory system since the WFS direct field approaches the target field (Gauthier *et al.*, 2005b). (It is possible to decompose the WFS reproduced field as the superposition of a direct field and a field which is reverberated by the room walls and furniture surfaces.) Direct-field reproduction is critical for accurate subjective localization (Blauert, 1999). Therefore, the active compensation introduced by AWFS includes a departure penalty from the WFS operators, thus motivating the adaptive WFS terminology. The WFS operators are defined as follows (Verheijen, 1997):

$$Q_{\mathrm{WFS}}(\mathbf{x}_l^{(l)},\omega) = -A(\omega)j\sqrt{\frac{jk}{2\pi}}\cos\Psi\frac{e^{jkr_o}}{\sqrt{r_o}} \times\sqrt{r^{(\mathrm{ref})}/(r^{(\mathrm{ref})}+r_o)}\Delta_l \tag{5}$$

and

$$q_{\mathrm{WFS}}(\mathbf{x}_l^{(l)},\omega) = -\frac{4\pi Q_{\mathrm{WFS}}(\mathbf{x}_l^{(l)},\omega)}{j\omega\rho}, \tag{6}$$

where $A(\omega)$ is the virtual source monopole amplitude, Ψ [rad] is the angle between the virtual source and the normal to the reproduction line at the reproduction source position $\mathbf{x}_l^{(l)}$, $r_o=\|\mathbf{x}_o-\mathbf{x}_l^{(l)}\|_2$ is the distance [m] between the virtual source (in $\mathbf{x}_o$) and the reproduction source (in $\mathbf{x}_l^{(l)}$) and $r^{(\mathrm{ref})}=\|\mathbf{x}^{(\mathrm{ref})}-\mathbf{x}_l^{(l)}\|_2$ is the distance [m] between the reproduction source and the reference line along the line r_o. These geometric parameters are illustrated in Fig. 1. In Eq. (5), Δ_l [m] is the reproduction source separation ($\Delta_l=\|\mathbf{x}_l^{(l)}-\mathbf{x}_{l+1}^{(l)})\|_2$). Equation (5) expresses the WFS monopole source amplitude $Q_{\mathrm{WFS}}(\mathbf{x}_l^{(l)},\omega)$ to produce $p^{(\mathrm{im})}(\mathbf{x},\omega)$ while Eq. (6) transforms WFS monopole amplitudes in source strength amplitudes $q_{\mathrm{WFS}}(\mathbf{x}_l^{(l)},\omega)$ (Pierce, 1991).

The stationary phase approximation in WFS also leads to the definition of a "reference line" (Verheijen, 1997) on which the amplitude of the reproduction error should theoretically be zero in comparison with points outside the reference line, again for free field. Indeed, the stationary phase approximation in WFS implies that the reproduced sound field would be effectively reproduced only on a line in front of the reproduction source array. For WFS, this line is arbitrarily chosen and this corresponds to the WFS reference line. This line is shown in Fig. 1. In this paper, the reproduction source line L is defined as a circle, and the reference line is also defined as a circle, centered on the virtual source position $\mathbf{x}_o$ and passing through the source array center. The choice of a given reference line definition is not relevant to the AWFS fundamental idea, but it remains an important choice for the WFS definition (for numerical results reproducibility) since the reference line influences the WFS operators of Eq. (5).

For WFS, a reproduction source is active if $|\Psi|<\pi/2$ radian (see Fig. 1 for Ψ definition). An active source implies that $Q_{\mathrm{WFS}}(\mathbf{x}_l^{(l)},\omega)\neq 0$. Other sources, with $|\Psi|\geqslant\pi/2$, are not active [artificially forced to $Q_{\mathrm{WFS}}(\mathbf{x}_l^{(l)},\omega)=0$]. This highlights the fact that WFS solution implies that only a part of the reproduction source array is active for a given virtual source. A spatial weighting window is also used to progressively reduce the WFS strength amplitudes. This is applied to the reproduction sources closer to the end of the active part of the source WFS array. Such spatial windowing limits the finite aperture effects of the reproduction array. Here, the spatial weighting window is a rectangular window with half Hanning windows at its ends. The half Harming window extends over 0.5 m (Verheijen, 1997).

B. Definition of AWFS

Using the radiation model of Eqs. (1)–(3) with the target pressure field of Eq. (4), AWFS (adaptive wave field synthesis) is introduced as an optimal control approach with this cost function

$$J_{\mathrm{AWFS}} = \mathbf{e}^H\mathbf{e} + \lambda(\mathbf{q}-\mathbf{q}_{\mathrm{WFS}})^H(\mathbf{q}-\mathbf{q}_{\mathrm{WFS}}) \tag{7}$$

which must be minimized with λ being a real positive scalar. Superscript H denotes Hermitian transposition. Equation (7)

introduces the reproduction error vector $\mathbf{e}=\mathbf{p}^{(\mathrm{im})}-\mathbf{p}^{(\mathrm{rep})}=\mathbf{p}^{(\mathrm{im})}-\mathbf{Z}\mathbf{q}$, where $\mathbf{p}^{(\mathrm{im})}$ is the complex column vector of the target sound pressure at the M sensor locations, defined by Eq. (4) with $\mathbf{x}_m^{(m)}$ replacing $\mathbf{x}$. The first term in Eq. (7) represents the summation of the quadratic reproduction errors with respect to the target reproduced sound field, observed at the sensor locations. The second term in Eq. (7) is weighted by the regularization parameter λ and it represents a quadratic departure penalty of the solution $\mathbf{q}$ with respect to the fixed WFS solution $\mathbf{q}_{\mathrm{WFS}}$.

J_{AWFS} is a quadratic function of $\mathbf{q}$,

$$J_{\mathrm{AWFS}}=\mathbf{q}^H[\mathbf{Z}^H\mathbf{Z}+\lambda\mathbf{I}]\mathbf{q}-\mathbf{q}^H[\mathbf{Z}^H\mathbf{p}^{(\mathrm{im})}+\lambda\mathbf{q}_{\mathrm{WFS}}]-[\mathbf{p}^{(\mathrm{im})^H}\mathbf{Z}+\lambda\mathbf{q}_{\mathrm{WFS}}^H]\mathbf{q}+[\mathbf{p}^{(\mathrm{im})^H}\mathbf{p}^{(\mathrm{im})}+\lambda\mathbf{q}_{\mathrm{WFS}}^H\mathbf{q}_{\mathrm{WFS}}], \quad (8)$$

where $\mathbf{I}$ is the $L\times L$ identity matrix.

Although the reproduction system of Eq. (1) is underdetermined ($M<L$) (Nelson and Elliott, 1992), the WFS departure penalty term, weighted by λ, ensures that there will be a unique minimum of Eq. (7). That is, the Hessian matrix ($[\mathbf{Z}^H\mathbf{Z}+\lambda\mathbf{I}]$) of the quadratic Eq. (8) is positive definite provided that $\lambda>0$. This optimum source strength vector minimizes the cost function (Nelson and Elliott, 1992)

$$\mathbf{q}^{(\mathrm{opt})}=[\mathbf{Z}^H\mathbf{Z}+\lambda\mathbf{I}]^{-1}[\mathbf{Z}^H\mathbf{p}^{(\mathrm{im})}+\lambda\mathbf{q}_{\mathrm{WFS}}]. \quad (9)$$

When $\lambda\to\infty$ the solution reduces to the WFS operators $\mathbf{q}_{\mathrm{WFS}}$, when $\lambda\to 0$ the solution is the optimal control solution that minimizes the reproduction error in a least-mean-square (LMS) sense. These two situations are the limiting cases of AWFS.

III. INTERPRETATION OF AWFS

Equations (7) and (9) lead to insightful observations about the behavior of AWFS. It first should be noted that the second term of the cost function (7) includes a standard effort penalty obtained by $\lambda\mathbf{q}^H\mathbf{q}$ and a projection term $-2\lambda\mathbf{q}_{\mathrm{WFS}}^H\mathbf{q}$. The first term acts like a standard regularization while the second introduces a contribution of $\mathbf{q}_{\mathrm{WFS}}$ in the solution.

Rewriting Eq. (9) and extracting the $\mathbf{q}_{\mathrm{WFS}}$ contribution, one finds

$$\mathbf{q}^{(\mathrm{opt})}=[\mathbf{Z}^H\mathbf{Z}+\lambda\mathbf{I}]^{-1}[\mathbf{Z}^H\mathbf{e}_{\mathrm{WFS}}]+\mathbf{q}_{\mathrm{WFS}}, \quad (10)$$

where $\mathbf{e}_{\mathrm{WFS}}=\mathbf{p}^{(\mathrm{im})}-\mathbf{Z}\mathbf{q}_{\mathrm{WFS}}$ is the reproduction error of WFS at the sensors. This potentially includes the inherent WFS approximation errors reviewed in Sec. II A, along with the room effects.

Equation (10) illustrates that the optimal solution $\mathbf{q}^{(\mathrm{opt})}$ is the superposition of the optimal correction $[\mathbf{Z}^H\mathbf{Z}+\lambda\mathbf{I}]^{-1}[\mathbf{Z}^H\mathbf{e}_{\mathrm{WFS}}]$ on the fixed WFS solution $\mathbf{q}_{\mathrm{WFS}}$. The optimal correction matches a standard optimal solution for active control (including effort penalization), the difference is that the target is the WFS reproduction error $\mathbf{e}_{\mathrm{WFS}}$ replacing the target wave field $\mathbf{p}^{(\mathrm{im})}$. Although the WFS contribution $\mathbf{q}_{\mathrm{WFS}}$ activates only a limited set of sources, the optimal contribution ($[\mathbf{Z}^H\mathbf{Z}+\lambda\mathbf{I}]^{-1}[\mathbf{Z}^H\mathbf{e}_{\mathrm{WFS}}]$) will activate all the reproduction sources.

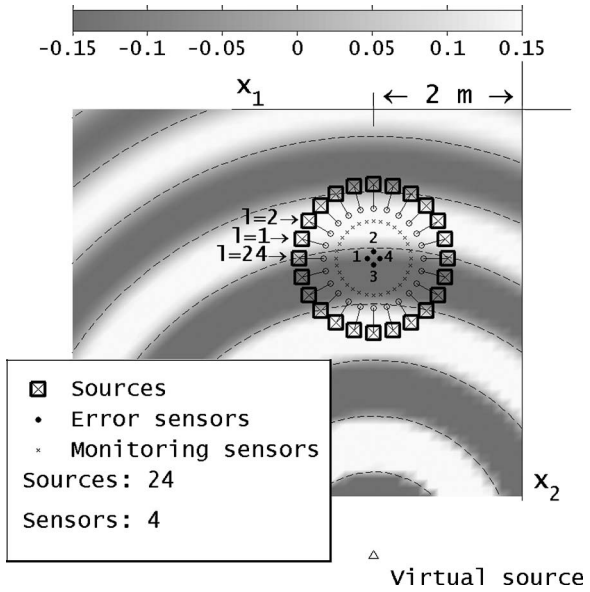


FIG. 2. System configuration in free field and real part of the target wave field at 220 Hz. Dashed lines represent zero-pressure contours of the target pressure field $p^{(\mathrm{im})}(\mathbf{x},w)$. —○, the normals to the source line. Δ, virtual source $\mathbf{x}_o=[2,6,0]$ meters.

IV. SIMULATION RESULTS

Simulation results are reported for a relatively small and compact hypothetical error sensor array prototype. (Such a prototype is currently under development for practical AWFS.) Simulations are performed in free field to illustrate the fundamental AWFS mechanisms.

The system configuration and the target pressure field $\mathbf{p}^{(\mathrm{im})}$ (at 220 Hz) are shown in Fig. 2. As illustrated, the system includes 24 sources, four error sensors, and 32 monitoring sensors used to quantify the reproduction quality outside the control region defined by the error sensor locations. The error sensors are separated by 0.175 m along x_1 and x_2. The source array has a diameter of 2 m and the source separation is 0.2611 m. The monitoring sensor array has a diameter of 1 m and the monitoring sensor separation is 0.098 m. All the arrays are centered around $x_1=x_2=2$ m. The virtual source is located at $\mathbf{x}_o=[2,6,0]$ meters.

Simulation results are globally quantified by E_{LS}

$$E_{\mathrm{LS}}=\frac{\mathbf{e}^H\mathbf{e}}{\mathbf{p}^{(\mathrm{im})^H}\mathbf{p}^{(\mathrm{im})}}. \quad (11)$$

E_{LS} is the global residual quadratic reproduction error normalized by the summation of the quadratic target pressures at the error sensor locations. $E_{\mathrm{LS}}=0$ characterizes perfect reproduction and $E_{\mathrm{LS}}=1$ corresponds to an error as large as the target field.

E_{LS} can also be computed for the M_m monitoring sensors using the reproduction errors at the monitoring sensor locations $\{\mathbf{x}_1^{(\mathrm{mon})}\cdots\mathbf{x}_{M_m}^{(\mathrm{mon})}\}$. The monitoring sensors are introduced to monitor the reproduction error outside the error sensor locations.

Results in terms of E_{LS} are shown in Figs. 3 and 4 for the configuration of Fig. 2 and others. Configuration (b) corresponds to Fig. 2. Configuration (a) corresponds to a unique

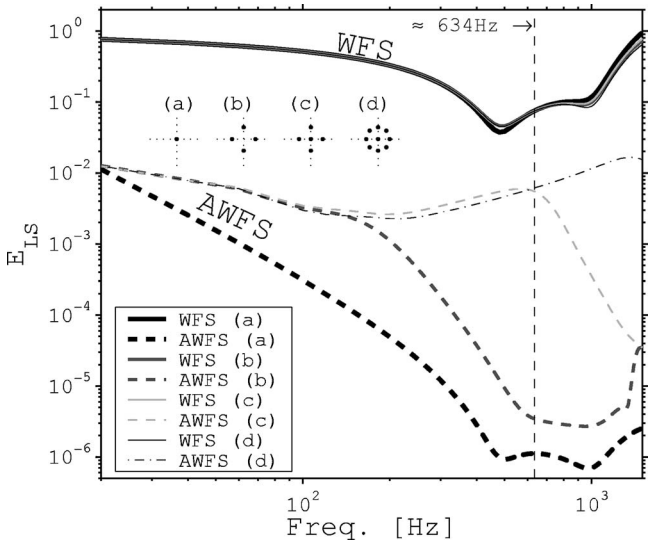


FIG. 3. E_{LS} at the error sensors for various approaches (WFS and AWFS) and error sensor configurations (shown as dot plots). Configuration (b) corresponds to Fig. 2; (a) corresponds to a unique sensor located at the center of the source array; (c) corresponds to configuration (b) plus an additional sensor at the center of the source array; (d) includes nine sensors approaching two tesseral quadrupolelike dispositions (45 degrees apart) plus a centered sensor.

sensor located at the center of the source array (at $x_1=x_2=2$ m). Configuration (c) is similar to (b) plus an additional sensor at the center of the source array. Configuration (d) includes nine sensors forming two tesseral quadrupoles [$\pi/4$ radian apart, the separation distance is the same as in case (b)] plus a centered sensor. In Fig. 3, WFS produces slightly different E_{LS} because the error sensor locations change: The WFS reproduced field is the same for the four configurations. In each case, the regularization parameter λ for AWFS was set to different values to take into account that an increased number of sensors reduces the importance of the penalization term in J_{AWFS} with a fixed λ. Case (a) was simulated with $\lambda=50$ while the other values were (b) $\lambda=200$, (c) $\lambda=250$, and (d) $\lambda=450$. Here, the λ values are not optimally chosen

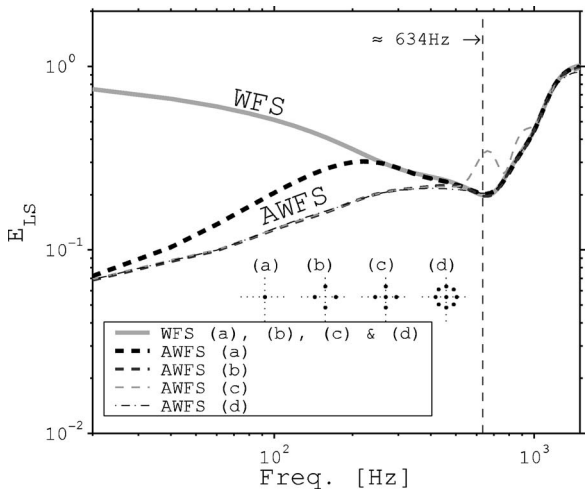


FIG. 4. E_{LS} at the monitoring sensors for the cases of Fig. 3 (the error sensor configurations are shown as dot plots).

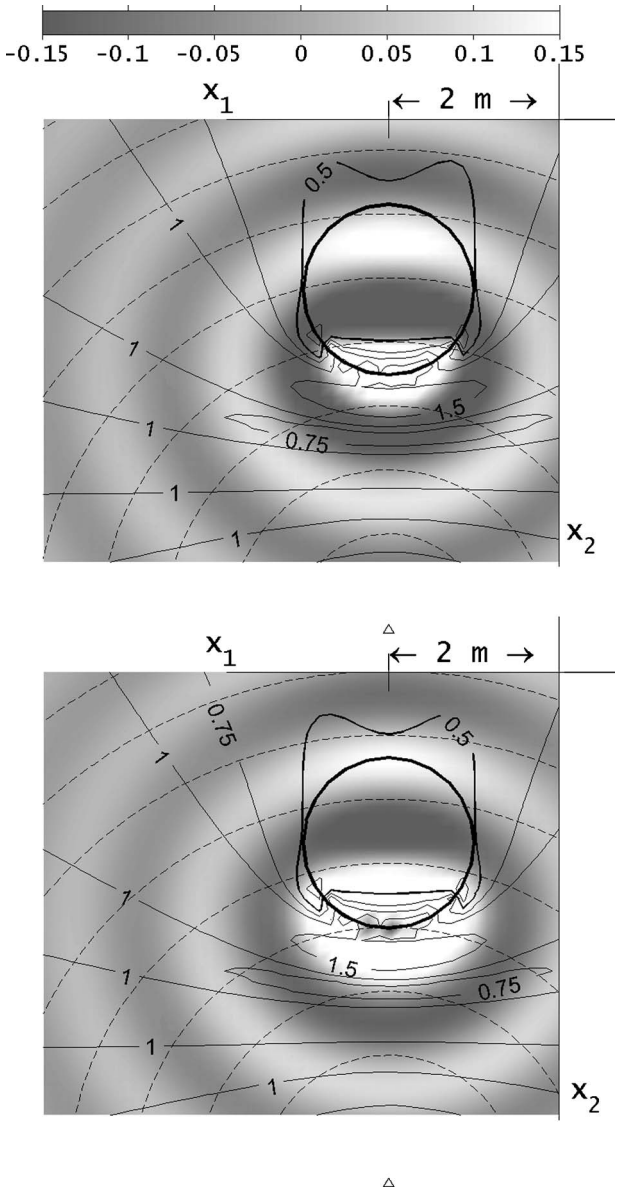


FIG. 5. Reproduced sound field at 220 Hz in free field WFS (real part on top, imaginary part on bottom). Dashed lines show zero-pressure contours of the target pressure field. Local normalized quadratic residual error is represented as continuous line contour plots. Reproduction sources are replaced by a thick source line for clarity. Δ, virtual source.

[with respect to inverse problem criteria such as the L-curve (Hansen, 1998; Nelson, 2001)], they were set to an arbitrary value multiplied by the number of sensors.

As shown in Fig. 3, E_{LS} at the error sensors is larger with WFS than AWFS. Also, the reproduction quality at the monitoring sensors with AWFS converges to the WFS results above approximately 634 Hz (Fig. 4). This frequency corresponds to the WFS spatial aliasing frequency which is defined in a first approximation by the source separation distance (Verheijen, 1997). WFS requires at least two reproduction sources per wavelength to avoid spatial aliasing.

Reproduced sound fields at 220 Hz for WFS, AWFS and reproduction error regularized minimization [$\lambda=50$ and $\mathbf{q}_{WFS}=\mathbf{0}$ in Eq. (9), that is optimal sound field reproduction without the WFS *a priori* solution] are shown in Figs. 5–7 for the configuration (b) shown in Fig. 2. $\lambda=50$ for all these

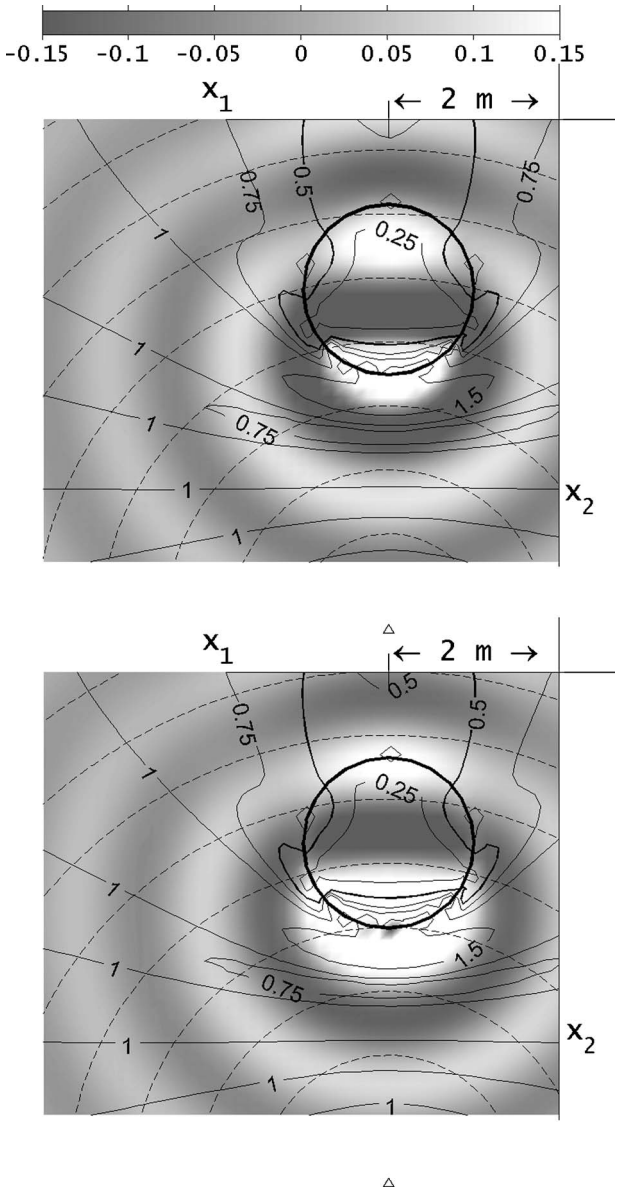


FIG. 6. Reproduced sound field at 220 Hz in free field by AWFS with $\lambda = 50$ (real part on top, imaginary part on bottom). Dashed lines show zero-pressure contours of the target pressure field. Local normalized quadratic residual error is represented as continuous line contour plots. Reproduction sources have been replaced by a thick source line for clarity. Δ, virtual source.

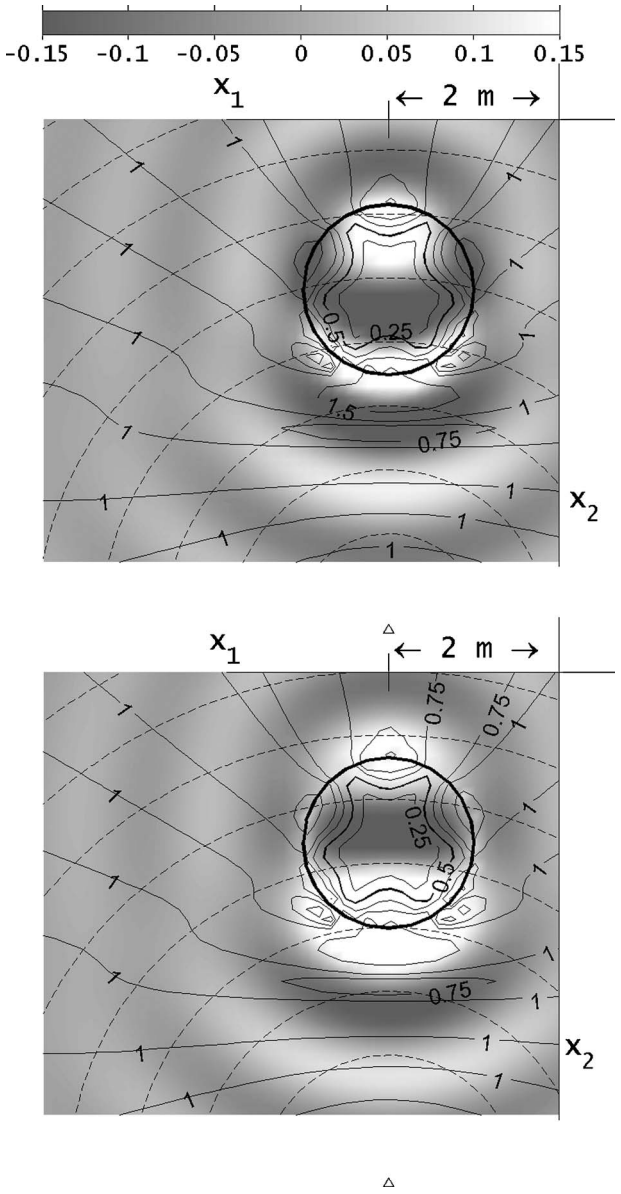


FIG. 7. Reproduced sound field in free field at 220 Hz by reproduction error minimization with $\lambda = 50$ and $\mathbf{q}_{\mathrm{WFS}} = \mathbf{0}$ (real part on top, imaginary part on bottom). Dashed lines show zero-pressure contours of the target pressure field. Local normalized quadratic residual error is represented as continuous line contour plots. Reproduction sources have been replaced by a thick source line for clarity. Δ, virtual source.

reproduced fields. The reproduced sound fields approach the target wave field. On these three figures, the spatial extent of effective reproduction is also shown. The size of this region varies with frequency. The spatial extent of effective reproduction is displayed via contour plots of the local quadratic normalized error $(|p^{(\mathrm{im})}(\mathbf{x},\omega) - p^{(\mathrm{rep})}(\mathbf{x},\omega)|^2 / |p^{(\mathrm{im})}(\mathbf{x},\omega)|^2)$. Contours are plotted only for the 0.25, 0.5, 0.75, 1, and 1.5 levels. The 0.5 level is suggested as the acceptable reproduction region limit for comparison purposes (Kirkeby and Nelson, 1993; Gauthier *et al.*, 2005a). AWFS introduces lower reproduction error than WFS in the source array center and the effective reproduction region is somewhat larger than the WFS effective region. AWFS in Fig. 6 also produces a more extended reproduction area (based on the 0.5 error level) than when simply minimizing the reproduction error at the error sensors with a control effort penalty (Fig. 7).

One of the benefits of AWFS is that it is possible to adjust the solution between a pure WFS realization (using a large penalization parameter) to a simple reproduction error minimization (with $\lambda \to 0$). This is shown in Fig. 8 on the basis of E_{LS} variations as a function of the penalization term λ. WFS, AWFS and reproduction error minimization regions are defined in Fig. 8. The E_{LS} values for regularization parameter $\lambda = \sigma_i^2$ (where σ_i are the transfer impedance matrix singular values) are plotted as circles. WFS corresponds to a large value of λ and E_{LS} does not significantly change with λ in this region. Reproduction error minimization corresponds to small values of λ and E_{LS} does not significantly change with λ in this region. The AWFS region is defined by the

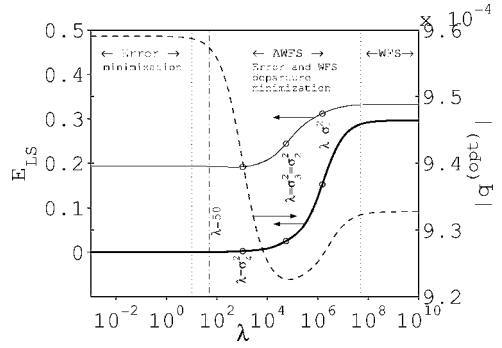


FIG. 8. E_{LS} at the error sensors (thick line) and at the monitoring sensors (thin line) and $(|\mathbf{q}^{(\mathrm{opt})}|)=\sqrt{\mathbf{q}^{(\mathrm{opt})H}\mathbf{q}^{(\mathrm{opt})}}$ (dashed line) at 220 Hz as a function of the regularization parameter λ. E_{LS} values for $\lambda=\sigma_i^2$ are shown as circles.

transition region between WFS and reproduction error minimization. The points for $\lambda=\sigma_i^2$ belong to this transition region.

For near-coincident error sensor arrays (relative to the wavelength) such as shown in Fig. 2, an under-regularized AWFS ($\lambda\rightarrow 0$) system can be regarded as a "closed-loop Ambisonics" reproduction system ("Ambisonics" systems are usually implemented as a open-loop systems, that is without any feedback or compensation), because such sensor arrays can be used to evaluate the pressure and two perpendicular pressure gradients in the horizontal plane. This relation between "Ambisonics" reproduction and AWFS is examined in Sec. VI A.

These simple simulations were used as an exemplary illustration of a simple AWFS behavior in free field. The more general underlying AWFS behavior, independent of the reproduction space, is exposed in the following section.

V. AWFS AND SINGULAR VALUE DECOMPOSITION OF Z

The singular value decomposition of the transfer impedance matrix is (Golub and van Loan, 1996)

$$\mathbf{Z}=\mathbf{U}_{M\times M}\boldsymbol{\Sigma}_{M\times L}\mathbf{V}^H_{L\times L}, \tag{12}$$

where subscripts indicate matrix dimensions. $\mathbf{U}$ is the matrix whose columns are the left singular vectors $\mathbf{u}_i$ of $\mathbf{Z}$, $\mathbf{V}$ is the matrix whose columns are the right singular vectors $\mathbf{v}_i$ of $\mathbf{Z}$ and $\boldsymbol{\Sigma}$ is the rectangular matrix of $\mathbf{Z}$ real positive singular values σ_i

$$\boldsymbol{\Sigma}=\begin{bmatrix}\sigma_1 & 0 & \cdots & & & 0\\ 0 & \sigma_2 & 0 & \cdots & & 0\\ & & \ddots & & & \\ 0 & \cdots & 0 & \sigma_M & \cdots & 0\end{bmatrix}_{M\times L}, \tag{13}$$

where $\sigma_i>\sigma_{i+1}$. The rank r of $\mathbf{Z}$ ($r\leqslant M<L$) is equal to the number of nonzero singular values. In Eq. (13) it was assumed that $r=M$ for convenience.

The matrices $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices ($\mathbf{U}^H\mathbf{U}=\mathbf{U}\mathbf{U}^H=\mathbf{V}^H\mathbf{V}=\mathbf{V}\mathbf{V}^H=\mathbf{I}$) since the left and right singular vectors are orthonormal vectors. SVD has well-known properties (Golub and van Loan, 1996; Hansen, 1998). Examples of SVD in acoustics are described by Nelson (2001); Williams *et al.* (2003); Bai and co-workers (2004; 2005). The span of the left singular vectors $\mathbf{u}_i$ constitutes the range of the matrix $\mathbf{Z}$ and the span of the last $L-r$ right singular vectors $\mathbf{v}_i$ is the null space of $\mathbf{Z}$ ($\mathbf{Z}\mathbf{v}_i=\mathbf{0}$ for $i>r$ and $i\leqslant L$) (Lancaster, 1985; Golub and van Loan, 1996) In order to explicitly relate left and right singular vectors to physical parameters, right singular vectors $\mathbf{v}_i$ and left singular vectors $\mathbf{u}_i$ will hereafter be, respectively, identified as "source modes" and "pressure modes." Each source and pressure modes pair with the corresponding singular value is here defined as a "radiation mode." This AWFS-specific terminology is different from the terminology introduced by Photiadis (1990) and from the more common definition of a radiation mode being a source distribution which radiates sound power independently from the other radiation modes.

Equations (1) and (12) result in this transformed radiation equation

$$\tilde{\mathbf{p}}^{(\mathrm{rep})}=\boldsymbol{\Sigma}\tilde{\mathbf{q}} \tag{14}$$

with the transformed source strengths $\tilde{\mathbf{q}}$ and acoustic pressures $\tilde{\mathbf{p}}$,

$$\tilde{\mathbf{p}}^{(\mathrm{im})}=\mathbf{U}^H\mathbf{p}^{(\mathrm{im})},\quad \tilde{\mathbf{p}}^{(\mathrm{rep})}=\mathbf{U}^H\mathbf{p}^{(\mathrm{rep})},\ \tilde{\mathbf{q}}=\mathbf{V}^H\mathbf{q}. \tag{15}$$

Since $\boldsymbol{\Sigma}$ is a pseudodiagonal rectangular matrix (the elements are nonzero only on the main diagonal), the radiation problem of Eq. (14) is a set of independent acoustical radiation between source modes $\mathbf{v}_i$ and pressure modes $\mathbf{u}_i$.

A. Nature of the under-determined set of equations $\mathbf{Zq}=\mathbf{p}^{(\mathrm{im})}$

In Sec. II B we treated the equation $\mathbf{Zq}=\mathbf{p}^{(\mathrm{im})}$ as a quadratic minimization. We now discuss the under-determined set of equations $\mathbf{Zq}=\mathbf{p}^{(\mathrm{im})}$. Since there are more unknown variables than equations ($L>M$), the set of equations $\mathbf{Zq}=\mathbf{p}^{(\mathrm{im})}$ cannot be solved in this form: there exists an infinite set of source strength values that will produce $\mathbf{p}^{(\mathrm{im})}$. Such under-determination is expressed as the contribution of left singular vectors in the solution

$$\mathbf{q}=\sum_{i=1}^{r}\frac{\tilde{p}_i^{(\mathrm{im})}}{\sigma_i}\mathbf{v}_i+\sum_{i=r+1}^{L}\alpha_i\mathbf{v}_i, \tag{16}$$

where α_i represents arbitrary coefficients which, with the vector basis of the null space of $\mathbf{Z}$, define the under-determined nature of $\mathbf{Zq}=\mathbf{p}^{(\mathrm{im})}$. Inserting (16) in $\mathbf{Zq}$ will effectively produce $\mathbf{p}^{(\mathrm{im})}$ regardless of α_i. A solution to such an under-determined problem is to cast it in a new quadratic minimization form which seeks to reduce the departure between the solution $\mathbf{q}$ and a given solution $\mathbf{q}_o$. If this *a priori* solution (Hansen, 1998) is simply 0 then the pseudoinversion can be applied to solve $\mathbf{Zq}=\mathbf{p}^{(\mathrm{im})}$ (Golub and van Loan, 1996). If the *a priori* solution $\mathbf{q}_o$ is not zero, a more general rule is obtained: the α_i in Eq. (16) correspond to the projection of $\mathbf{q}_o$ on $\mathbf{v}_i$

$$\alpha_i=\mathbf{v}_i^H\mathbf{q}_o\ \forall\ i>r\quad\text{and } i\leqslant L \tag{17}$$

and the first r contributions of $\mathbf{v}_i$ are determined by the first term in Eq. (16). This minimizes the cost function

$$J(\alpha_i) = \sum_{i=1}^{r} (\tilde{p}_i^{(\mathrm{im})}/\sigma_i - \mathbf{v}_i^H \mathbf{q}_o)^2 + \sum_{i=r+1}^{L} (\alpha_i - \mathbf{v}_i^H \mathbf{q}_o)^2. \quad (18)$$

In the next section, we proceed to show how the AWFS solution [expressed by Eq. (9)] corresponds to a modified version of Eqs. (16) and (17) where $\mathbf{q}_{\mathrm{WFS}}$ replace $\mathbf{q}_o$.

B. Independent radiation mode control for AWFS

Taking the AWFS solution of Eq. (9) and using the SVD of $\mathbf{Z}$, the AWFS solution can be expanded over the source modes $\mathbf{v}_i$, $\mathbf{q}^{(\mathrm{opt})} = \sum_{i=1}^{L} \tilde{q}_i \mathbf{v}_i$, as follows:

$$\tilde{q}_i = \begin{cases} \dfrac{\sigma_i}{(\sigma_i^2 + \lambda_i)} \tilde{p}_i^{(\mathrm{im})} + \dfrac{\lambda_i}{(\sigma_i^2 + \lambda_i)} \mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}} \quad \forall \quad i \leqslant r \leqslant M, \\ \mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}} \quad \forall\, \lambda_i > 0, \quad \forall\, i > r \text{ and } i \leqslant L, \\ 0 \quad \forall\, \lambda_i = 0, \quad \forall\, i > r \text{ and } i \leqslant L, \end{cases} \quad (19)$$

where the independent contributions of $\mathbf{v}_i$ in $\mathbf{q}^{(\mathrm{opt})}$ were individually regularized with λ_i. AWFS thus proceeds to control each radiation mode independently. The AWFS solution of Eq. (9) implicitly assumed an identical regularization parameter for all sources modes, $\lambda_i = \lambda \ \forall i$. Note that Eq. (19) was slightly modified in comparison to Eq. (9) for the particular case $\lambda_i = 0$: Eq. (19) is forced to the pseudoinverse definition by the addition of the third case in Eq. (19) (Hansen, 1998). The reproduction error is expanded

$$\mathbf{e} = \sum_{i=1}^{M} \tilde{e}_i \mathbf{u}_i \quad (20)$$

with

$$\tilde{e}_i = \frac{\lambda_i (\tilde{p}_i^{(\mathrm{im})} - \sigma_i \mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}})}{(\sigma_i^2 + \lambda_i)} \quad \forall\, i \leqslant r. \quad (21)$$

Equation (19) shows that even if the multichannel system must generate the L source inputs in $\mathbf{q}^{(\mathrm{opt})}$, the total number of degrees of freedom is in fact r. In the first case [$i \leqslant r \leqslant M$ in Eq. (19)], the $\tilde{q}_i$ are a combination of reproduction error minimization with $\mathbf{q}_{\mathrm{WFS}}$ projection on the corresponding source mode $\mathbf{v}_i$. This combination is weighted by λ_i. When $\lambda_i = 0$, Eq. (19) provides the same solution as Eqs. (16) and (17) for $i \leqslant r \leqslant M$. In the second case [$i > r$ and $i \leqslant L$ in Eq. (19)], the corresponding $\tilde{q}_i$ are solely defined by the projection of $\mathbf{q}_{\mathrm{WFS}}$ on $\mathbf{v}_i$ if $\lambda_i > 0$. This shows that even if such source modes produce no sound pressure at sensor locations, they contribute to the solution in order to better match the WFS sound field reproduction outside the control region. These results explain how a part of the WFS operators is modified by AWFS and how the other fixed part is defined by the projection of $\mathbf{q}_{\mathrm{WFS}}$ on the null space vector basis. Since these observations stem from an analytical development and since the SVD exist for any complex matrix $\mathbf{Z}$, it is possible to extent this interpretation to a any real situation.

Equation (19) shows the discontinuity in the solution $\mathbf{q}^{(\mathrm{opt})}$ as $\lambda \to 0$, specifically in the components $\tilde{q}_i$ corresponding to the null space of $\mathbf{Z}$. This discontinuity in the solution would not be observable in the reproduced field at the error sensors, because the span of these source modes is the null space of the transfer impedance matrix. However, anywhere else such discontinuity would be observable in the reproduced sound field.

Equations (19) and (21) give insights into how each radiation modes are affected by $\lambda_i = \lambda$. Rewriting these equations for $i \leqslant r$ in terms of transformed WFS reproduction error ($\tilde{e}_{i\mathrm{WFS}} = \mathbf{u}_i^H \mathbf{p}^{(\mathrm{im})} - \sigma_i \mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}} = \mathbf{u}_i^H (\mathbf{p}^{(\mathrm{im})} - \mathbf{Z}\mathbf{q}_{\mathrm{WFS}})$), one finds

$$\tilde{q}_i = \frac{\sigma_i}{(\sigma_i^2 + \lambda_i)} \tilde{e}_{i\mathrm{WFS}} + \mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}} \quad \forall\, i \leqslant r \quad (22)$$

and

$$\tilde{e}_i = \frac{\lambda_i}{(\sigma_i^2 + \lambda_i)} \tilde{e}_{i\mathrm{WFS}} \quad \forall\, i \leqslant r. \quad (23)$$

When $\lambda_i \ll \sigma_i^2$, the solution $\tilde{q}_i$, for the control of the radiation mode i, is dominated by WFS reproduction error $\tilde{e}_{i\mathrm{WFS}}$ minimization ($\tilde{q}_i - \mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}}$ is large and $\tilde{e}_i$ is small). When $\lambda_i \gg \sigma_i^2$, the solution $\tilde{q}_i$ is dominated by WFS departure penalty ($\tilde{q}_i - \mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}}$ is small and $\tilde{e}_i$ is large). This illustrates that for a regularization parameter λ as in Eqs. (7) and (9), the amount of penalization is different for each radiation mode; some radiation modes can be over-penalized ($\lambda \geqslant \sigma_i^2$) and some others under-penalized ($\lambda \leqslant \sigma_i^2$). This suggests that the behavior of AWFS with a global penalization ($\lambda_i = \lambda$) is a complex contribution of the radiation modes. This defines the AWFS transition band in Fig. 8.

VI. ANALYSIS OF SIMULATION RESULTS WITH SVD

The behavior of independent AWFS is illustrated through numerical results shown in Figs. 9–14 for free-field situations. Our goal is to simply exemplify the analytical results presented in the preceding section.

Figure 9 shows the first four source and pressure modes for the configuration of Fig. 2 at 220 Hz. There is 24 sources modes and four pressure modes in this configuration. The last 20 source modes produce no sound pressure at the sensor array. Figure 9 shows that the source modes have a single circumferential harmonic component, and the corresponding pressure modes provide a multipole expansion of the sound field at the sensor array. Moreover, one notes that the third source mode has a $\pi/4$ radian spatial phase difference with the second source mode (a similar trend is observed for the corresponding pressure modes). The source modes producing the $\mathbf{Z}$ null space are shown in Fig. 10. The null space source modes are spatially high-pass filtered impulses. This filtering removes the low-order source mode components from the spatial impulses.

Figure 11 shows the produced sound field for the first six source modes $\mathbf{v}_i$ at 220 Hz. The first source mode produces uniform sound pressure at the sensor array [subfigure (a)]. The next two source modes [subfigures (b) and (c)] produce perpendicular pressure gradients. The next source mode [subfigure (d)] produces a tesseral quadrupole field compo-

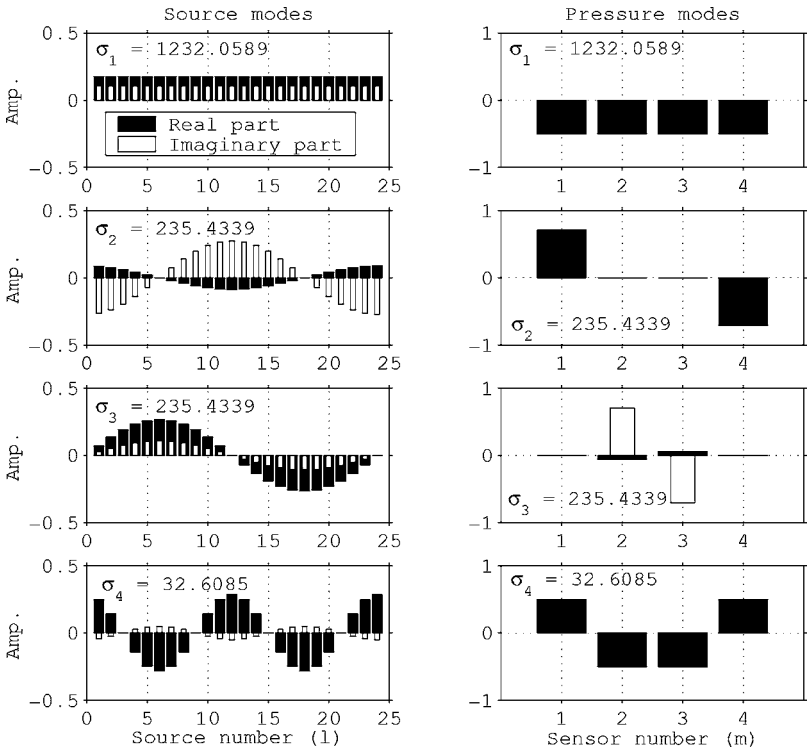


FIG. 9. Source modes $\mathbf{v}_i$ at 220 Hz on the left-hand side and corresponding pressure modes $\mathbf{u}_i$ (for $\sigma_i \neq 0$, that is $i \leqslant r = M$) on the right-hand side. Source and sensor numbers are shown in Fig. 2.

nent. The last two source modes produce no sound pressure at the sensors [subfigures (e) and (f)] since the corresponding source modes belong to the nullspace. These two source modes are shown on the first line in Fig. 10.

Figure 12 shows multipole representations of the pressure modes $\mathbf{u}_i$ at 220 Hz. The corresponding free-field directivities are also shown as polar plots. These directivity patterns were computed with plane wave responses at the sensors weighted by the complex pressure modes $\mathbf{u}_i^H$ as sensor gains. This figure illustrates that the $i=1$ pressure mode corresponds to sound pressure, the $i=2$ pressure mode approximates pressure gradient along x_1 (dipole sensitivity in the low frequency limit), the $i=3$ pressure mode gives a pressure gradient along x_2 (dipole sensitivity in the low frequency limit) and the $i=4$ pressure mode corresponds to a tesseral quadrupole finite approximation. Since Eq. (19) sug-

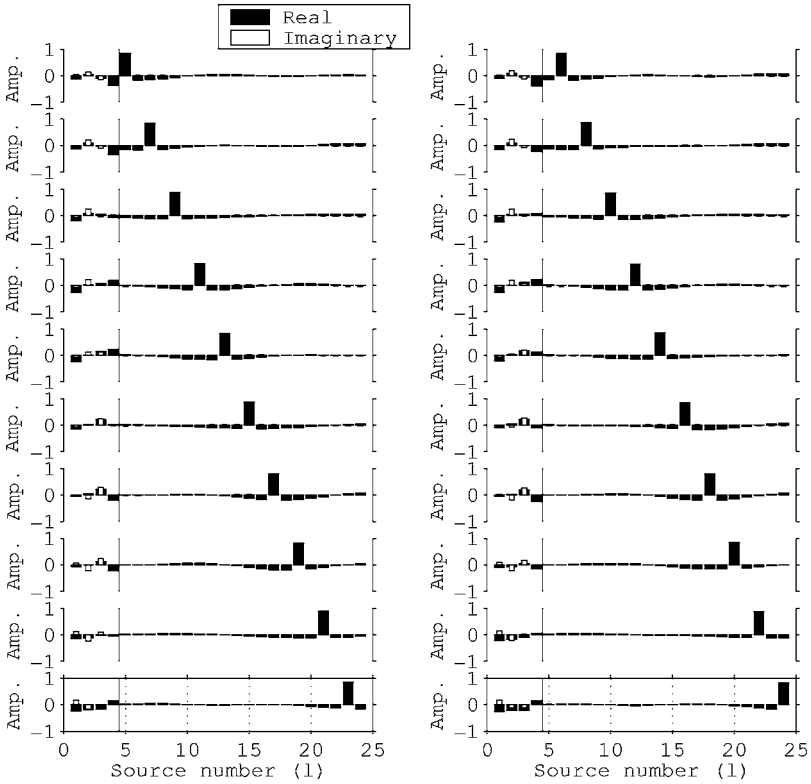


FIG. 10. The null space source modes $\mathbf{v}_i$ at 220 Hz. Increasing i proceeds from left to right and from top to bottom.

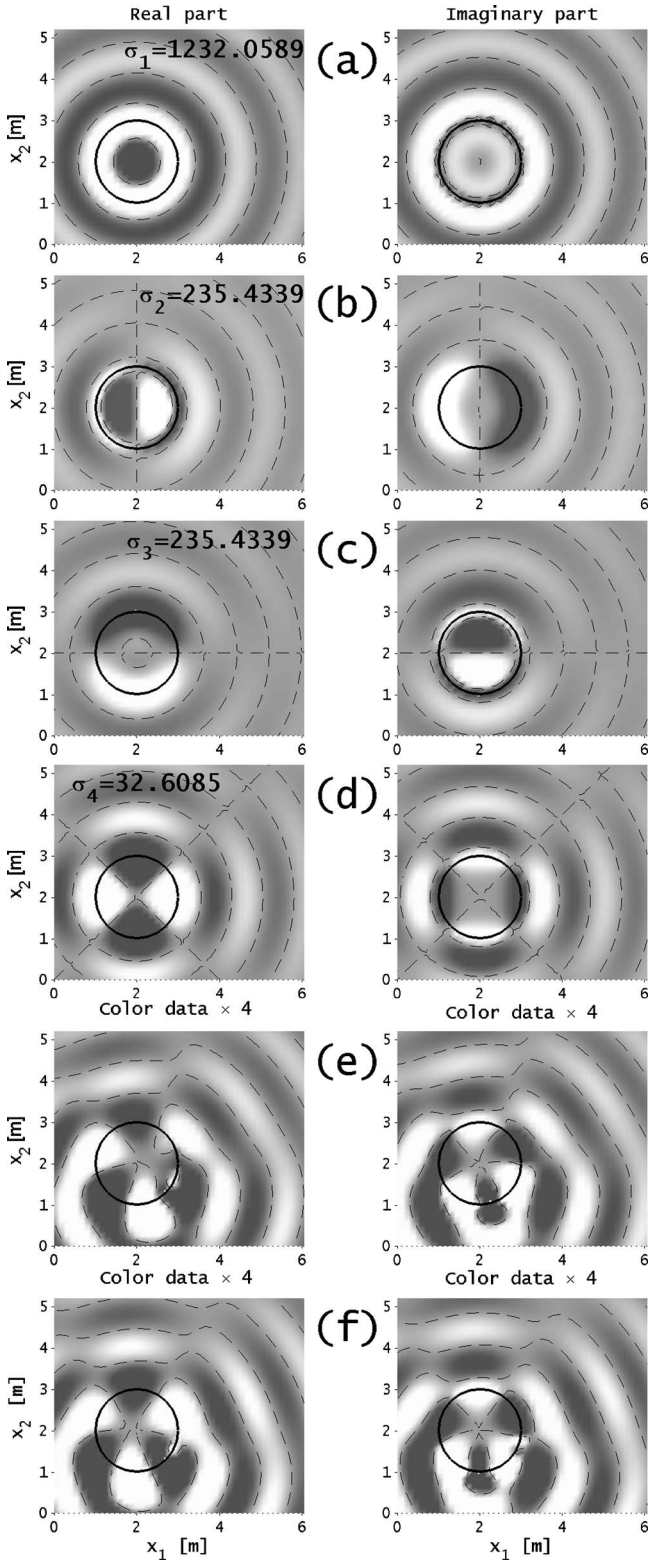


FIG. 11. Sound field radiated by the first six source modes. Source modes (a)–(d) generate the range of the transfer impedance matrix. Source modes (e) and (f) belong to the null space of the transfer impedance matrix. Dashed lines show zero-pressure contours of the pressure fields.

gests independent control of the reproduced pressure modes, one can conclude, from the pressure mode shapes of Fig. 12, that AWFS [Eqs. (7) and (9)] effectively operates on the independent control of pressure, pressure gradients and tesseral quadrupole pressure modes at the sensor array.

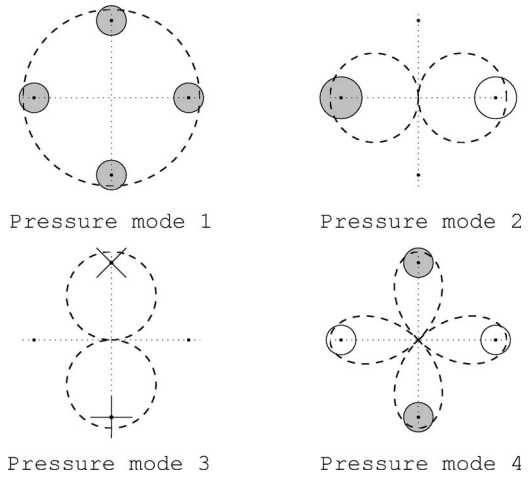


FIG. 12. Multipole representations of the first $M=4$ pressure modes and corresponding free field directivity.· sensor position; ○, positive real part; filled ○, negative real part; +, positive imaginary part; ×, negative imaginary part. Symbol diameter illustrates the magnitude of the corresponding value. Dashed lines, free field directivity.

Figure 13 shows the source strength vectors for WFS and AWFS used to produce the sound fields in Figs. 5 and 6 along with the projections of $\mathbf{q}_{\mathrm{WFS}}$ and $\mathbf{q}^{(\mathrm{opt})}$ on the source modes $\mathbf{v}_i$. By comparing the left graphs of Fig. 13, it is clear that WFS activates only a part of the reproduction sources while AWFS significantly activates nearly all sources. This illustrates the additive combination of WFS and reproduction error minimization. As shown in the right graphs of Fig. 13, AWFS only changes the first $r=M$ source mode contributions $\tilde{q}_i$ in comparison with WFS, this was shown in Eq. (19). This illustrates that the general AWFS minimization [Eqs. (7) and (9)] effectively gives the SVD solution of Eq. (19). Even if deriving (19) from (9) is straightforward using SVD, this numerical example is used to illustrate this equivalence.

A. AWFS and "Ambisonics"

The next simulations are introduced to draw the link between AWFS, using a compact centered sensor array as shown in Fig. 2, and "Ambisonics".

For the first simulation, the regularization parameter λ was set to the small value of 1×10^{-4} and the WFS solution $\mathbf{q}_{\mathrm{WFS}}$ was set to 0 in Eq. (9). A small penalization parameter λ is one of the limiting cases of AWFS. This corresponds to the minimization of the reproduction errors with a small penalty of the control effort. The optimal source strengths $\mathbf{q}^{(\mathrm{opt})}$ at 220 Hz and the corresponding projections on the source modes $\mathbf{v}_i$ are shown in Fig. 14(a). Figure 14(a) shows that only the first $r=M$ source modes are active, this is consistent with Eq. (19) when $\mathbf{q}_{\mathrm{WFS}}=0$. As explained below, such a type of source distribution is somewhat related to "Ambisonics" panning functions.

A short description of "Ambisonics" is introduced here. A more complete theoretical development is included in the appendix. "Ambisonics" is an open-loop reproduction technology from the 1970s introduced by Gerzon (Rumsey, 2001). In the most simple case, the target wave field $p^{(\mathrm{im})}(\mathbf{x},\omega)$ is defined as a plane wave in free field

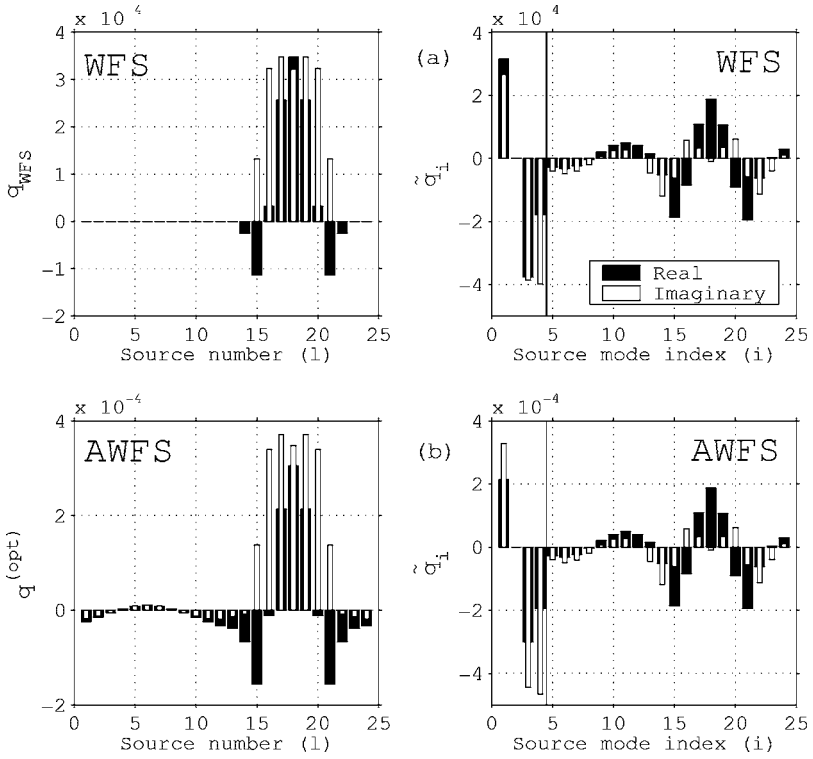


FIG. 13. Left, complex source strengths for (a) WFS ($\mathbf{q}_{\mathrm{WFS}}$ corresponding to Fig. 5) and (b) AWFS ($\mathbf{q}^{(\mathrm{opt})}$ corresponding to Fig. 6). Right, complex source mode contributions $\tilde{q}_i$ of (a) WFS and (b) AWFS solutions obtained from the orthogonal projection ($\tilde{q}_i=\mathbf{v}_i^H\mathbf{q}^{(\mathrm{opt})}$ or $\tilde{q}_i=\mathbf{v}_i^H\mathbf{q}_{\mathrm{WFS}}$) of the source strengths on the source modes.

$$p^{(\mathrm{im})}(R,\phi,\omega)=A(\omega)e^{jkR\cos(\phi-\theta)}, \tag{24}$$

where space coordinates are now described in terms of cylindrical coordinates (R,ϕ,z), and θ is the direction of propagation of the plane wave.

The reproduction system is assumed to be in free field and the reproduction sources (uniformly distributed over 2π radians) produce a set of plane waves at the origin of the coordinate system

$$p^{(\mathrm{rep})}(R,\phi)=\sum_{l=1}^{L}A_l e^{jkR\cos(\phi-\phi_l)}, \tag{25}$$

where A_l is the unknown contribution of the plane reproduction wave l whose direction of propagation is ϕ_l.

Using Fourier-Bessel series for both the target sound pressure field [Eqs. (24) and (A2)] and the reproduced field [Eqs. (25) and (A8)], it is possible to minimize (in an LMS

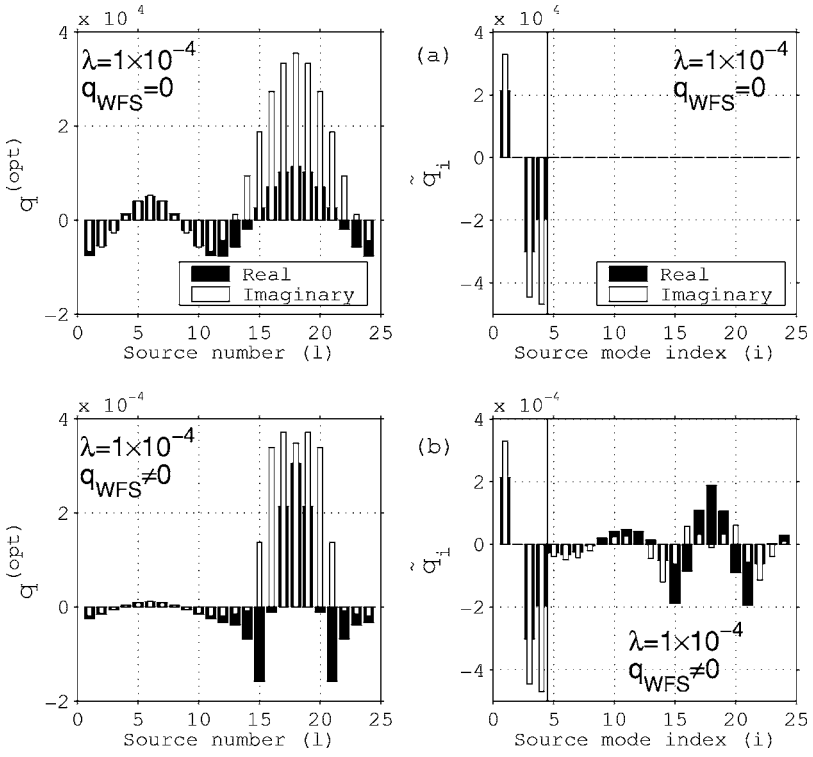


FIG. 14. Left, complex source strengths for optimal control (a) $\mathbf{q}^{(\mathrm{opt})}$ with $\lambda=1\times10^{-4}$ and $\mathbf{q}_{\mathrm{WFS}}=\mathbf{0}$, (b) $\mathbf{q}^{(\mathrm{opt})}$ with $\lambda=1\times10^{-4}$ and $\mathbf{q}_{\mathrm{WFS}}\neq\mathbf{0}$. Right, corresponding source mode contributions $\tilde{q}_i$ obtained from the orthogonal projections ($\tilde{q}_i=\mathbf{v}_i^H\mathbf{q}^{(\mathrm{opt})}$) of the source strengths on the source modes.

sense) the reproduction error over each of the coefficient of the Fourier-Bessel series. This gives a panning function (amplitude A_l of plane waves reproduction sources as function of θ and ϕ_l)

$$A_l = \frac{A}{L}\left[1 + 2\sum_{n=1}^{N} \cos(n(\theta - \phi_l))\right]. \tag{26}$$

This is a circular *sinc* function of ϕ_l for which the series truncature order N defines the number and amplitudes of the side lobes.

From this, one gets a panning function as to how (in an *a posteriori* operation) distribute a given signal to a circular source array to create, under the free field and plane wave assumptions, the target wave field. Since the Fourier-Bessel coefficients [see Eq. (A5)] correspond to low-order directivity patterns of microphones (first order "Ambisonics," $N=1$, gives three coefficients in the two series: monopole and two perpendicular dipole patterns), it is possible to consider the limiting case of AWFS with $\lambda \to 0$ as a "closed-loop Ambisonics" realization. The "closed-loop" appears since AWFS would be implemented with real microphones for reproduction error minimization while "Ambisonics" is a LMS solution that is implemented as an open-loop system. This is first supported by the pressure modes of Fig. 12 which shows that AWFS in this case corresponds to independent reproduction control of pressure, two perpendicular sound pressure gradients and a tesseral quadrupole approximation of the sound field at the error sensors in the low frequency limit. However, this connection between "Ambisonics" and AWFS using compact sensor array is essentially conceptual. This arises from that AWFS and "Ambisonics" use (1) different radiation models (spherical waves and plane waves for the reproduction system, respectively) and (2) different sensing devices (discrete sensor array for AWFS and perfectly coincident directive patterns for "Ambisonics").

The previous remarks show that AWFS (while letting $\lambda \to 0$ and forcing the WFS solution to zero) can be interpreted as a closed-loop "Ambisonics" realization. In the next example, we proceed in explaining how the AWFS realization (with $\mathbf{q}_{\mathrm{WFS}} \neq 0$ and for the specific sensor configuration shown in Fig. 2) is an hybrid solution between WFS and a "closed-loop Ambisonics" realization. This corresponds to "closed-loop modified Ambisonics." The "modified Ambisonics" arises from the superposition of the filtered WFS solution and the low-order closed-loop "Ambisonics" solution.

Figure 14(b) corresponds to the limiting case of AWFS with a vanishing regularization parameter $\lambda = 1 \times 10^{-4}$ and $\mathbf{q}_{\mathrm{WFS}}$ corresponding to the WFS solution [Eqs. (5) and (6)]. For the first four source modes contributions ($\tilde{q}_t$ with $i = 1,2,3,4$), the solution with $\mathbf{q}_{\mathrm{WFS}} \neq \mathbf{0}$ is almost identical to the solution with $\mathbf{q}_{\mathrm{WFS}} = \mathbf{0}$. This was to be expected from Eq. (19) since when $\lambda \to 0$, $\tilde{q}_i \approx \tilde{p}_i^{(\mathrm{im})}/\sigma_i$ for $i \leqslant r$. The higher-order source mode ($r < i \leqslant L$) contributions $\tilde{q}_i$ are the WFS projections ($\mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}}$) on $\mathbf{v}_i$ and are independent of λ. This appears when comparing the two graphs on the right-hand side of Fig. 14.

The previous results suggested a conceptual connection between AWFS in the configuration of Fig. 2 and "Ambisonics." In the next example, the nearly exact correspondence of AWFS, using Eq. (19), and the "Ambisonics" panning function of Eq. (26) will be shown. For this simulation, the AWFS radiation model is replaced with a reproduction source radiation model consistent with the "Ambisonics" formalism (plane waves as reproduction sources). The sensor array is however kept identical to the one shown in Fig. 2. Eq. (19) is also applied in the following conditions: First, the *a priori* solution $\mathbf{q}_{\mathrm{WFS}}$ is set to $\mathbf{0}$. Second, the effort penalization parameter λ is set to 0. Third, the $\tilde{q}_i$ components are only computed for $i = 1,2,3$ (these first three pressure modes correspond to uniform pressure and two pressure gradients sensing at the sensor array). This would correspond to first order "Ambisonics" using $N=1$ in Eq. (26). Fourth, the target wave field is a plane wave [see Eq. (24)] for which the direction of propagation corresponds to reproduction source number 12, see Fig. 2 for the angular plane source distributions which correspond to the reproduction source angular positions). The corresponding AWFS and "Ambisonics" solutions are shown in Fig. 15. The AWFS and "Ambisonics" source distributions are almost identical and correspond to a raised cosine [see Eq. (26) with $N=1$]. Despite appearances, the solutions are not exactly the same. The difference is due to the AWFS sensor array which can only provide a finite approximation to monopole and dipole sensitive patterns. The AWFS solution of Eq. (19) matches the "Ambisonics" solution of Eq. (A16) when the radiation models are consistent.

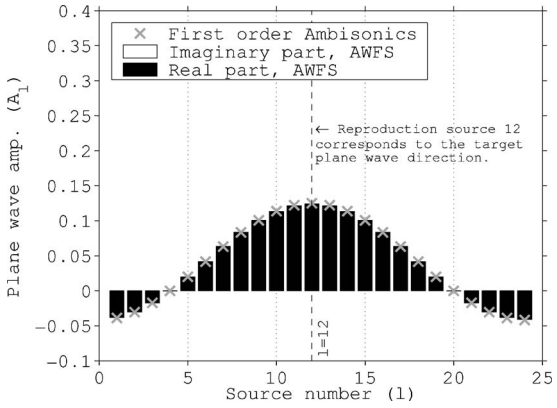


FIG. 15. Reproduction source amplitudes for plane wave reproduction with first order "Ambisonics" and truncated modified AWFS.

In summary, when the regularization parameter λ is large, AWFS simply converges towards WFS. When λ is small, the first ($i \leqslant r$) source modes contributions $\tilde{q}_i$ approach a closed-loop "Ambisonics" realization with a small WFS departure penalty, while all the other source modes (which create the nullspace) contributions are determined by the projection of WFS solution $\mathbf{q}_{\mathrm{WFS}}$ on the corresponding source modes $\mathbf{v}_i$. This precisely defines what was described earlier as "closed-loop modified Ambisonics." Briefly stated, the limiting cases of AWFS implementation produce (1) the WFS solution or (2) a closed-loop "Ambisonics" solution superposed to the spatially filtered WFS solution.

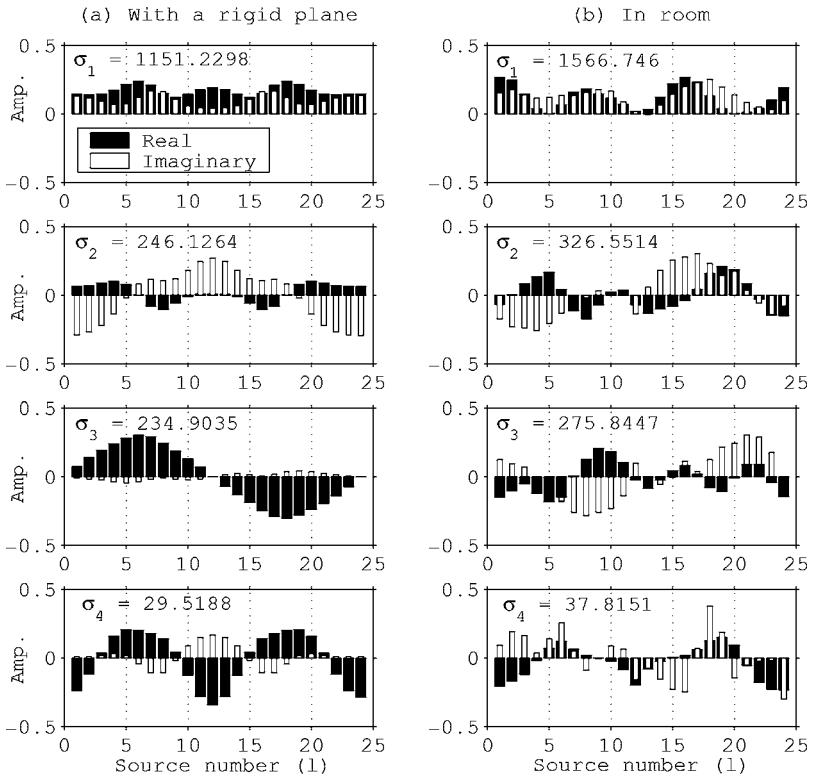


FIG. 16. First four source modes and singular values. (a) with a rigid plane in x_2-x_3. (b) in a room of 6.05 $\times 5.2 \times 2$ m with a reverberation time of approximately 0.28 s.

B. Including reflections and room effects

The preceding paragraphs showed the fundamental theoretical mechanisms of AWFS on the basis of analytical developments and illustrative simulation results in free field. However, AWFS has the potential to compensate for the effects of the listening space. Although this paper does not primarily address AWFS in room situation, some details are introduced to extend the significance of the preceding sections. In room, the compensation is more significant since the free-field assumptions of WFS are no longer respected and the accuracy of the reproduction by WFS is reduced (Gauthier *et al.*, 2005b).

The main goal of this section is to illustrate that even if both source modes and singular values are altered by the presence of boundaries, the pressure modes still provide pressure and pressure gradients estimates at the error sensors. Interestingly, this, with Eq. (19) (which is valid for any **Z**), extends the relation between AWFS and "Ambisonics" to sound field reproduction in rooms: independent control of pressure, pressure gradients and tesseral quadrupole pressure modes is achieved by AWFS in the configuration shown in Fig. 2 even in theoretical non-free-field conditions.

Figure 16(a) shows the first four source modes at 220 Hz when a rigid wall replaces the x_2-x_3 plane. The transfer impedance matrix was computed using the method of images. The corresponding pressure modes are shown as multipole representations in Fig. 17. Even if the pressure modes of Fig. 17 are slightly different to the free-field source modes of Fig. 12, the pressure modes still give finite difference approximations of pressure, pressure gradients and tesseral quadrupole as shown in Fig. 17. Figure 16(b) shows the first source modes when the system is located in a room for which the transfer impedance is based on a modal model including 6137 modes [the model is fully described by Gauthier *et al.* (2005a), see also Morse and Ingard (1968)]. The room is rectangular and its dimensions are $6.05 \times 5.2 \times 2$ m the room corner is located at $x_1=x_2=x_3=0$. The source and sensor plane is at $x_3=1.2192$ m. The reverberation time is approximately 0.28 s. The corresponding pressure modes are shown as a multipole representation in Fig. 18. Even though the source modes are distorted in comparison with the free field or rigid wall situation, the pressure modes still give finite difference approximations of pressure, perpendicular pressure gradients and tesseral quadrupole at the sensor array as shown in Fig. 18. By comparing Figs. 9

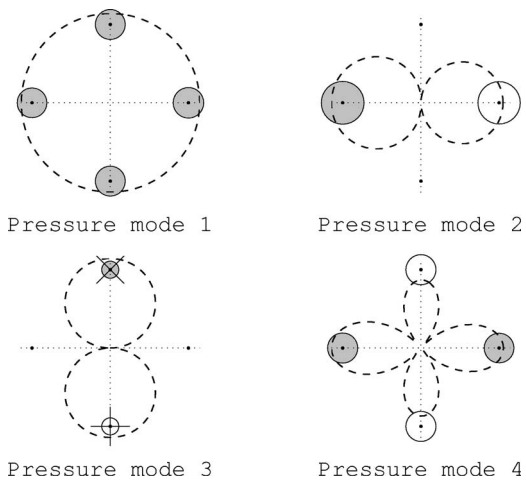


FIG. 17. Multipole representations of the first $M=4$ pressure modes and corresponding free field directivity for the source modes in Fig. 16(a). ·, sensor position; ○, positive real part; filled ○, negative real part; +, positive imaginary part; ×, negative imaginary part. Symbol diameter illustrates the magnitude of the corresponding value. Dashed lines, free field directivity.

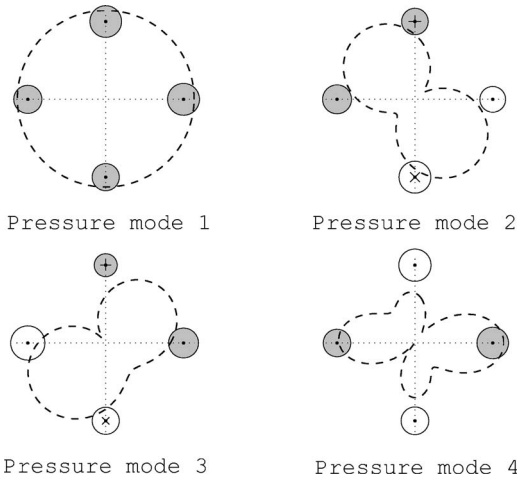


FIG. 18. Multipole representations of the first $M=4$ pressure modes and corresponding free field directivity for the source modes in Fig. 16(b). ·, sensor position; ○, positive imaginary part; filled ○, negative real part; +, positive imaginary part; ×, negative real part. Symbol diameter illustrates the magnitude of the corresponding value. Dashed lines, free field directivity.

and 16(b), one notes that the room singular values are significantly higher than the free-field and the semi-infinite space singular values. This demonstrates that for a given source strength vector $\mathbf{q}$, the sound pressures $\mathbf{p}^{(\mathrm{rep})}$ produced in a closed volume are larger than in free field.

According to these examples, independent control of radiation mode as pressure and higher-order sound pressure field derivatives is still a valid fundamental interpretation of Eq. (19) in the presence of a wall or a room. This suggests that AWFS based on independent radiation mode control as shown by Eq. (19) is a promising avenue for practical adaptive sound field reproduction which should somehow keep a link with "Ambisonics" and WFS.

VII. IMPLEMENTATION OF AWFS BY INDEPENDENT FEEDFORWARD CONTROL OF RADIATION MODES

This section describes a possible implementation of AWFS based on adaptive signal processing. As noted in Eq. (19), each source mode contribution in $\tilde{q}_i$ in $\mathbf{q}^{(\mathrm{opt})}$ is solely defined by the corresponding $\tilde{p}^{(\mathrm{im})}, \sigma_i, \lambda_i$ and $\mathbf{v}_i^H \mathbf{q}_{\mathrm{WFS}}$. It is possible to exploit this independence of the radiation modes for practical implementations of AWFS.

In a classical adaptive minimization of Eq. (7), a centralized multichannel adaptive algorithm would be implemented (Elliott, 2001) using a feedforward architecture. This is illustrated in Fig. 19. A set of z-transformed adaptive filters $\mathbf{w}(z)$ are adapted to reduce the quadratic summation of the M reproduction errors $\mathbf{e}_M(n)$. The discrete filters $\mathbf{w}(z)$ are fed by a reference signal $x(n)$ and produce the L command signals $\mathbf{y}_L(n)$ for the L reproduction sources. Here, $\mathbf{Z}(z)$ is the only physical system: All other operations are achieved in a signal processing domain. This includes $\hat{\mathbf{Z}}(z)$ which is an identified approximation of $\mathbf{Z}(z)$. The target field definition at the error sensors is obtained by passing the reference signal $x(n)$ through the target pressure modeler $\mathbf{a}(z)$. The WFS solution departure penalty [see Eq. (7)] is automatically taken into

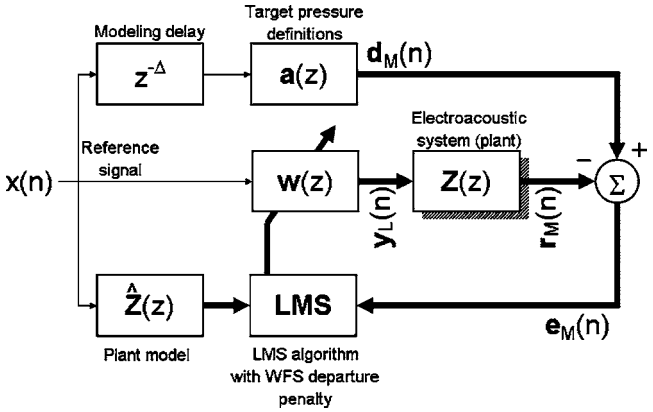


FIG. 19. Block diagram of a possible AWFS least-mean-square adaptive digital processing implementation using FXLMS multichannel algorithm. Subscripts denote signal dimensions.

account by the LMS adaptation which is based on a discrete time representation of the AWFS cost function of Eq. (7). Using M sensors, this will be a multiple-input-multiple-output (MIMO) $L \times M$ system. The computational burden involved in the block diagram of Fig. 19 becomes prohibitive as L, M and the adaptive filters order increase (Elliott, 2001; Bai and Elliott, 2004). Moreover, this may slow down the gradient LMS algorithm convergence as L and M increase (Bai and Elliott, 2004).

Another approach is illustrated as a block diagram in Fig. 20. $\mathbf{Z}(z)$ is again the only physical system. With this diagram, we seek to produce a set of M (with $M \leqslant L$) independent SISO (single-input–single-output) adaptive systems $\tilde{\mathbf{w}}_i(z)$. Here, each independent adaptive system generates the signal sent to each source mode $\mathbf{v}_i$ for $i < r \leqslant M$. This type of independent control using SVD was investigated by Bai and Elliott (2004) for cross-talk cancelation but without any *a priori* solution like $\mathbf{q}_{\mathrm{WFS}}$. This independent radiation mode-control of AWFS is also a modification of the principal-component LMS (PC-LMS) for active noise control of tonal disturbance (Cabell and Fuller, 1999; Elliott, 2000; Cabell *et al.*, 2001). The PC-LMS is based on SVD, but for AWFS

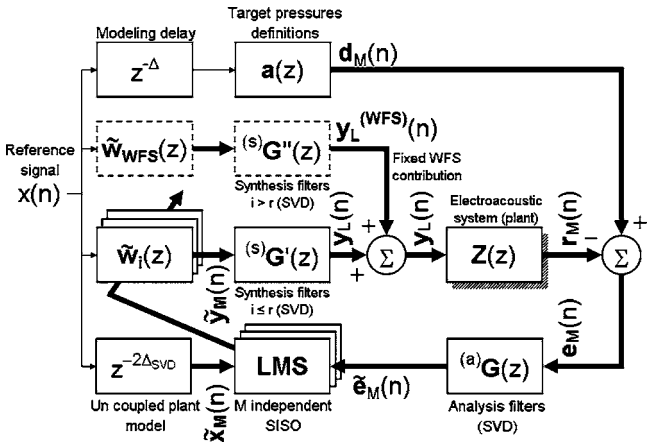


FIG. 20. Block diagram of a possible AWFS least-mean-square adaptive digital processing implementation based on independent radiation mode control. Subscripts denote signal dimensions.

the SVD decoupling must operate over the entire frequency range of interest (Bai and Elliott, 2004).

This implementation requires additional MIMO (multiple-input–multiple-output) synthesis and MIMO analysis filters, ${}^{(s)}\mathbf{G}'(z)$ and ${}^{(a)}\mathbf{G}(z)$, respectively, which must be generated in *a priori* identification stage. In this case, an identification of frequency response functions (FRF) $\hat{Z}_{ml}(\omega)$ between each source l and each sensor m must be achieved. The SVD of the corresponding transfer matrix for each frequency provides source modes $\mathbf{v}_i(\omega)$ and corresponding pressure modes $\mathbf{u}_i(\omega)$. The inverse Fourier transform of $\mathbf{v}_i(\omega)/\sigma_i(\omega)\,\forall\, i \leqslant r \leqslant M$ and $\mathbf{u}_i(\omega)$ generates the z-transformed discrete-time impulse responses (IR) matrices ${}^{(s)}\mathbf{G}'(z)$ and ${}^{(a)}\mathbf{G}(z)$, respectively. The ${}^{(s)}\mathbf{G}'(z)$ matrix generates the inputs of the individual reproduction sources from the source mode inputs with a delay of Δ_{SVD} samples (Bai and Elliott, 2004). Note that division by $\sigma_i(\omega)$ implies plant whitening along uncoupling. Conversely, the ${}^{(a)}\mathbf{G}(z)$ matrix is used to generate the pressure mode errors from the individual sensor errors, again with a delay of Δ_{SVD} samples (Bai and Elliott, 2004). The second synthesis filter matrix ${}^{(s)}\mathbf{G}''(z)$ is created from the null space source modes to produce the fixed part of the AWFS solution. The delay block (Fig. 20) of $2\Delta_{\mathrm{SVD}}$ samples then represents the uncoupled plant model. This type of realization was more extensively described by Bai and Elliott (2004) for cross-talk cancelation. Also, because of the WFS departure penalty in the AWFS definition, an additional set of fixed filters produce the outputs for the $i > r$ (and $i \leqslant$)L WFS fixed source mode contributions $\tilde{q}_i$. This *a priori* solution creates a difference with the work of Bai and Elliott (2004).

All the preceding remarks should be subject of further investigations and experimental verifications. This is topic of current research on practical AWFS in a real reproduction space, that is including a reflective environment.

VIII. CONCLUSION AND PERSPECTIVES

This paper introduced AWFS in the physical domain for active sound field reproduction. AWFS is based on the minimization of the quadratic reproduction errors with a quadratic departure penalty from the WFS solution. Although the AWFS practical interest would be for room compensation, this theoretical paper was mainly concerned with the fundamental AWFS behavior via general analytical considerations and simple free-field simulations.

AWFS can be interpreted as an additive combination of (1) a fixed WFS solution and (2) an optimal regularized solution which seeks to minimize the WFS reproduction error. Contrary to WFS, the AWFS active contribution is spread over the entire source array.

SVD of the transfer impedance matrix shown that, for AWFS, independent control operates on each radiation mode. In the case of a four sensor array, this leads to independent control of a multipole expansion of the sound field at the error sensor locations. For a transfer impedance matrix including either a rigid wall reflection or room effects, the pressure modes still provide finite difference approximations of pressure, pressure gradients and tesseral quadrupole as pressure modes. In real situation, we thus expect that AWFS control will effectively and independently work on similar pressure modes at the sensor array since the analytical results of Sec. V B apply for any transfer impedance matrix $\mathbf{Z}$.

A practical usefulness of AWFS arises from the possibility to move from WFS to closed-loop modified "Ambisonics." AWFS converges toward (1) WFS when the regularization parameter λ is large and (2) closed-loop modified "Ambisonics" when λ is low. In the latter case, it was shown that low order source mode contributions approach an "Ambisonics" problem and that higher order source mode (creating the null space) contributions are imposed by the projection of the WFS solution on the null space basis of $\mathbf{Z}$.

SVD of the transfer impedance matrix suggests an independent radiation mode control implementation of AWFS. This could be achieved with $r \leqslant M$ independent LMS adaptive algorithms which would independently control the source mode contributions in a real adaptive control application. Work is in progress to implement AWFS using the currently developed AWFS adaptive algorithms for a configuration like the one shown in Fig. 2. AWFS performance in-room situation is being therefore investigated from a practical and signal processing view point.

Most of the analytical developments for AWFS can be applied to active noise or vibration control. To the authors knowledge, the use of *a priori* solution like $\mathbf{q}_{\mathrm{WFS}}$ have not been used for active control, this could be of potential interest. Eqs. (7) and (16) to (18) are also typically encountered in inverse problems theory (Hansen, 1998). Inverse problems are of importance in acoustics (Nelson, 2001; Gérard *et al.*, 2005), where the use of *a priori* solutions could also be applied.

ACKNOWLEDGMENTS

This work was supported by NSERC (Natural Sciences and Engineering Research Council of Canada), NATEQ (Fond Québecois de la Recherche sur la Nature et les Technologies), VRQ (Valorisation Recherche Québec), and Université de Sherbrooke and was conducted within CIRMMT (Centre for Interdisciplinary Research in Music Media and Technology, McGill University). The first author (P.-A. G.) acknowledges Anthony Gérard from Université de Sherbrooke for interesting discussions.

APPENDIX: "AMBISONICS" REVIEW

The target wave field $p^{(\mathrm{im})}(\mathbf{x},\omega)$ is defined as a plane wave in free field

$$p^{(\mathrm{im})}(R,\phi,\omega) = A(\omega)e^{jkR\cos(\phi-\theta)} \tag{A1}$$

with cylindrical coordinates (R,ϕ,z), and θ is the direction of propagation of the plane wave. This target plane wave field is represented by a Fourier-Bessel series,

$$p^{(\mathrm{im})}(R,\phi,\omega) = A(\omega)J_0(kR) + 2A(\omega)\sum_{n=1}^{\infty} j^n J_n(kR)\cos(n(\phi-\theta)), \quad \text{(A2)}$$

where J_n is the Bessel function of the first kind and order n. Rewriting Eq. (A2) gives

$$p^{(\mathrm{im})}(R,\phi) = AJ_0(kR) + 2A\sum_{n=1}^{\infty} j^n J_n(kR)\cos(n\theta)\cos(n\phi) + 2A\sum_{n=1}^{\infty} j^n J_n(kR)\sin(n\theta)\sin(n\phi). \quad \text{(A3)}$$

In practice, this series is truncated and expressed in vector form

$$p^{(\mathrm{im})}(R,\phi) = A\mathbf{c}^T\mathbf{h} \quad \text{(A4)}$$

with

$$\mathbf{c}^T = \sqrt{2}[1/\sqrt{2} \quad \cos(\theta) \quad \sin(\theta) \quad \cdots \quad \cos(n\theta) \quad \sin(n\theta) \quad \cdots] \quad \text{(A5)}$$

and

$$\mathbf{h}^T = \sqrt{2}[J_0(kR)/\sqrt{2} \quad j\cos(\phi)J_1(kR) \quad j\sin(\phi)J_1(kR) \quad \cdots \quad j^n\cos(n\phi)J_n(kR) \quad j^n\sin(n\phi)J_n(kR) \quad \cdots], \quad \text{(A6)}$$

where T denotes real transposition. Equation (A6) is independent of the target wave field definition while Eq. (A5) includes the target wave field direction θ. The Fourier-Bessel coefficients $\mathbf{c}$ are proportional to the outputs produced by harmonic directivity patterns on which the plane target field would impinge [see Eq. (A5)]. In $\mathbf{c}$, the first coefficient produces omnidirectional sensing, the second and the third produce perpendicular dipole patterns and this continues with higher order circumferential harmonics.

The reproduction system is assumed to be in free field and the reproduction sources (uniformly distributed over 2π radians) produce a set of plane waves at the origin of the coordinate system

$$p^{(\mathrm{rep})}(R,\phi) = \sum_{l=1}^{L} A_l e^{jkR\cos(\phi-\phi_l)}, \quad \text{(A7)}$$

where A_l is the unknown contribution of the plane reproduction wave l whose direction of propagation is ϕ_l. The reproduced field is also expressed as truncated Fourier-Bessel series

$$p^{(\mathrm{rep})}(R,\phi) = \sum_{l=1}^{L} A_l\mathbf{c}_l^T\mathbf{h} = \mathbf{A}^T\mathbf{C}^T\mathbf{h} \quad \text{(A8)}$$

with

$$\mathbf{A}^T = [A_1 \cdots A_L], \quad \text{(A9)}$$

$$\mathbf{C} = [\mathbf{c}_1 \cdots \mathbf{c}_L] \quad \text{(A10)}$$

$$\mathbf{c}_l = \sqrt{2}[1/\sqrt{2} \quad \cos(\phi_l) \quad \sin(\phi_l) \quad \cdots \quad \cos(n\phi_l) \quad \sin(n\phi_l) \quad \cdots]^T. \quad \text{(A11)}$$

The final step is the minimization of the sum of the quadratic reproduction error $|(\mathbf{A}\mathbf{c}^T - \mathbf{A}^T\mathbf{C}^T)\mathbf{h}|^2$, that is the reproduced wave field Fourier-Bessel coefficients $\mathbf{A}^T\mathbf{C}^T$ must approach, in a least-mean-square (LMS) sense, the target wave field Fourier-Bessel coefficients $A\mathbf{c}^T$. The minimization problem can readily be solved using the generalized pseudo-inverse (Golub and van Loan, 1996) denoted by †

$$\mathbf{A} = A\mathbf{C}^\dagger\mathbf{c} \quad \text{(A12)}$$

with

$$\mathbf{C}^\dagger = \mathbf{V}_c\boldsymbol{\Sigma}_c^+\mathbf{U}_c^H, \quad \text{(A13)}$$

where the SVD of the matrix $\mathbf{C}$ is $\mathbf{U}_c\Sigma_c\mathbf{V}_c^H$ and

$$[\boldsymbol{\Sigma}_c^+]^T = \begin{bmatrix} 1/\sigma_{1_c} & 0 & \cdots & & & 0 \\ 0 & 1/\sigma_{2_c} & 0 & \cdots & & 0 \\ & & \ddots & & & \\ 0 & \cdots & 0 & 1/\sigma_{N_c} & \cdots & 0 \end{bmatrix}_{N\times L}, \quad \text{(A14)}$$

where it was assumed that $N \leqslant L$. $\sigma_{1_c}, \ldots, \sigma_{N_c}$ are the singular values of $\mathbf{C}$. When all the "Ambisonics" reproduction sources are uniformly distributed over 2π radians, Eq. (A12) becomes

$$\mathbf{A} = A\frac{\mathbf{C}^T}{L}\mathbf{c} \quad \text{(A15)}$$

and this gives the following panning function (amplitude A_l of plane waves reproduction sources as function of θ and ϕ_l):

$$A_l = \frac{A}{L}\left[1 + 2\sum_{n=1}^{N}\cos(n(\theta-\phi_l))\right]. \quad \text{(A16)}$$

This corresponds to circular sinc function of ϕ_l.

AES Staff Writer (**2005**). "Multichannel audio systems and techniques," J. Audio Eng. Soc. **53**, 329–335.

Asano, F., and Swanson, D. C. (**1995**). "Sound equalization in enclosures using modal reconstruction," J. Acoust. Soc. Am. **98**, 2062–2069.

Bai, M. R., and Elliott, S. J. (**2004**). "Preconditioning multichannel adaptive filtering algorithms using EVD- and SVD-based signal prewhitening and system decoupling," J. Sound Vib. **270**, 639–655.

Bai, M. R., and Kwuen-Yieng, O. (**2005**). "Head-related transfer function (HRTF) synthesis based on a three-dimensional array model and singular valued composition," J. Sound Vib. **281**, 1093–1115.

Berkhout, A. J. (**1988**). "A holographic approach to acoustic control," J. Audio Eng. Soc. **36**, 977–995.

Berkhout, A. J., de Vries, D., and Vogel, P. (**1993**). "Acoustic control by wave field synthesis," J. Acoust. Soc. Am. **93**, 2764–2778.

Betlehem, T., and Abhayapala, T. D. (**2005**). "Theory and design of sound field reproduction in reverberant rooms," J. Acoust. Soc. Am. **117**, 2100–2111.

Blauert, J. (**1999**). *Spatial Hearing, The Psychophysics of Human Sound Localization* (MIT Press, Cambridge).

Cabell, R. H., and Fuller, C. R. (**1999**). "A principal component algorithm for feedforward active noise and vibration control," J. Sound Vib. **227**, 159–181.

Cabell, R. H., Palumbo, D., and Vipperman, J. (**2001**). "A principal component feedforward algorithm for active noise control: Flight test results,"

IEEE Trans. Control Syst. Technol. **9**, 76–83.
Corteel, E., Horbach, U., and Pellegrini, R. S. (**2002**). "Multichannel inverse filtering of multiexciter distributed mode loudspeakers for wave field synthesis," *Proceedings of the 112th AES Convention*, convention paper 5611.
Davis, M. F. (**2003**). "History of spatial coding," J. Audio Eng. Soc. **51**, 554–569.
Elliott, S. J. (**2000**). "Optimal controllers and adaptive controllers for multichannel feedforward control of stochastic disturbances," IEEE Trans. Signal Process. **48**, 1053–1060.
Elliott, S. (**2001**). *Signal Processing for Active Control* (Academic, London).
Epain, N., Friot, E., and Rabau, G. (**2004**). "Indoor sonic boom reproduction using ANC," *Proceedings of Active 2004*.
Fuster, L., López, J. J., González, A., and Zuccarello, P. D. (**2005**). "Room compensation using multichannel inverse filters for wavefield synthesis systems," AES 118th Convention, convention paper 6401.
Garas, J. (**1999**). *Adaptive 3D Sound Systems* (Technische Universiteit Eindhoven, Eindhoven).
Gauthier, P.-A., Berry, A., and Woszczyk, W. (**2005a**). "Sound-field reproduction in-room using optimal control techniques: simulations in the frequency domain," J. Acoust. Soc. Am. **117**, 662–678.
Gauthier, P.-A., Berry, A., and Woszczyk, W. (**2005b**). "Évaluation objective de la synthèse de champs acoustiques et recherches récentes sur la reproduction de champs acoustiques en salle," *73e Congrès de l'ACFAS*, Chicoutimi, Québec, Canada.
Gérard, A., Berry, A., and Masson, P. (**2005**). "Control of tonal noise from subsonic fans. Part 1: Reconstruction of aeroacoustic sources from far-field acoustic pressure," J. Sound Vib. **288**, 1049–1075.
Golub, G. H., and van Loan, C. F. (**1996**). *Matrix Computations* (John Hopkins University Press, Baltimore).
Hansen, C. (**1998**). *Rank-Deficient and Discrete Ill-Posed Problems* (SIAM, Philadelphia).
Hulsebos, E., Schuurmans, T., de Vries, D., and Boone, R. (**2003**). "Circular microphone array for discrete multichannel audio recording," *Proceedings of the 114th AES Convention*, convention paper 5716.
Ise, S. (**1999**). "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse system," Acta Acust. **85**, 78–87.
Junger, M. C., and Felt, D. (**1993**). *Sound, Structures, and Their Interactions* (Acoustical Society of America, Woodbury).
Kirkeby, O., and Nelson, P. A. (**1993**). "Reproduction of plane wave sound fields," J. Acoust. Soc. Am. **94**, 2992–3000.
Kirkeby, O., Nelson, P. A., Orduña-Bustamante, F., and Hamada, H. (**1996**). "Local sound field reproduction using digital signal processing," J. Acoust. Soc. Am. **100**, 1584–1593.
Lancaster, P., and Tismenetsky, M. (**1985**). *The Theory of Matrices* (Academic, Orlando).
Malham, D. G., and Myatt, A. (**1995**). "3-D sound spatialization using ambisonic techniques," Comput. Music J. **19**, 58–70.
Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).
Nelson, P. A. (**1994**). "Active control of acoustic fields and the reproduction of sound," J. Sound Vib. **177**, 447–477.
Nelson, P. A. (**2001**). "A review of some inverse problems in acoustics," Int. J. Acoust. Vib. **6**, 118–134.
Nelson, P. A. (**2002**). "Active control for virtual acoustics," *Proceedings of Active* 2002, pp. 67–90.
Nelson, P. A., and Elliott, S. J. (**1992**). *Active Control of Sound* (Academic, London).
Nelson, P. A., and Kahana, Y. (**2001**). "Spherical harmonics, singular-value decomposition and the head-related transfer function," J. Sound Vib. **239**, 607–637.
Nicol, R., and Emerit, M. (**1999**). "3D-sound reproduction over an extensive listening area: A hybrid method derived from holophony and ambisonic," *Proceedings of the AES 16th International Conference*, pp. 436–453.
Photiadis, D. M. (**1990**). "The relationship of singular value decomposition to wave-vector filtering in sound radiation problems," J. Acoust. Soc. Am. **88**, 1152–1159.
Pierce, A. D. (**1991**). *Acoustics: An introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY).
Poletti, M. A. (**2005**). "Effect of noise and transducer variability on the performance of circular microphone arrays," J. Audio Eng. Soc. **53**, 371–384.
Poulin, J. (**1957**). "Son et espace," *La revue musicale. Vers une musique expérimentale, sous la direction de Pierre Schaeffer, special number 236*, 105–114 (Paris: Richard-Masse).
Rumsey, F. (**2001**). *Spatial Audio* (Focal, Oxford).
Santillań, A. O. (**2001**). "Spatially extended sound equalization in rectangular rooms," J. Acoust. Soc. Am. **110**, 1989–1997.
Schobben, D. W. E., and Aarts, R. M. (**2005**). "Personalized multi-channel headphones sound reproduction based on active noise control," Acust. Acta Acust. **91**, 440–449.
Spors, S., Kuntz, A., and Rabenstein, R. (**2003**). "An approach to listening room compensation with wave field synthesis," *Proceedings of the AES 24th International Conference*, pp. 70–82.
Spors, S., Renk, M., and Rabenstein, R. (**2005**). "Limiting effects of active room compensation using wave field synthesis," AES 118th Convention, convention paper 6400.
Uto, S., and Hamada, H. (**1995**). "Transform domain adaptive audio equalizer for active control of sound field," *Proceedings of Active 95*.
Verheijen, E. N. G. (**1997**). Sound reproduction by wave field synthesis, Ph.D. thesis, Delft University of Technology, Delft.
Williams, E. G., Houston, B. H., and Herdic, P. C. (**2003**). "Fast Fourier transform and singular value decomposition formulations for patch near field acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333.
Woszczyk, W., Cooperstock, J., Roston, J., and Martens, W. (**2005**). "Shake, rattle, and roll: Getting immersed in multisensory, interactive music via broadband networks." J. Audio Eng. Soc. **53**, 336–344.

Transient response of a hollow cylinder with radial and axial material inhomogeneity

Michael El-Raheb
ATK Mission Research, Laguna Hills, California 92653

(Received 3 August 2005; revised 24 January 2006; accepted 13 February 2006)

The transient response of a hollow finite cylinder with radial and axial material inhomogeneity to a pulse of short duration is analyzed. The problem is solved in two steps. Radial inhomogeneity is treated first by dividing the cylinder into coaxial segments each with a constant modulus E which is allowed to vary from segment to segment. Transfer matrices relating variables at the two radial interfaces of a segment combine to satisfy continuity of stress and displacement at these interfaces. The Galerkin method is then utilized to treat the general case with both radial and axial inhomogeneity adopting the eigenfunctions of the radially inhomogeneous cylinder as trial functions. Features of both static and transient responses resemble that for a weakening material along the axis: displacement increases and stress reduces in proportion to the reduction in modulus. Spatial attenuation of transient stress cannot be replicated when geometric dimensionality is reduced from 3-D axisymmetric to 2-D plane-strain. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2184287]

PACS number(s): 43.40.At [DF] Pages: 2738–2749

I. INTRODUCTION

The dynamic response of solid and hollow elastic cylinders has been studied extensively in the literature as it applies to a variety of engineering and science problems. Vibrations of homogeneous hollow plane-strain cylinders were analyzed by Gasis[1] and Bird *et al.*[2] The frequency response of finite cylinders was analyzed by Grishenko and Meleshko,[3] Batard *et al.*,[4] and Grinchenko.[5] Soldatos and Ye[6] treated anisotropic laminated cylinders, and Hussein and Heyliger[7] considered layered piezoelectric cylinders. Cheung *et al.*[8] analyzed the 3-D vibration of solid and hollow cylinders by the Chebyshev-Ritz method. The free vibration of a hollow infinite multilayered cylinder was addressed by Sastry and Munjal[9] and Wang *et al.*[10] The free vibration of finite multilayered hollow cylinders was treated by Heyliger and Jilani[11] and Wang and Gong.[12] Whittier and Jones[13] studied the propagation of longitudinal and torsional waves in a bimaterial solid cylinder composed of an inner homogeneous core bonded to an outer homogeneous annular cylinder of different properties. Reuter[14] studied flexural waves in bimaterial cylinders. Keck and Armenakas[15] presented an exact solution for longitudinal waves in an infinitely long composite hollow cylinder made of three different transversely isotropic layers. The transient response of a multilayered hollow cylinder with time-dependent boundary conditions was treated by Lee[16] using a hybrid numerical method. Paul and Murali[17] determined the axisymmetric dynamic response of poro-elastic cylinders. Soldatos[18] presented a compilation of more than 150 references on the frequency response of solid and annular elastic cylinders. Yin and Yue[19] analyzed the plane-strain axisymmetric problem with multiple annular layers using Laplace transforms to integrate time dependence. Steinberg[20] formulated the inverse spectral problem to determine properties of a cylinder with inhomogeneous materials. No references were found treating the transient response of a finite hollow cylinder with both radial and axial inhomogeneity.

Analytical treatment of transient wave propagation in finite cylinders with radial and axial material inhomogeneity was not found in the literature. The reason is that these applications usually involve complicated geometries and constitutive laws that can be readily handled by general-purpose computer programs based on finite element, finite difference, finite volume, and smooth particle hydrodynamics (SPH). A comparison of results from these numerical programs with an independent analysis is needed to provide insight into the physics of the problem and phenomena that may not be gained from a purely numerical model. For transient phenomena following external loads of short duration, one simulation may take several hours on a fast workstation. To quickly validate results from these laborious models, engineers tend to simplify geometry and properties by considering limiting cases like that of a homogeneous plane-strain cylinder. Understanding the validity of these approximations is necessary since comparing results from a discretization model with narrowly valid approximations may lead to more uncertainty in the results. An overly simplified model does not always yield a valid result, yet a complicated discretization model needs consistent verification. The middle ground of an analysis that is neither a "black box" nor a simplification helps validate trends and, most importantly, provide understanding. Balancing the degree of simplification by including realistic physical effects in the analytical model rapidly increases its mathematical complexity.

Over the past 5 years, use of wave mechanics to explain mechanisms of tissue trauma has generated interest among researchers in bioengineering and medicine. Unfortunately, experimental measurements of tissue properties and their response to rapid transients with high strain rates are difficult if not impossible to achieve with acceptable accuracy. The

 0001-4966/2006/119(5)/2738/12/$22.50

standard predictive tool for this problem is a nonlinear finite element model. Comparison with a fundamentally different approach like analysis adds confidence to and understanding of results. While the analytical model is not a substitute for a general numerical model, it serves to understand and verify its results. When uncertainty arises from a peculiar numerical result, an independent analytical confirmation becomes necessary.

An analytical model for propagation of transient stress waves in tissue during projectile penetration was developed by El-Raheb.[21] The model consisted of a hollow homogeneous cylinder with radial velocity prescribed at the cylinder's inner boundary over a fixed axial footprint equal to the projectile's length. As long as the projectile's speed is much smaller than the speed of stress waves in the material, the approximation of a time-independent footprint is valid. Such is the case when projectile speed is 100 m/s while dilatational speed in tissue material is 1.7 km/s. However, for a projectile speed in excess of 500 m/s, projectile motion must be considered by allowing the footprint to lengthen with time at the projectile's speed. El-Raheb[22] extends the analysis in Ref. 21 to include a time-delayed prescribed radial displacement where length of the footprint varies with time.

This work addresses the effect on propagation of axial and radial material inhomogeneity from layers and radial tearing. Since real tissue inhomogeneity is too complicated to model, the analysis is restricted to axial symmetry. The problem is solved in two steps. First, radial or $E(r)$ inhomogeneity is treated in Sec. II by dividing the cylinder into coaxial segments each with constant modulus E_j that is allowed to vary from segment to segment. Transfer matrices relating variables at the two radial interfaces of a segment combine to satisfy continuity of stress and displacement at these interfaces. The Galerkin method is then utilized to treat the general problem with both radial and axial inhomogeneity or $E(r,z)$ adopting the eigenfunctions of the radially inhomogeneous cylinder as trial functions. The coupled problem is developed in Sec. III. Transient strain histories calculated with a nonlinear finite-element model compared favorably with the present analysis that is based on linear small displacement theory. This surprising result may be attributed to the tissue's comparative weakness.

Since the projectile's strength and acoustic impedance are much greater than those of tissue, the excitation transmitted over the inner cylinder boundary at the projectile-tissue interface can be approximated by prescribed radial and axial time-dependent motions rather than an unknown pressure excitation. In turn, the excited boundary is mixed meaning that part of the boundary over the footprint has a prescribed motion while the other part is traction-free. The influence method developed by El-Raheb[21] is employed to convert the segment of boundary where motion is prescribed to one where traction is prescribed. A summary of the steps is presented in Sec. VI.

The method of static-dynamic superposition developed by Berry and Naghdi[23] is applied to solve for transient response from each ring traction. The static solution and static-dynamic projection integral are developed in Secs. IV and V. This method accurately predicts stresses near and at the excited boundary and also converges rapidly with the number of terms in the modal expansion.

Section VII discusses results of the static and dynamic models. A comparison of histories from the different models explains how radial and axial inhomogeneities affect response.

Although the Galerkin method is standard, most treatments do not include the techniques in the present analysis and listed below:

(a) transfer matrices for segmenting radial inhomogeneity,
(b) influence method for modeling prescribed motion rather than prescribed pressure, and
(c) static-dynamic superposition for accurately representing the forcing function.

II. RADIAL INHOMOGENEITY

Consider a hollow cylinder with radial and axial coordinates (r,z) with origin on the cylinder's axis at its left face, inner and outer radii (r_p, r_o), length l, radial and axial displacements $\mathbf{u}=\{u,w\}^T$, Lamé constants (λ,μ), and mass density ρ. In all equations to follow subscripts to variables denote components and not partial derivatives. Also for conciseness, a bracket subscripted by j implies that in that bracket parameters λ,μ,k_e,k_s,C_k and functions φ_{mn},ψ_{mn} are also subscripted by j. Assume that Young's modulus $E(r,z)$ is the product of two functions in the form

$$E(r,z)=E_o\xi(r)\eta(z). \tag{1}$$

For radial inhomogeneity only, i.e., $\eta(z)=1$, the problem is solved analytically by dividing the region $r_p\leqslant r\leqslant r_o$ into N_r coaxial segments

$$r_j\leqslant r\leqslant r_{j+1},\quad 1\leqslant j\leqslant N_r,$$

$$\Delta r_j=r_{j+1}-r_j,\quad r_j=(r_o-r_p)/N_r. \tag{2}$$

Assume that $\xi(r)$ is constant over each segment but varies from segment to segment. Since axial symmetry holds, the linear elasto-dynamic equations of the jth segment are

$$\mathbf{D}_j\mathbf{u}_j-\rho\partial_{tt}\mathbf{u}_j=\mathbf{0},$$

$$D_{11,j}=[(\lambda+2\mu)\hat{\nabla}_1^2+\mu\partial_{zz}]_j,\quad D_{12,j}=(\lambda+\mu)_j\partial_{rz},$$

$$D_{21,j}=[(\lambda+\mu)\partial_z(\partial_r+1/r)]_j,$$

$$D_{22,j}=[\mu\hat{\nabla}_0^2+(\lambda+2\mu)\partial_{zz}]_j, \tag{3}$$

$$\lambda_j=\lambda_o\xi(r_j),$$

$$\mu_j=\mu_o\xi(r_j),\quad \hat{\nabla}_n^2\equiv\partial_{rr}+1/r\partial_r-n^2/r^2,\quad n=0,1.$$

λ_o,μ_o are Lamé constants that depend on E_o only, and t is time. For harmonic motions in time with radian frequency ω and simply supported boundary conditions at $z=0$ and $z=l$, Eq. (3) admits the solution (Ref. 21)

$$\{u(r,z,t),w(r,z,t)\}_j^T=\{\bar{u}(r)\sin(k_zz),\bar{w}(r)\cos(k_zz)\}_j^Te^{\hat{i}\omega t},$$

$$\bar{u}_j(r) = -k_{e,j}(C_1 J_1(k_e r) + C_2 Y_1(k_e r))_j + k_z(C_3 J_1(k_s r) + C_4 Y_1(k_s r))_j,$$

$$\bar{w}_j(r) = k_z(C_1 J_0(k_e r) + C_2 Y_0(k_e r))_j + k_{s,j}(C_3 J_0(k_s r) + C_4 Y_0(k_s r))_j, \tag{4a}$$

$$k_{e,j}^2 = \omega^2/c_{d,j}^2 - k_z^2, \quad c_{d,j}^2 = (\lambda + 2\mu)_j/\rho,$$
$$k_{s,j}^2 = \omega^2/c_{s,j}^2 - k_z^2, \quad c_{s,j}^2 = \mu_j/\rho. \tag{4b}$$

(c_d, c_s) are dilatational and shear speeds of sound, J_n and Y_n are Bessel functions of the first and second kind, $\hat{i}=\sqrt{-1}$, $k_z = m\pi/l$, where m is an integer axial wave number, and C_{kj} are independent constants in the solution of the jth coaxial segment.

The constitutive relations are

$$\sigma_{rr,j} = \lambda_j \Delta_j + 2\mu_j \partial_r u_j, \quad \sigma_{\theta\theta,j} = \lambda_j \Delta_j + 2\mu_j u_j/r,$$
$$\sigma_{zz,j} = \lambda_j \Delta_j + 2\mu_j \partial_z w_j, \quad \tau_{rz,j} = \mu_j(\partial_z u_j + \partial_r w_j), \tag{5}$$
$$\Delta_j = \partial_r u_j + u_j/r + \partial_z w_j.$$

Substituting (4a) in (5) yields

$$\begin{Bmatrix} \sigma_{rr} \\ \sigma_{\theta\theta} \\ \sigma_{zz} \\ \tau_{rz} \end{Bmatrix}_j (r,z,t) = \begin{Bmatrix} \bar{\sigma}_{rr}(r)\sin(k_z z) \\ \bar{\sigma}_{\theta\theta}(r)\sin(k_z z) \\ \bar{\sigma}_{zz}(r)\sin(k_z z) \\ \bar{\tau}_{rz}(r)\cos(k_z z) \end{Bmatrix}_j e^{\hat{i}\omega t}, \tag{6a}$$

$$\begin{aligned}\bar{\sigma}_{rr,j}(r) = {} & [-((\lambda+2\mu)k_e^2 + \lambda k_z^2)J_0(k_e r) \\ & + 2\mu k_e^2 J_1(k_e r)/(k_e r)]_j C_{1j} + [-((\lambda+2\mu)k_e^2 \\ & + \lambda k_z^2)Y_0(k_e r) + 2\mu k_e^2 Y_1(k_e r)/(k_e r)]_j C_{2j} \\ & + 2\mu_j k_{s,j} k_z [J_0(k_s r) - J_1(k_s r)/(k_s r)]_j C_{3j} \\ & + 2\mu_j k_{s,j} k_z [Y_0(k_s r) - Y_1(k_s r)/(k_s r)]_j C_{4j},\end{aligned} \tag{6b}$$

$$\begin{aligned}\bar{\sigma}_{\theta\theta,j}(r) = {} & -[\lambda(k_z^2 + k_e^2)J_0(k_e r) + 2\mu k_e^2 J_1(k_e r)/(k_e r)]_j C_{1j} \\ & -[\lambda(k_z^2 + k_e^2)Y_0(k_e r) + 2\mu k_e^2 Y_1(k_e r)/(k_e r)]_j C_{2j} \\ & + 2\mu_j k_{s,j} k_z [C_3 J_1(k_s r) + C_4 Y_1(k_s r)]_j/(k_{s,j} r),\end{aligned} \tag{6c}$$

$$\begin{aligned}\bar{\sigma}_{zz,j}(r) = {} & -((\lambda+2\mu)k_z^2 + \lambda k_e^2)_j [C_1 J_0(k_e r) + C_2 Y_0(k_e r)]_j \\ & - 2\mu_j k_{s,j} k_z [C_3 J_0(k_s r) + C_4 Y_0(k_s r)]_j,\end{aligned} \tag{6d}$$

$$\begin{aligned}\bar{\tau}_{rz,j}(r) = {} & -2\mu_j k_{e,j} k_z [C_1 J_1(k_e r) + C_2 Y_1(k_e r)]_j \\ & - \mu_j(k_{s,j}^2 - k_z^2)[C_3 J_1(k_s r) + C_4 Y_1(k_s r)]_j.\end{aligned} \tag{6e}$$

The functional form along z in (4a) and (6a) assumes that $\sigma_{zz} \equiv u = 0$ at the two axial ends of the cylinder $z = 0, l$. Although this condition is an approximation to a traction-free boundary where $\sigma_{zz} \equiv \tau_{rz} = 0$ at $z = 0, l$, nevertheless the effect of boundary conditions on stress waves remote from these boundaries is negligible prior to reflections from these boundaries, especially when the material's sound speed is low.

For each coaxial segment, express the state vector $\mathbf{S}_j = \{\bar{\sigma}_{rr}, \bar{\tau}_{rz}, \bar{u}, \bar{w}\}_j^T$ in terms of the constant vector $\mathbf{C}_j = \{C_1, C_2, C_3, C_4\}_j^T$:

$$\mathbf{S}_j(r) = \mathbf{B}_j(r)\mathbf{C}_j. \tag{7}$$

$\mathbf{B}_j(r)$ is a matrix with coefficients the functions multiplying $\{C_1, C_2, C_3, C_4\}_j^T$ in (4) and (6). Evaluating (7) at the two interfaces $r = r_j$ and $r = r_{j+1}$ of the jth segment and then eliminating $\mathbf{C}_j$ determines the 4×4 transfer matrix $\mathbf{T}_j$ relating state vectors at the two interfaces of a segment:

$$\mathbf{S}_j(r_{j+1}) = \mathbf{T}_j \mathbf{S}_j(r_j), \quad \mathbf{T}_j = \mathbf{B}_j^{-1}(r_j)\mathbf{B}_j(r_{j+1}), \tag{8a}$$

$$\mathbf{C}_j = \mathbf{B}_j^{-1}(r_j)\mathbf{S}_j(r_j). \tag{8b}$$

Enforcing continuity of $\mathbf{S}_j$ at interfaces of segments and homogeneous traction boundary conditions at $r = r_p$ and $r = r_o$,

$$\bar{\sigma}_{rr}(r_p) = \bar{\tau}_{rz}(r_p) = 0,$$
$$\bar{\sigma}_{rr}(r_o) = \bar{\tau}_{rz}(r_o) = 0, \tag{9}$$

yields the global transfer matrix $\mathbf{T}_G$ in tri-diagonal block form and the corresponding global $\mathbf{S}_G$ which is the ensemble of all $\mathbf{S}_j$:

$$\mathbf{T}_G \cdot \mathbf{S}_G = \mathbf{0} \Rightarrow \det|\mathbf{T}_G| = 0,$$
$$\mathbf{S}_G = \{\mathbf{S}_1(r_1), \mathbf{S}_2(r_2), \ldots, \mathbf{S}_j(r_j), \ldots, S_{N_r}(r_{N_r})\}^T. \tag{10}$$

Equation (10) determines the eigenset $\{\mathbf{S}_G; \omega\}$ and in turn $\mathbf{C} = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_j, \ldots, \mathbf{C}_{N_r}\}^T$ from (8b). For the jth coaxial segment, the eigenfunction expansions for u_j and w_j take the form

$$u_j(r,z,t) = \sum_n \sum_m a_{mn}(t)\varphi_{mn,j}(r)\sin(k_{zm} z),$$
$$w_j(r,z,t) = \sum_n \sum_m a_{mn}(t)\psi_{mn,j}(r)\cos(k_{zm} z). \tag{11}$$

$\phi_{mn,j} = \{\varphi_{mn}, \psi_{mn}\}_j^T$ is the displacement subvector of $\mathbf{S}_j$ corresponding to the nth eigenfunction with m axial half waves of the jth segment while their functional form is given by (4a).

III. RADIAL AND AXIAL INHOMOGENEITIES

When adding axial inhomogeneity, the equations of the jth coaxial segment become

$$\mathbf{D}_j \mathbf{u}_j + \tilde{\mathbf{D}}_j \mathbf{u}_j - \rho \partial_{tt} \mathbf{u}_j = \mathbf{0}. \tag{12}$$

$\mathbf{D}_j$ is similar to the differential operator matrix defined in (3) with the constants λ_j and μ_j replaced by the functions $\lambda_j \eta(z)$ and $\mu_j \eta(z)$, and $\tilde{\mathbf{D}}_j$ is the differential operator matrix defined by

$$\tilde{D}_{11,j} = \mu_j \eta'(z)\partial_z, \quad \tilde{D}_{12,j} = \mu_j \eta'(z)\partial_r,$$
$$\tilde{D}_{21,j} = \lambda_j \eta'(z)(\partial_r + 1/r), \quad \tilde{D}_{22,j} = (\lambda_j + 2\mu_j)\eta'(z)\partial_z. \tag{13}$$

$(\,)'$ is the derivative with respect to the argument. The Galerkin method is utilized to solve (12). Eigenfunctions of the radially inhomogeneous cylinder developed in Sec. II are

utilized as trial functions. Substituting (11) in (12) and noting that $\mathbf{D}_j\boldsymbol{\phi}_{mn,j}=-\rho\omega_{mn}^2\boldsymbol{\phi}_{mn,j}$ yields equations of the jth coaxial segment

$$\sum_l\sum_k [-\rho(\ddot{a}_{kl}+\eta(z)\omega_{kl}^2 a_{kl})\varphi_{kl,j}\sin(k_{zk}z) + a_{kl}\mu_j(k_{zk}\varphi_{kl}+\psi'_{kl})_j\eta'(z)\cos(k_{zk}z)] = p_r\bar{\delta}(r-r_p)(H(z-z_1)-H(z-z_2))f_p(t), \tag{14a}$$

$$\sum_l\sum_k [-\rho(\ddot{a}_{kl}+\eta(z)\omega_{kl}^2 a_{kl})\psi_{kl,j}\cos(k_{zk}z) + a_{kl}(\lambda(\varphi'_{kl}+\varphi_{kl}/r) - (\lambda+2\mu)k_{zk}\psi_{kl})_j\eta'(z)\sin(k_{zk}z)] = 0. \tag{14b}$$

$\dot{(\,)}$ is time derivative. a_{kl} and ω_{kl} are not subscripted by j because these modal quantities are the same for all segments. Multiplying (14a) by $\varphi_{mn,j}\sin(k_{zm}z)$ and (14b) by $\psi_{mn,j}\cos(k_{zm}z)$, adding the two equations, and then integrating over the cylinder volume including all segments while enforcing orthogonality of the $\{\varphi_{mn}(r),\psi_{mn}(r)\}^T$ set produces

$$\mathbf{Ka}+\mathbf{M}\ddot{\mathbf{a}}=\mathbf{0}, \tag{15a}$$

$$M_{mn,kl}=-0.5\rho l\delta_{mn,kl}\sum_j\int_{r_j}^{r_{j+1}}(\varphi_{mn}^2+\psi_{mn}^2)_j r\,dr \tag{15b}$$

$$K_{mn,kl}=\sum_{q=1}^{4}R_{mn,kl}^{(q)}Z_{mk}^{(q)}, \tag{15c}$$

$$\begin{aligned}
R_{mn,kl}^{(1)}&=-\rho\sum_j\int_{r_j}^{r_{j+1}}\omega_{kl}^2\varphi_{kl,j}\varphi_{mn,j}r\,dr,\\
R_{mn,kl}^{(2)}&=-\rho\sum_j\int_{r_j}^{r_{j+1}}\omega_{kl}^2\psi_{kl,j}\psi_{mn,j}r\,dr,\\
Z_{mk}^{(1)}&=\int_0^l\eta(z)\sin(k_{zk}z)\sin(k_{zm}z)dz,\\
Z_{mk}^{(2)}&=\int_0^l\eta(z)\cos(k_{zk}z)\cos(k_{zm}z)dz,\\
R_{mn,kl}^{(3)}&=\sum_j\int_{r_j}^{r_{j+1}}\mu_j(k_{zk}\varphi_{kl}+\psi'_{kl})_j\varphi_{mn,j}r\,dr,\\
R_{mn,kl}^{(4)}&=\sum_j\int_{r_j}^{r_{j+1}}[\lambda(\varphi'_{kl}+\varphi_{kl}/r)-(\lambda+2\mu)k_{zk}\psi_{kl}]_j\psi_{mn,j}r\,dr,\\
Z_{mk}^{(3)}&=\int_0^l\eta'(z)\cos(k_{zk}z)\sin(k_{zm}z)dz,\\
Z_{mk}^{(4)}&=\int_0^l\eta'(z)\sin(k_{zk}z)\cos(k_{zm}z)dz=Z_{km}^{(3)}.
\end{aligned} \tag{15d}$$

$\delta_{mn,kl}$ is the Kronecker delta. Equation (15a) is diagonalized by solving the eigenvalue problem

$$[\mathbf{K}-\mathbf{M}\omega^2]\mathbf{a}=\mathbf{0}. \tag{16}$$

This determines the coupled eigenfrequencies ω_{ci}, coupling coefficients a_{mni}, and orthogonal set $\{\Phi_{ci},\Psi_{ci}\}$ of the ith coupled mode. In terms of coupled modes, the eigenfunction expansion takes the form

$$\begin{aligned}
u(r,z,t)&=\sum_i c_i(t)\Phi_{ci}(r,z),\\
w(r,z,t)&=\sum_i c_i(t)\Psi_{ci}(r,z),
\end{aligned} \tag{17a}$$

$$\begin{aligned}
\Phi_{ci}(r,z)&=\sum_n\sum_m a_{mn,i}\sum_j\varphi_{mn,j}(r)(H(r-r_j)-H(r-r_{j+1}))\sin(k_{zm}z),\\
\Psi_{ci}(r,z)&=\sum_n\sum_m a_{mn,i}\sum_j\psi_{mn,j}(r)(H(r-r_j)-H(r-r_{j+1}))\cos(k_{zm}z).
\end{aligned} \tag{17b}$$

Consider the inhomogeneous boundary conditions at $r=r_p$:

$$\begin{aligned}
\sigma_{rr}(r_p,z;t)&=p_r[H(z-z_1)-H(z-z_2)]f_p(t),\\
\tau_{rz}(r_p,z;t)&=0,\\
\sigma_{rr}(r_o,z;t)&=\tau_{rz}(r_o,z;t)=0.
\end{aligned} \tag{18}$$

H is the Heaviside function, $f_p(t)$ is the time dependence of the forcing function, and p_r is the magnitude of the uniform applied pressure. To satisfy (23), the static-dynamic superposition method developed by Berry and Naghdi[23] is utilized. Total displacement is expressed as a superposition of two solutions

$$\mathbf{u}_T(r,z;t)=\mathbf{u}_s(r,z)f_p(t)+\mathbf{u}(r,z,t). \tag{19}$$

$\mathbf{u}_s$ is static displacement satisfying the inhomogeneous boundary conditions

$$\begin{aligned}
\sigma_{rrs}(r_p,z)&=[H(z-z_1)-H(z-z_2)],\\
\tau_{rzs}(r_p,z)&=0,\\
\sigma_{rrs}(r_o,z)&=0,\quad \tau_{rzs}(r_o,z)=0
\end{aligned} \tag{20}$$

(see Sec. V), and $\mathbf{u}$ is the dynamic eigenfunction expansion (17a). In this way, the first term yields finite σ_{rr} at $r=r_p$. In addition to accuracy of variables at and near the boundary, this method converges rapidly. Substituting (19) in (15a) and enforcing orthogonality of $\{\Phi_{ci},\Psi_{ci}\}$ determines uncoupled equations in $c_i(t)$:

$$\ddot{c}_i(t)+\omega_{ci}^2c_i(t)=p_rN_{cfi}\ddot{f}_p(t)/N_{cii}, \tag{21a}$$

$$\begin{aligned}
N_{cii}&=\rho\int_0^l\int_{r_p}^{r_o}(\Phi_{ci}^2+\Psi_{ci}^2)r\,dr\,dz\\
&=0.5\rho l\sum_n\sum_m a_{mn,i}^2\sum_j\int_{r_j}^{r_{j+1}}(\varphi_{mn}^2+\psi_{mn}^2)_j r\,dr,
\end{aligned}$$

$$N_{cfi} = -\rho \sum_j \int_{r_j}^{r_{j+1}} \int_0^l (\Phi_{ci} u_s + \Psi_{ci} w_s) r \, dr \, dz. \tag{21b}$$

The static-dynamic projection integral N_{cfi} is derived in Sec. V.

The effect of material visco-elasticity may be approximated in Eq. (21a) by including a velocity proportional term factored by frequency-dependent viscous damping coefficients $2\zeta_i(\omega_{ci})\dot{c}_i(t)$. A rigorous treatment of dynamic visco-elasticity would require computing a complex stiffness matrix **K** in Eq. (16) as it includes a complex frequency-dependent modulus. This yields the evaluation of a complex eigenvalue problem (see Ref. 24) which is beyond the scope of the present study.

IV. STATIC SOLUTION

The static-dynamic superposition method expressed by Eq. (19) requires the static solution $\mathbf{u}_s$ of the inhomogeneous cylinder. A method similar to the dynamic solution in Sec. III is adopted to solve the static problem. For the jth coaxial segment, express displacement as

$$u_{s,j}(r,z) = \left[u_{sr}(r,z) + \sum_n \sum_m \tilde{a}_{smn} \varphi_{mn}(r) \sin(k_{zm} z) \right]_j,$$
$$w_{s,j}(r,z) = \left[w_{sr}(r,z) + \sum_n \sum_m \tilde{a}_{smn} \psi_{mn}(r) \cos(k_{zm} z) \right]_j, \tag{22}$$

where static displacement (u_{sr}, w_{sr}) and eigenfunction $(\varphi_{kl}, \psi_{kl})$ are those of the radially inhomogeneous cylinder. The static equations are

$$[\mathbf{D}_j + \tilde{\mathbf{D}}_j] \mathbf{u}_{s,j} = \mathbf{0}. \tag{23}$$

The differential operators $\mathbf{D}_j$ and $\tilde{\mathbf{D}}_j$ are given by Eqs. (3) and (13), respectively. Substituting (28) in (29) yields

$$\sum_l \sum_k \tilde{a}_{skl} [-\rho \eta(z) \omega_{kl}^2 \varphi_{kl,j} \sin(k_{zk} z) + \mu_j (k_{zk} \varphi_{kl} + \psi'_{kl})_j \eta'(z) \cos(k_{zk} z)] = -[\tilde{D}_{11} u_{sr} + \tilde{D}_{12} w_{sr}]_j, \tag{24a}$$

$$\sum_l \sum_k \tilde{a}_{skl} [-\rho \eta(z) \omega_{kl}^2 \psi_{kl,j} \cos(k_{zk} z) + (\lambda(\varphi'_{kl} + \varphi_{kl}/r) - (\lambda + 2\mu) k_{zk} \psi_{kl})_j \eta'(z) \sin(k_{zk} z)] = -[\tilde{D}_{21} u_{sr} + \tilde{D}_{22} w_{sr}]_j. \tag{24b}$$

Note that in (24) $\mathbf{D}_j \mathbf{u}_{srj} = \mathbf{0}$ since $\mathbf{u}_{srj}$ is the static solution of the radially inhomogeneous cylinder. Multiplying (24a) by $\varphi_{mn,j} \sin(k_{zm} z)$, (24b) by $\psi_{mn,j} \cos(k_{zm} z)$, integrating over the cylinder's volume, and then adding the two equations produces

$$\mathbf{K} \tilde{\mathbf{a}}_s = \mathbf{p}_s, \tag{25}$$

where the stiffness matrix **K** is given by (15c) and the right-hand vector $\mathbf{p}_s$ is defined by

$$p_{s,mn} = -\sum_j \int_0^l \int_{r_j}^{r_{j+1}} [\tilde{D}_{11} u_{sr} + \tilde{D}_{12} w_{sr}]_j \varphi_{mn,j} \times \sin(k_{zm} z) r \, dr \, dz - \sum_j \int_0^l \int_{r_j}^{r_{j+1}} \times [\tilde{D}_{21} u_{sr} + \tilde{D}_{22} w_{sr}]_j \psi_{mn,j} \cos(k_{zm} z) r \, dr \, dz. \tag{26}$$

Equation (25) yields the coupling coefficient vector $\mathbf{a}_s = \{a_{skl}\}_j^T$. Static displacement (u_{sr}, w_{sr}) of a coaxial segment is derived by El-Raheb[1] and given below for completeness:

$$u_{sr,j}(r,z) = \sum_m \varphi_{srm,j}(r) \sin(k_{zm} z), \tag{27a}$$
$$\varphi_{srm,j}(r) = [C_1 I_1(k_z r) + C_2 K_1(k_z r) + C_3(k_z r I_0(k_z r) - I_1(k_z r)) - C_4(k_z r K_0(k_z r) + K_1(k_z r))]_{m,j},$$
$$w_{sr,j}(r,z) = \sum_m \psi_{srm,j}(r) \cos(k_{zm} z),$$
$$\psi_{srm,j}(r) = [C_1 I_0(k_z r) - C_2 K_0(k_z r) + C_3(\alpha_1 I_0(k_z r) + k_z r I_1(k_z r)) + C_4(-\alpha_1 K_0(k_z r) + k_z r K_1(k_z r))]_{m,j}, \tag{27b}$$
$$\alpha_{1j} = (\lambda + 3\mu)_j / (\lambda + \mu)_j.$$

I_n and K_n are the modified Bessel functions of the first and second kind. The transfer matrix approach utilized in Sec. II is applied to the radially inhomogeneous static problem. Since the z functions are orthogonal, solutions for each k_{zk} decouple and coefficients $C_{q,k}$ $(q=1,4)$ are determined by satisfying boundary conditions (20).

V. STATIC-DYNAMIC PROJECTION INTEGRAL

Substituting (17), (22), and (27) in (21b) and enforcing orthogonality of the z functions yields an expression for N_{cfi} in terms of constituent functions, i.e., modes and static solution of the radially inhomogeneous problem

$$N_{cfi} = \sum_n \sum_m (a_{mn,i} N_{rfmn} - \tilde{a}_{smn} a_{mn,i} N_{rmn}),$$
$$N_{rfmn} = -0.5 \rho l \sum_j \int_{r_j}^{r_{j+1}} (\varphi_{srm} \varphi_{mn} + \psi_{srm} \psi_{mn})_j r \, dr, \tag{28}$$
$$N_{rmn} = 0.5 \rho l \sum_j \int_{r_j}^{r_{j+1}} (\varphi_{mn}^2 + \psi_{mn}^2)_j r \, dr.$$

N_{rfmn} and N_{rmn} are the static-dynamic projection integral and generalized mass of the cylinder with purely radial inhomogeneity developed in Sec. II. This is the limit when $\tilde{a}_{smn} = 0$ and $a_{mn,i} = \delta_{mn,i}$ in (28), and with generalized coordinates a_{mn} in (11) satisfying

$$\ddot{a}_{mn}(t) + \omega_{mn}^2 a_{mn}(t) = p_r N_{rfmn} f_p(t) / N_{rmn}. \tag{29}$$

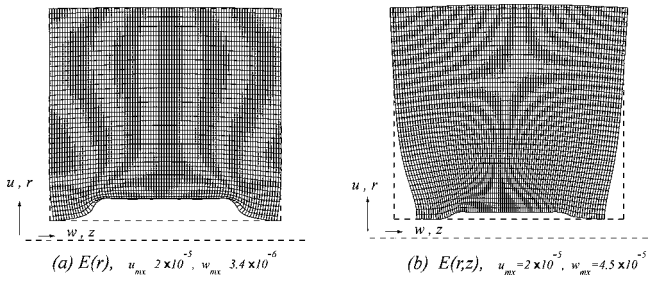


FIG. 1. Static deformation from prescribed u at $r=r_p$. (a) $E=E(r)$. (b) $E=E(r,z)$.

VI. PRESCRIBED MOTION WITH TIME DELAY

When the forcing function is a radial motion prescribed over part of the inner cylindrical boundary while the remaining part of the boundary is traction-free, the problem converts to a mixed boundary value whose solution is briefly outlined below while details may be found in Ref. 21.

Divide the cylindrical surface $\{r=r_p, z_a \leqslant z \leqslant z_b\}$ into n_p+1 equidistant ring stations with constant increment Δz_p

$$z_1, z_2, \ldots, z_l, \ldots, z_{n_p+1}, \quad z_l = z_a + (l-1)\Delta z_p,$$
$$\Delta z_p = (z_b - z_a)/n_p, \quad z_1 = z_a, \quad z_{n_p+1} = z_b. \tag{30}$$

Assume a uniform pressure of unit intensity to act over each ring segment $z_l \rightarrow z_{l+1}$ with arbitrary time dependence $\tilde{f}_p(t)$. For the kth pressure segment, evaluating $u_{lk}(r_p, z_{cl}; t) \equiv U_{clk}(t)$ at the central point $z_{cl}=(z_l+z_{l-1})/2$ of the lth segment yields coefficients of the influence matrix $U_{clk}(t)$. Total displacement at the boundary $r=r_p$ is then

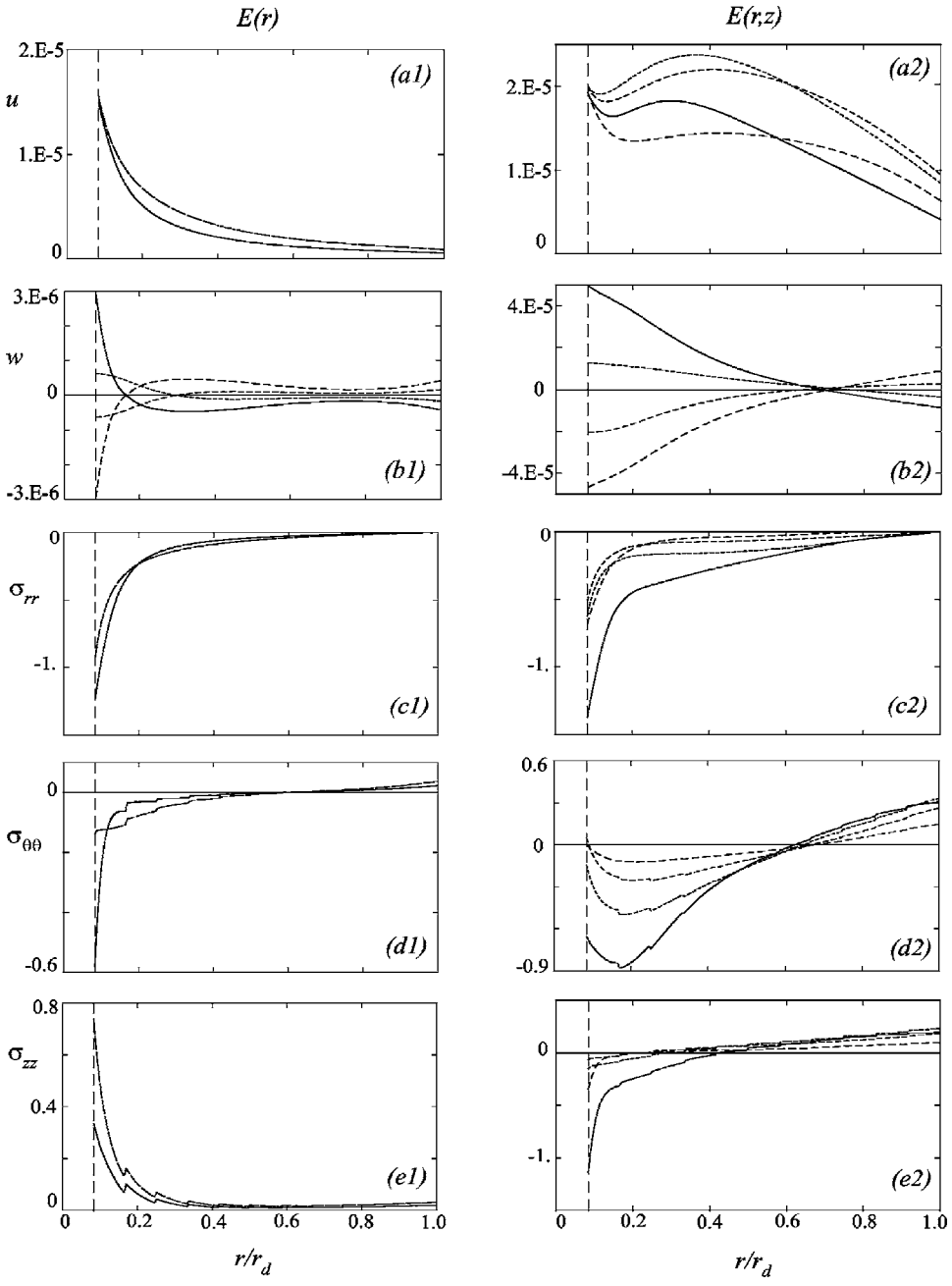


FIG. 2. Distribution of static variables along r from prescribed u at $r=r_p$. (a1)–(e1) $E=E(r)$. (a2)–(e2) $E=E(r,z)$.———z/l=0.2, --- 0.4, - - - - 0.6, - - - 0.8.

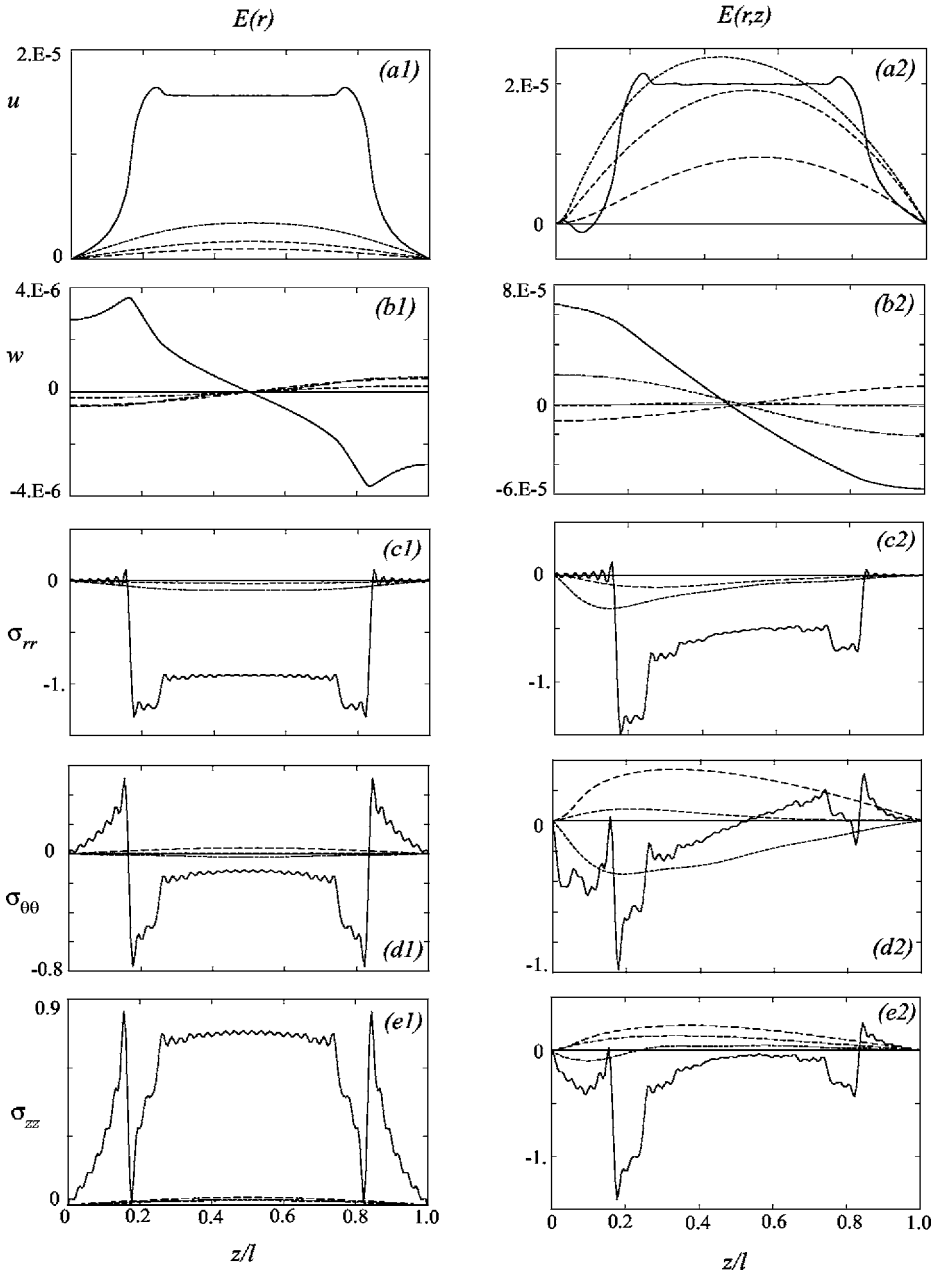


FIG. 3. Distribution of static variables along z from prescribed u at $r=r_p$. (a1)–(e1) $E=E(r)$. (a2)–(e2) $E=E(r,z)$.———$r/r_p=1$, --- 5, - - - - 8, - - - 12.

$$u(r_p,z_{cl};t)=\sum_{k=1}^{n} U_{clk}(t)\tilde{p}_k(t), \quad 1\leqslant l\leqslant n. \tag{31}$$

$\tilde{p}_k(t)$ are time-dependent weights determined from the constraint that $u(r_p,z_{cl};t)=u_p(t)$ where $u_p(t)$ is instantaneous prescribed displacement. This superposition applies to all dependent variables $\mathbf{S}$,

$$\mathbf{S}(r,z;t)=\sum_{k=1}^{n} \mathbf{S}_{\mathbf{k}}(r,z;t)\tilde{p}_k(t), \quad 1\leqslant l\leqslant n, \tag{32}$$

where $\mathbf{S}_k$ is the value of the variable from the kth pressure segment. Since $\tilde{f}_p(t)$ is arbitrary, numerical experiments have shown that a simple ramp is appropriate:

$$\tilde{f}_p(t)=(t/\Delta t_1)[H(t)-H(t-\Delta t_1)]+H(t-\Delta t_1), \tag{33}$$

where Δt_1 is a rise time in the prescribed motion profile. Enforcing the condition of prescribed displacement $u_p(t)$ at each time step yields a set of simultaneous equations in the instantaneous weights $\tilde{p}_k(t)$:

$$\sum_{k=1}^{n_p} U_{clk}(t)\tilde{p}_k(t)=u_p(t), \quad 1\leqslant l\leqslant n_p. \tag{34}$$

When the projectile moves, the axial length of the footprint varies instantaneously. Then, $f_p(t)$ in (21a) is modified to a time-delayed form that depends not only on t but also on z as follows:

$$f_p(\hat{t}), \quad \hat{t} = (t - z/V_p)H(t - z/V_p). \tag{35}$$

V_p is projectile speed. The significance of (35) is that at some station z, $f_p(\hat{t})$ acts only when $t \geqslant z/V_p$, otherwise it vanishes, thus the term "time-delayed."

VII. RESULTS

Before discussing results of tissue response, a comparison of the present analysis with an example by Yin and Yue[19] is appropriate although their analysis concerns the special case of a plane-strain bi-material cylinder. In this example, the inner annulus is made of steel and the outer annulus of aluminum. Inner radius a_1 and thickness of each layer h_1, h_2 are all equal. The ratio of extensional propagation speeds is $c_{d1}/c_{d2}=1.1$, the density ratio is $\rho_1/\rho_2=2.724$, and the Poisson ratio for both layers is $\nu=0.25$. An axisymmetric prescribed pressure with a step function time dependence $p(t)=p_oH(t)$ is applied at the inner perimeter $r=a_1$ while the outer perimeter is traction-free, where $H(t)$ is the Heaviside function. Plots of normalized radial displacement $\tilde{u}=u(\lambda_1+2\mu_1)/(a_1p_o)$ and normalized circumferential stress $\tilde{\sigma}_{\theta\theta}=\sigma_{\theta\theta}/(2(1-\nu)p_o)$ along the normalized radius $\tilde{r}=(r-a_1)/(2h_1)$ at the normalized time $\tilde{t}\equiv t/(h_1/c_{d1}+h_2/c_{d2})=1$ are identical to the solid thick lines in the second and third plots of Fig. 2 of Ref. 19. In the present analysis, 150 terms were used in the modal expansion. This large number is needed since in this example the rise time of $H(t)$ is zero, meaning that its frequency content is infinite. In spite of this demanding requirement, the present analysis converges rapidly. Although this comparison is limited to the case of plane strain, it validates the radial dependence of a segmented infinite cylinder adopting the transfer matrix method.

In all numerical results to follow, the geometry and material properties of the hollow gelatin cylinder are

$$r_p = 0.635 \text{ cm } (=0.25 \text{ in.}), \quad r_o = 7.62 \text{ cm } (=3 \text{ in.}), \tag{36a}$$

$$l = 7.62 \text{ cm } (=3 \text{ in.}),$$

$$E_o = 3.1 \times 10^9 \text{ dyn/cm}^2 \ (=4.5 \times 10^4 \text{ lb/in.}^2)$$

$$\rho = 0.93 \text{ g/cm}^3 \ (=8.7 \times 10^{-5} \text{ lbs}^2/\text{in.}^4), \quad \nu = 0.48 \tag{36b}$$

$$c_d = 1.71 \text{ km/s } (=6.74 \times 10^4 \text{ in./s}),$$

$$c_s = 0.34 \text{ km/s} (=1.32 \times 10^4 \text{ in./s}).$$

E_o is Young's modulus of the homogeneous cylinder. This low modulus and the choice of $\nu=0.48$ for an almost incompressible material are consistent with a gelatin simulent of biological tissue used in experiments. Motion prescribed over the inner boundary extends over the interval $z_a \leqslant z \leqslant z_b$ where $z_a=1.27$ cm (=0.5 in.) and $z_b=6.35$ cm (=2.5 in.). This interval is divided into 14 equidistant ring segments needed in the influence method.

Functions $\xi(r)$ and $\eta(z)$ in Eq. (1) defining radial and axial material inhomogeneity are given by

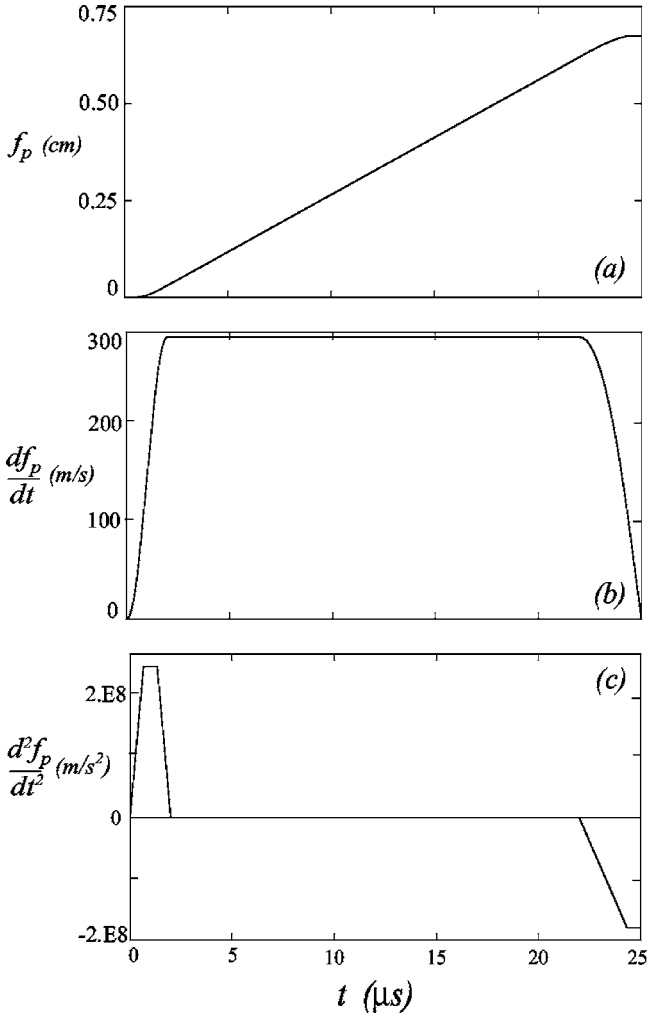


FIG. 4. Prescribed motion. (a) f_p, (b) df_p/dt, and (c) d^2f_p/dt^2.

$$E(r,z) = E_o\xi(r)\eta(z),$$

$$\xi(r) = \alpha + (1-\alpha)(r - r_p)/(r_o - r_p), \quad \alpha = 0.444, \tag{36c}$$

$$\eta(z) = 1 - 0.75(z/l) + 0.18(z/l)^2.$$

$\xi(r)$ is discretized into $N_r=12$ steps of equal interval

$$\xi(r) = \xi(r_{cj})[H(r - (j-1)\Delta r) - H(r - j\Delta r)],$$

$$(j-1)\Delta r \leqslant r \leqslant j\Delta r, \tag{36d}$$

$$r_{cj} = (2j-1)\Delta r/2, \quad \Delta r = (r_o - r_p)/N_r, \quad 1 \leqslant j \leqslant N_r.$$

The reduced modulus at $r=r_p$ accounts for radial tearing caused by the projectile's motion weakening the material locally. The reduction in modulus axially accounts for weaker tissue behind muscle.

Since the static solution is needed for determining transient response, static results are discussed first. Figures 1(a) and 1(b) show the static deformed cylinder's generator for prescribed radial displacement with $E(r)$ and $E(r,z)$. Note that for the same magnitude of prescribed u_o, w_{mx} from $E(r,z)$ is an order of magnitude larger than that from $E(r)$ because of reduced stiffness along the axis. At $r=r_p$, w is such that it shortens cylinder length while at $r=r_o$ the reverse

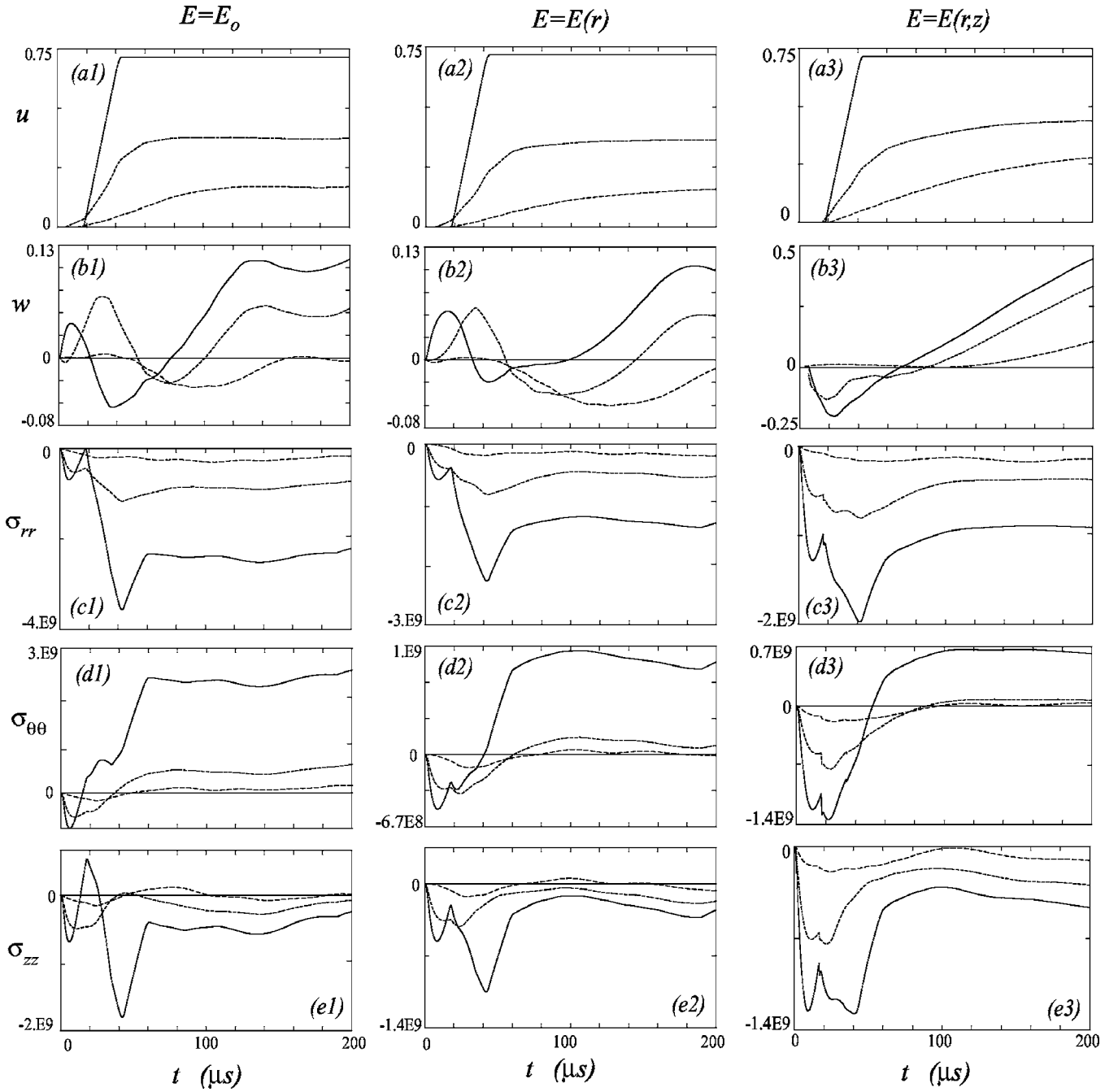


FIG. 5. Histories of cylinder at $z=2$ cm.———$r/r_p=1$, --- 2, - - - 4. (a1)–(e1) $E=E_o$. (a2)–(e2) $E=E(r)$. (a3)–(e3) $E=E(r,z)$.

holds. A clearer interpretation of static results comes from distribution of the dependent variable along r and z.

Figure 2 plots variables along r with z as a parameter for both $E(r)$ and $E(r,z)$. In order to satisfy the prescribed u condition at $r=r_p$, the resulting applied traction varies along z. For comparative purposes, the magnitude of applied traction is then adjusted by a constant factor so that in both cases average pressure is unity. For $E(r)$, u [Fig. 2(a1)] is symmetric about $z=l/2$ and diminishes steeply along r reaching small values at $r=r_o$. For $E(r,z)$, u [Fig. 2(a2)] is asymmetric about $z=l/2$ from modulus asymmetry and maintains its magnitude throughout l. The same applies to w [Fig. 2(b1,b2)] except that for $E(r,z)$, w_{mx} is an order of magnitude larger than its value for $E(r)$. σ_{rr} is comparable for both $E(r)$ and $E(r,z)$ cases [Fig. 2(c1,c2)], while $\sigma_{\theta\theta}$ follows the same trend as u [Fig. 3(d1,d2)]. Comparing Fig. 2(e1) to (e2) reveals that although σ_{zz}'s magnitude remains almost the same, it changes from tensile for $E(r)$ to compressive for $E(r,z)$.

Figure 3 plots static variable along z with r as a parameter. u [Fig. 3(a1,a2)] follows the uniform prescribed u_o. Magnitude differs for each case due to the normalization explained earlier. At $r>r_p$, u_{mx} in Fig. 3(a2) is almost an order of magnitude larger than that in Fig. 3(a1) and the same applies to w_{mx} in Fig. 3(b1,b2). σ_{rr} in Fig. 3(c1) is symmetric about $z=l/2$ while σ_{rr} in Fig. 3(c2) is not. Note the steep rise in magnitude of σ_{rr} near z_a and z_b typical of the traction distribution necessary to achieve a uniform pre-

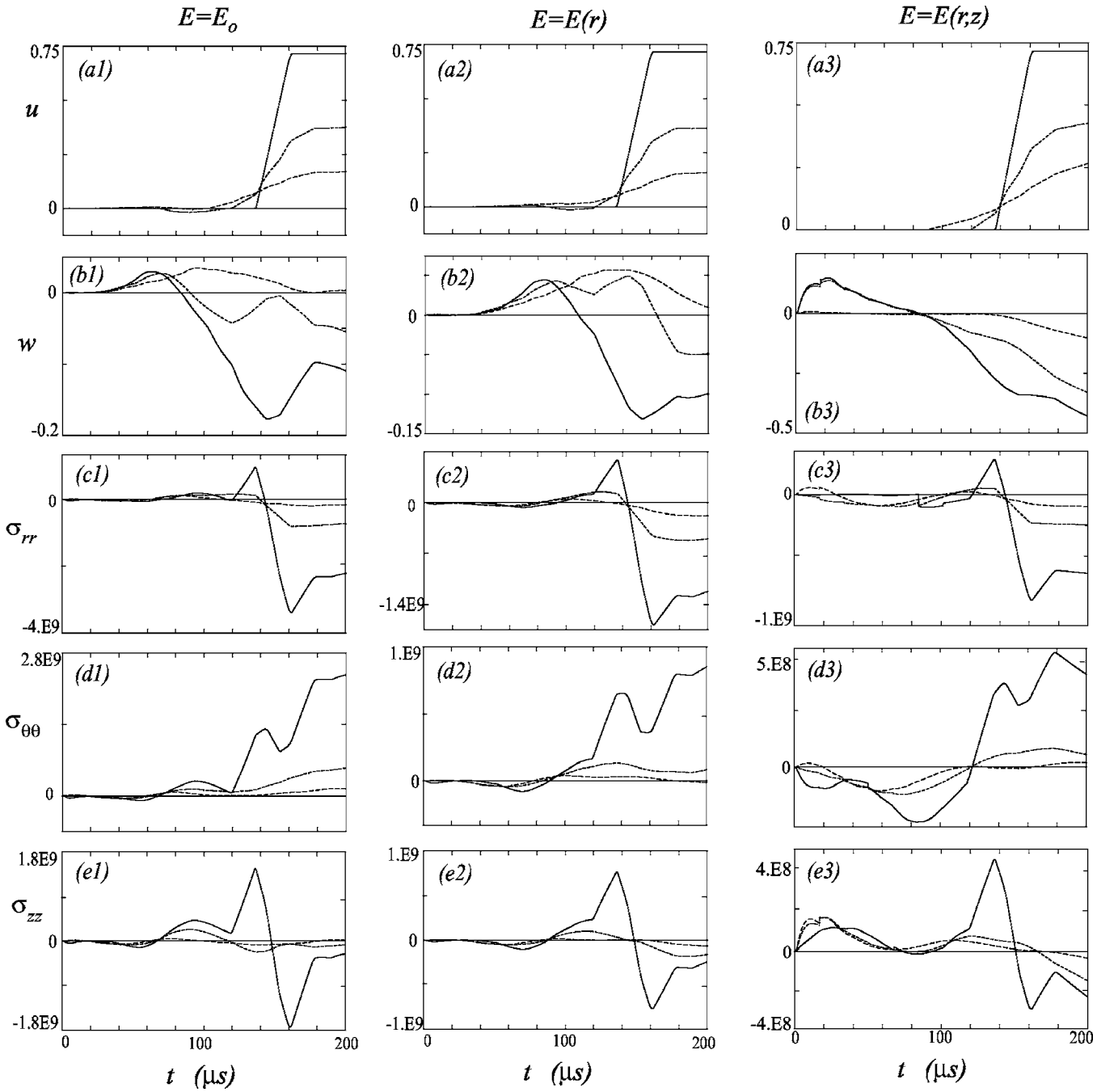


FIG. 6. Histories of cylinder at $z=5.6$ cm.———$r/r_p=1$, --- 2, - - - 4 (a1)–(e1) $E=E_o$. (a2)–(e2) $E=E(r)$. (a3)–(e3) $E=E(r,z)$.

scribed u_o (see Ref. 1, Fig. 1). $\sigma_{\theta\theta}$ [Fig. 3(d1, d2)] is discontinuous at z_a and z_b while its magnitude follows the same steep rise as σ_{rr} near these limits. As noted in Fig. 2(e1, e2), σ_{zz} changes from tensile for $E(r)$ to compressive for $E(r,z)$ and is discontinuous at z_a and z_b. For $z_a<z<z_b$ magnitude of σ_{zz} for $E(r)$ is almost ten times larger than that for $E(r,z)$.

In calculating transient response, prescribed motion follows the profiles in Figs. 4(a)–4(c) where the maximum uniform velocity attained after a 2-μs transient is $V_p=300$ m/s (=984 ft/s). The acceleration profile is made of six linear segments:

(1) Linear acceleration: $\ddot{f}_{p1}(t)=\alpha_1 t, \quad 0\leqslant t\leqslant t_1,$

(2) Constant acceleration: $\ddot{f}_{p2}(t)=\alpha_1 t_1, \quad t_1\leqslant t\leqslant t_2,$

(3) Linear deceleration: $\ddot{f}_{p3}(t)=\ddot{f}_{p2}(t_2)-\alpha_2(t-t_2),$

$t_2\leqslant t\leqslant t_3,$

(4) Constant velocity: $\ddot{f}_{p4}(t)=0, \quad \dot{f}_{p4}(t_3)=V_p,$

$t_3\leqslant t\leqslant t_4,$

(5) Linear deceleration: $\ddot{f}_{p5}(t)=-\alpha_3(t-t_4),$

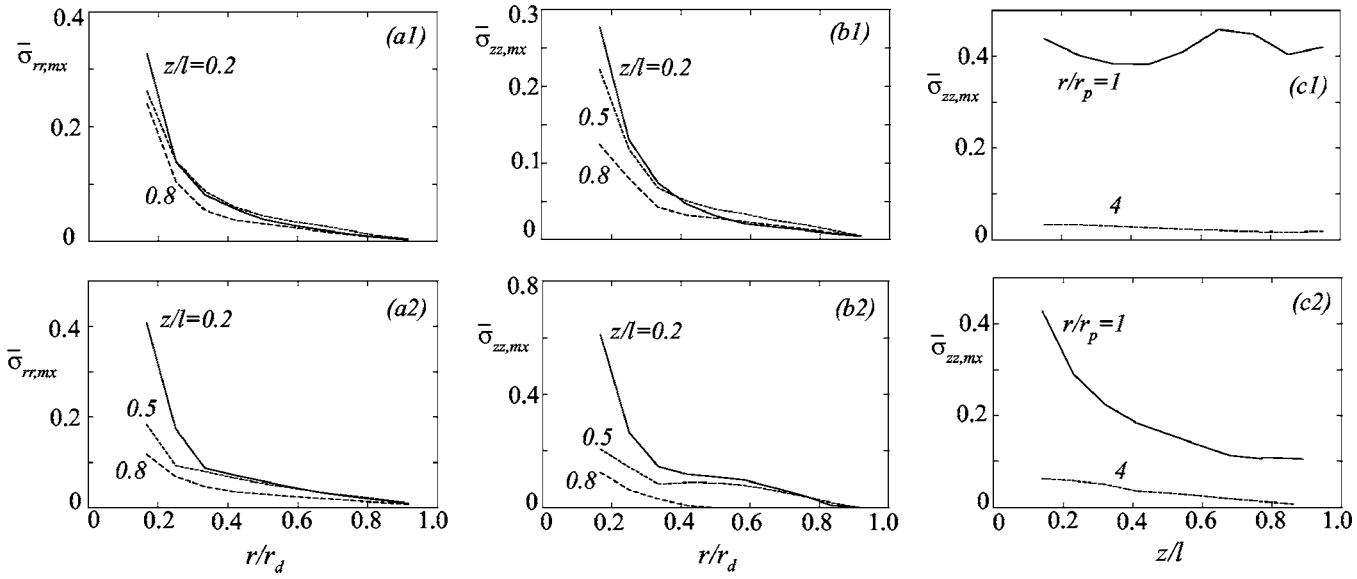


FIG. 7. Peak stress attenuation in 3-D axisymmetric model. (a1), (b1), (c1) Homogeneous $E=E_o$. (a2), (b2), (c2) Inhomogeneous $E=E_o\xi(r)\eta(z)$.

$$t_4 \leqslant t \leqslant t_5,$$

(6) Constant deceleration: $\ddot{f}_{p6}(t) = -\alpha_3(t_5 - t_4)$,

$$t_5 \leqslant t \leqslant t_6.$$

t_6 is the time when $\dot{f}_p(t)=0$. Assuming that the first three time intervals are equal ($\Delta t_1=\Delta t_2=\Delta t_3$, $\Delta t_i=t_i-t_{i-1}$) and $\alpha_2=\alpha_1$, then α_1 is determined by assigning the constant velocity V_p to $\dot{f}_{p4}(t_3)$. In the analysis to follow $\Delta t_{1,3}\equiv\Delta t_1+\Delta t_2+\Delta t_3=2\ \mu s$, $\Delta t_{3,4}=20\ \mu s$, and $\alpha_3=1/3\alpha_1$.

Figure 5 plots the history of variables at $z/l=0.266$ for three cases:

(1) Case A, Fig. 5(a1)–(e1), homogeneous cylinder with $E=E_o$.
(2) Case B, Fig. 5(a2)–(e2), radial inhomogeneity with $E=E(r)$.
(3) Case C, Fig. 5(a3)–(e3), radial and axial inhomogeneity $E=E(r,z)$.

At $r=r_p$, u [Fig. 5(a1, a2, a3)] follows the time-delayed prescribed $f_p(t)$ in Fig. 4(a) for all cases. For $r>r_p$, u for case C is larger than the other two cases. The same applies to w [Fig. 5(b1)–(b3)] where for case C it is an order of magnitude larger than the other two cases as observed for the static case in Fig. 3(a2) and (b2). This follows the reduction in modulus along z as prescribed by (36c). σ_{rr} (Fig. 5(c1)–(c3)) peaks at 10 μs, then at 40 μs, and then follows an almost flat plateau. Its magnitude is largest for case A, followed by case B and case C, since stiffness of these cases follows the same order. The second peak happens 20 μs after arrival time of the time-delayed prescribed motion. $\sigma_{\theta\theta}$ follows the same trend as σ_{rr} in magnitude and shape [Fig. 5(d1)–(d3)]. Unlike the static case in Figs. 2(e2) and 3(e2), σ_{zz} [Fig. 5(e1)–(e3)] is compressive for all cases. After the second peak at 40 μs, σ_{zz} reaches an almost flat plateau with magnitude for case A the same as case C, trailed by that for case B. The first peak at 10 μs occurs during an almost hydrostatic stress state where $\sigma_{rr}\approx\sigma_{\theta\theta}\approx\sigma_{zz}$.

Figure 6 plots the history of variables for the three cases when $z/l=0.733$. At $r=r_p$, u [Fig. 6(a1)–(a3)] follows the time-delayed prescribed motion that reaches $z/l=0.733$ at the arrival time of 140 μs. For cases A and B at $r>r_p$, u starts rising near 120 μs just before arrival time. The same observation applies to w [Fig. 6(b1)–(b3)]. For case C, the magnitude of w is four times larger than the other two cases. σ_{rr} peaks at 160 μs following arrival time [Fig. 6(c1)–(c3)] with largest magnitude for case A, followed by case B, then case C. For case C, σ_{rr} is four times smaller than in case A. The same observation applies to $\sigma_{\theta\theta}$ [Fig. 6(d1)–(d3)] and σ_{zz} [Fig. 6(e1)–(e3)]. An observation common to all variables is quiescence prior to arrival time.

The effect on peak transient stresses of inhomogeneity is shown in Fig. 7. For the homogeneous cylinder with $E=E_o$, Figs. 7(a1) and (b1) plot $\bar{\sigma}_{rr,mx}=\sigma_{rr,mx}(r,z_s)/\sigma_{rr,mx}(r_p,z_s)$ and $\bar{\sigma}_{zz,mx}=\sigma_{zz,mx}(r,z_s)/\sigma_{zz,mx}(r_p,z_s)$ versus r/r_d for $r>r_p$ at three axial positions $z_s/l=0.2,0.5,0.8$, while Fig. 7(c1) plots $\bar{\sigma}_{zz,mx}=\sigma_{zz,mx}(r_s,z)/\sigma_{zz,mx}(r_s,0)$ versus z/l at two radial positions $r_s/r_p=1$ and 4. An explanation that $\bar{\sigma}_{rr,mx}$ and $\bar{\sigma}_{zz,mx}$ are substantially higher at $r=r_p$ and $z=0$ is that immediately following impact, a hydrostatic state of stress prevails which disperses quickly. The hydrostatic condition replicates the one-dimensional environment, thus explaining the higher stresses. Both $\bar{\sigma}_{rr,mx}$ and $\bar{\sigma}_{zz,mx}$ fall steeply with r as with the static case [Figs. 2 (c1) and (e1)]. As expected, after a sharp fall from the value at $z=0$, $\bar{\sigma}_{zz,mx}$ is almost constant along z for a fixed value of r_s as with the static case [Fig. 3(e1) for $0.25\leqslant z/l\leqslant 0.75$]. For the inhomogeneous cylinder [Figs. 7(a2)–(b2)] $\bar{\sigma}_{rr,mx}$ and $\bar{\sigma}_{zz,mx}$ are higher than those of the homogeneous cylinder with greater sensitivity to z_s. Moreover, $\bar{\sigma}_{zz,mx}$ [Fig. 7(c2)] falls along z as with the static case [Fig. 3(e2) for $0.25\leqslant z/l\leqslant 0.75$].

A simplification to the 3-D axisymmetric models in Secs. II and III are the radial and axial plane-strain models

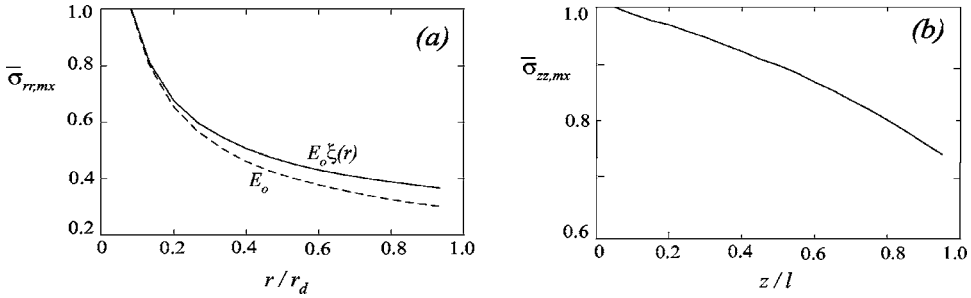


FIG. 8. Peak stress attenuation in plane-strain models. (a) Radial model.———$E=E_o\xi(r)$, --- $E=E_o$. (b) Axial model.———$E=E_o\eta(z)$.

which are limits when either axial or radial dependences are eliminated. The radial plane-strain model has the same r_p and r_o as the cylinder with $E(r)=E_o\xi(r)$. The axial plane-strain model has the same l as the cylinder with $E(z)=E_o\eta(z)$. Figure 8(a) plots distribution of $\bar{\sigma}_{rr,mx}$ along r for the inhomogeneous case (solid line) and the homogeneous case (dashed line). Comparing Fig. 8(a) with Fig. 7(a1, a2) reveals that in plane-strain stress falls more gradually than in 3-D. The greater attenuation in 3-D is caused by dispersion. It is evident that coupling between axial and radial motions makes the cylinder's response features unique in that the plane-strain models cannot duplicate.

VIII. CONCLUSION

The effect on transient response of radial and axial inhomogeneity is analyzed. Radial inhomogeneity is treated adopting transfer matrices of coaxial segments each with constant modulus. Axial inhomogeneity is then included employing the eigenfunctions of the radially inhomogeneous problem as a trial function in a Galerkin procedure. Noteworthy results are the following:

(1) Static response includes the principal features of transient response concerning stress and displacement.
(2) Static stress reduces sharply with r for the homogeneous and radially inhomogeneous cases. However, in the case of axial inhomogeneity stress maintains its value at the excited boundary further along r.
(3) A reduction in modulus along r and z reduces transient stress proportionately and increases w substantially.
(4) All transient variables are quiescent prior to arrival time of the time-delayed prescribed motion.
(5) In both static and dynamic response, prescribed motion raises radial boundary traction near the edges of the footprint along the cylinder's inner boundary.
(6) The simplified plane-strain models cannot replicate transient stress attenuation of the inhomogeneous cylinder as coupling between axial and radial motions makes the cylinder unique in response features.

[1]D. Gazis, "Exact analysis of the plane-strain vibration of thick-walled hollow cylinders," J. Acoust. Soc. Am. **30**, 786–794 (1958).
[2]J. Bird, R. Hart, and F. McClure, "Vibration of thick-walled hollow cylinders: Exact numerical solutions," J. Acoust. Soc. Am. **32**, 1403–1412 (1960).
[3]V. Grinchenko and V. Meleshko, "Axisymmetric vibrations of an elastic cylinder of finite length," Sov. Phys. Acoust. **24**, 488 (1978).
[4]H. Batard and G. Quentin, "Acoustical resonances of solid elastic cylinders: Parametric study," J. Acoust. Soc. Am. **91**, 581 (1992).
[5]V. Grinchenko, "Eigenforms and eigenfrequency spectrum of finite elastic cylinder," J. Acoust. Soc. Am. **105**, 1392 (1999).
[6]K. Soldatos and J. Ye, "Wave propagation in anisotropic laminated hollow cylinders of infinite extent," J. Acoust. Soc. Am. **96**, 3744–3752 (1994).
[7]M. Hussein and P. Heyliger, "Three-dimensional vibrations of layred piezoelectric cylinders," J. Eng. Mech. **124**, 1294 (1998).
[8]Z. Cheung, Y. Lo, and S. Au, "3-D vibration analysis of solid and hollow circular cylinders via Chebyshev-Ritz method," Comput. Methods Appl. Mech. Eng. **192**(13–14), 1575–1589 (2003).
[9]J. Sastry and M. Munjal, "Response of a multi-layered infinite cylinder to a plane wave excitation by means of transfer matrices," J. Sound Vib. **209**(1), 99–121 (1998).
[10]H. Wang, H. Ding, and Y. Chen, "Dynamic solution of a multilayered orthotropic piezoelectric hollow cylinder for axisymmetric plane strain problems," Int. J. Solids Struct. **42**(1), 85–102 (2004).
[11]P. Heyliger and A. Jilani, "The free vibrations of inhomogeneous elastic cylinders and spheres," Int. J. Solids Struct. **29**(22), 2689–2708 (1992).
[12]X. Wang and Y. Gong, "An elastodynamic solution for multilayered cylinders," Int. J. Eng. Sci. **30**(1), 25–33 (1992).
[13]J. Whittier and J. Jones, "Axially symmetric wave propagation in a two-layered cylinder," Int. J. Solids Struct. **3**, 657–675 (1967).
[14]R. Reuter, "Dispersion of flexural waves in circular bimaterial cylinders—Theoretical treatment," J. Acoust. Soc. Am. **46**, 643–648 (1969).
[15]H. Keck and A. Armenakas, "Wave propagation in transversely isotropic layered cylinders," J. Eng. Mech. **97**, 541–558 (1971).
[16]Z. Lee, "Hybrid numerical method applied to 3-D multilayered hollow cylinder with time-dependent boundary conditions," Appl. Math. Comput. **150**(1), 25–43 (2004).
[17]H. Paul and V. Murali, "Dynamic response of axisymmetric poroelastic cylindrical bone," J. Acoust. Soc. Am. **98**, 2978–2986 (1995).
[18]K. Soldatos, "Review of three-dimensional dynamic analyses of circular cylinders and cylindrical shells," Appl. Mech. Rev. **47**(10), 501–516 (1994).
[19]X. Yin and Z. Yue, "Transient plane-strain response of multilayered elastic cylinders to axisymmetric impulse," J. Appl. Mech. **69**, 825–835 (2002).
[20]L. Steinberg, "Inverse spectral problems for inhomogeneous elastic cylinders," J. Elast. **38**, 133–151 (1995).
[21]M. El-Raheb, "Wave propagation in a hollow cylinder due to prescribed velocity at the boundary, Int. J. Solids Struct., **41**(18–19), 5051–5069 (2004).
[22]M. El-Raheb, "Transient response in a finite hollow cylinder from time-delayed prescribed motion at the boundary, J. Sound Vib. **285**, 491–500 (2005).
[23]J. Berry and P. Naghdi, "On the vibration of elastic bodies having time dependent boundary conditions, Q. Appl. Math. **14**, 43–50 (1956).
[24]M. El-Raheb, "Dynamics of two-dimensional composites of elastic and viscoelastic layers, J. Acoust. Soc. Am. **112**, 1445–1455 (2002).

Three-dimensional simulation of electron-acoustic detections

Chun-ming Gao, Shu-yi Zhang, Zhong-ning Zhang, and Xiu-ji Shui
Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China

(Received 29 April 2005; revised 9 January 2006; accepted 15 February 2006)

In electron-acoustic (EA) detection, a periodically modulated electron beam is impinging on a sample to excite thermo-acoustic signals, and a PZT transducer bonded on the back surface of the sample is used to detect the thermo-acoustic signals. Based on the conditions of the EA detection experiment, theoretical analyses of 3-D temperature distributions in the sample, 3-D elastic fields in the sample and piezoelectric transducer, and piezoelectric output signals of the transducer are given. Generally, the vibration modes and the electrical output signals of the transducer are hard to calculate because of the complicated structures, i.e., complicated boundary conditions, thus a finite element method (FEM) is used to simulate the electron-acoustic signals. According to the corresponding FEM model, the frequency spectra of the detected EA signals can be calculated. Meanwhile, the related experimental measurements for several homogeneous samples are carried out. Comparing the theoretical simulated results with those of the experiments, it shows that both are in good agreement with each other. Finally, using this FEM model, the dependences of the electron-acoustic signals on the thermal and elastic properties of several samples are investigated.
 [DOI: 10.1121/1.2185491]

PACS number(s): 43.40.Ey, 43.40.Sk, 43.58.Ls [AJZ] Pages: 2750–2757

I. INTRODUCTION

Scanning electron-acoustic microscopy (SEAM) was established by Brandis *et al.*[1] and Cargill III[2] independently in 1980. Since then, the applications of SEAM for characterizing the microstructures of materials and devices have been widely developed, such as doping regions in semiconductors,[3] magnetic domains in ferromagnetic and microwave materials,[4] ferroelectric domains in ferroelectric ceramics,[5] residual stresses in ceramics[6] and metals,[7] surface and subsurface structures in biological materials,[8–10] and so on. The advantages of SEAM are the special imaging abilities for local different thermal or elastic features of surface and subsurface structures with high resolution, by which laminated images of micro-inhomogeneous materials can be obtained nondestructively.[7,10] It is well known that the similar results can also be obtained by conventional acoustic microscopy (AM).[11] However, due to the thermal wavelength being about one thousand times smaller than the acoustic wavelength and operating in the same frequency and the same material, the operating frequency in SEAM can be decreased three orders of magnitude for obtaining the same special resolution.[12]

On the other hand, several theoretical studies on electron-acoustic (EA) signal generation have been performed. Bresse *et al.*[13,14] evaluated the electron-acoustic signal generation in semiconductors and metallic layers by one-dimensional model; Rosencwaig *et al.*[15] and Qian *et al.*[16,17] analyzed the mechanisms of the acoustic signal generation by focused electron beam in a three-dimensional model, but did not give the numerically calculated results. Comparing with the successful SEAM imaging experiments, the theoretical research is quite deficient.

In this work, based on the experimental conditions of SEAM,[7,10] 3-D analyses of the temperature distributions in the sample, the elastic fields in the sample, and the piezoelectric transducer induced by focused electron beam impinging are given, and then a finite element method (FEM) is used for simulating the EA signals detected by a PZT transducer. Meanwhile, several experiments for homogeneous samples have also been carried out. Comparisons of the experimental measurements and theoretical simulations, which verify the theoretical modeling and the simulation results of EA signals, are available. Finally, the dependences between the EA signals on the properties of the samples are obtained, which illustrate that the EA signals can display different structures and properties of materials, and then the imaging mechanisms of inhomogeneous materials by SEAM can be demonstrated. In addition, the properties and parameters of the materials can also be evaluated by SEAM signals.

II. THEORY

Based on the experimental method, the theoretical model of the EA detection with the cylindrical coordinate is shown in Fig. 1, in which a periodically modulated electron beam with frequency f is impinging on the surface of the sample to excite the thermo-acoustic waves, and the thermo-acoustic waves are detected by a circular piece of PZT transducer.

In Fig. 1, L_s and r_s are the thickness and radius of the sample, L_t and r_t are the thickness and radius of the transducer, P_1 and P_2 are the pressure on the upper and back sides of the transducer, respectively, and r_p is the aperture radius of the brass cap.

 0001-4966/2006/119(5)/2750/8/$22.50

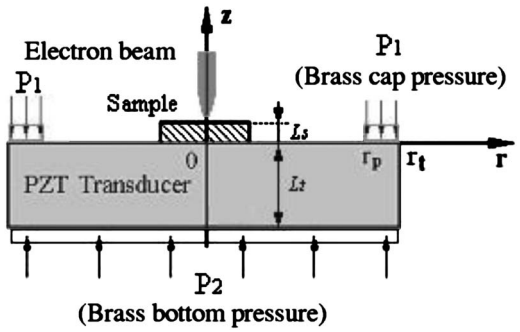


FIG. 1. Theoretical model of EA detection.

The thermal diffusion equation in the sample can be expressed as

$$\nabla^2 T_s(r,z) - \frac{i\omega}{D_s} T_s(r,z) = -\frac{1}{k_s} H(r,z),$$
$$D_s = k_s/\rho_s c_s, \tag{1}$$

where $T_s(r,z)$ is the temperature variation induced by the electron beam, D_s, k_s, ρ_s, and c_s are the thermal diffusivity, thermal conductivity, density, and specific heat of the sample, respectively, and $H(r,z)$ is the power density of the heat source,[16]

$$H(r,z) = \frac{\eta P_0 \beta}{2\pi r_0^2} \exp(-\beta|z - z_0|)\exp(-r^2/r_0^2), \tag{2}$$

where P_0 is the power and r_0 is the focused radius of the electron beam, η is the efficiency of the incident electron beam converted to heat, β is the electron energy attenuation coefficient, and z_0 is the electron penetration depth in the sample.

In the SEAM detection, the sample is placed under vacuum at room temperature, therefore the thermal boundary conditions at the upper and lateral surfaces of the sample are assumed to be adiabatic. Meanwhile, the thermal wave in the samples is damping very quickly and the propagation distance is about one wavelength. In our experimental situation, the modulation frequency is several hundreds kHz, the wavelength of the thermal wave is approximately several μm, and the thickness and radius of the sample are several hundreds to several thousands of μm, which is much larger than the wavelength of the thermal wave. Thus the boundary conditions can be written as

$$\left.\frac{\partial T_s(r,z)}{\partial z}\right|_{z=0} = 0, \qquad \left.\frac{\partial T_s(r,z)}{\partial z}\right|_{z=Is} = 0,$$
$$\left.\frac{\partial T_s(r,z)}{\partial r}\right|_{r=r_s} = 0. \tag{3}$$

Using the boundary conditions (3), the temperature distributions in the sample can be obtained by solving Eq. (1).

For an isotropic solid sample, the thermoelastic equation can be expressed as[16]

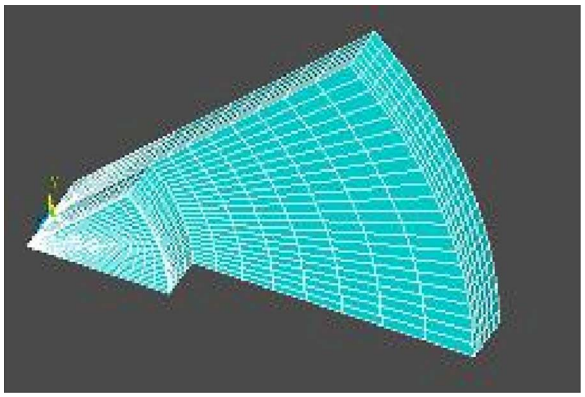


FIG. 2. (Color online) Meshed model of detection component.

$$\mu_s \nabla^2 \bar{u}_s(r,z) + (\lambda_s + \mu_s)\nabla[\nabla \cdot \bar{u}_s(r,z)] + \rho_s \omega^2 \bar{u}_s(r,z) = \alpha_{TS}(3\lambda_s + 2\mu_s)\nabla T_s(r,z), \tag{4}$$

where $\bar{u}_s(r,z)$ is the displacement field, λ_s and μ_s are the Lamé constants, and α_{TS} is the thermal expansion coefficient of the sample.

For the piezoelectric transducer made by PZT-5 ceramic, due to the elastic constants of the material[18] approach to those of isotropic materials, the transducer can be approximately assumed as elastic isotropic, thus the elastic equation can be expressed as

$$\mu_t \nabla^2 \bar{u}_t(r,z) + (\lambda_t + \mu_t)\nabla[\nabla \cdot \bar{u}_t(r,z)] + \rho_t \omega^2 \bar{u}_t(r,z) = 0, \tag{5}$$

where $\bar{u}_t(r,z)$ is the displacement field, λ_t and μ_t are the Lamé constants, and ρ_t is the density of the transducer. Based on Eqs. (1)–(5) and the boundary conditions of continuous displacement and stress in the interface of the sample and transducer, the displacement and stress fields in the transducer can be determined.

In addition, for the PZT transducer with axial polorization, the piezoelectric equations are[18]

$$\bar{\tau} = K\bar{u}_t - K^z\bar{E},$$
$$\bar{D}_e = K^z\bar{u}_t + K^d\bar{E}, \tag{6}$$

where $\bar{\tau}$ is the stress, $\bar{u}_t$ is the strain, $\bar{E}$ is the electric field, $\bar{D}_e$ is the electric displacement, K, K^z, and K^d are the stiffness coefficients, the piezoelectric constants, and dielectric coefficients of the transducer, respectively. Then the output voltage of the PZT transducer can be obtained.

According to the experimental conditions, the boundary conditions can be written as follows:

The stresses are free on both the upper surface of the sample and a part of the upper surface of the transducer (see Fig. 1), i.e.,

$$\tau_{zs} = \tau_{rzs} = 0 \quad (\text{at } z = L_s \text{ and } 0 < r < r_s),$$
$$\tau_{zt} = \tau_{rzt} = 0 \quad (\text{at } z = 0 \text{ and } r_s < r < r_P), \tag{7}$$

where τ_{zs} and τ_{rzs} are the normal and shear stress components of the sample and τ_{zt} and τ_{rzt} are the normal and shear

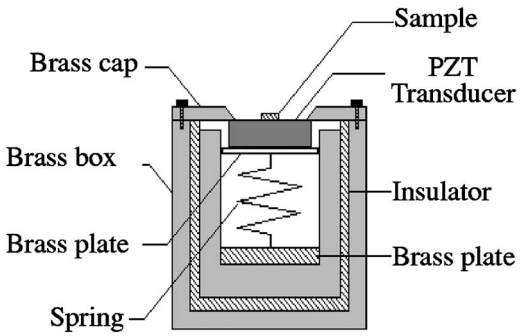


FIG. 3. Structure of piezoelectric detection component.

stress components of the transducer, respectively.

The stresses at the upper surface and the backside of the PZT transducer can be expressed as

$$\begin{aligned}
&\tau_{zt} = p_1 \quad (\text{at } z = 0 \text{ and } r_P < r < r_t),\\
&\tau_{rzt} = 0 \quad (\text{at } z = 0 \text{ and } r_P < r < r_t),\\
&\tau_{zt} = -p_2 \quad (\text{at } z = -L_t \text{ and } 0 < r < r_t),\\
&\tau_{rzt} = 0 \quad (\text{at } z = -L_t \text{ and } 0 < r < r_t).
\end{aligned} \tag{8}$$

At the interface between sample and transducer, the displacements and stresses are considered continuous,

$$\begin{aligned}
&u_{rs} = u_{rt}, \quad u_{zs} = u_{zt} \quad (\text{at } z = 0 \text{ and } 0 < r < r_s),\\
&\tau_{rs} = \tau_{rt}, \quad \tau_{zs} = \tau_{zt} \quad (\text{at } z = 0 \text{ and } 0 < r < r_s),
\end{aligned} \tag{9}$$

where the subscripts r, z represent r and z directions, and s, t represent the sample and transducer. According to the temperature distribution in the sample obtained from Eqs. (1)–(3) and the complicated boundary conditions shown in Eqs. (7)–(9), the thermo-elastic equations (4) and (5) are difficult to solve by analytical method, and then the EA signals cannot be obtained analytically.

Therefore, a finite element method (FEM) is employed to solve the temperature distributions, displacement fields, and EA signals numerically.

III. FEM ANALYSIS

The general FEM form of the thermal diffusivity equation can be written by[19]

$$\begin{bmatrix} C_1 & 0 & \cdots & 0 \\ 0 & C_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & C_m \end{bmatrix} \begin{bmatrix} \dot{\phi}_1 \\ \dot{\phi}_2 \\ \vdots \\ \dot{\phi}_m \end{bmatrix} + \begin{bmatrix} K_1 & 0 & \cdots & 0 \\ 0 & K_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & K_m \end{bmatrix} \times \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_m \end{bmatrix} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix}, \tag{10}$$

where C_i is the heat-capacity matrix, $\dot{\phi}_i = d\phi_i/dt$, ϕ_i is the temperature array, K_i is the heat-conduction matrix, and P_i is the temperature-load array. The K_i, C_i, and P_i can be expressed as[20]

$$\begin{aligned}
K_i &= \sum \int_{V_i} k_i \left(\left(\frac{\partial N}{\partial r} \right)^* \left(\frac{\partial N}{\partial r} \right) + \left(\frac{\partial N}{\partial z} \right)^* \left(\frac{\partial N}{\partial z} \right) \right.\\
&\quad \left. - \left(\frac{\partial N}{\partial z} \right)^* \left(\frac{N}{r} \right) \right) dV_i,\\
C_i &= \sum \int_{V_i} \rho_i c_i N^* N \, dV_i,\\
P_i &= \sum \int_{V_i} \rho_i Q_i N^* \, dV_i,
\end{aligned} \tag{11}$$

where the k_i, ρ_i, and c_i are the heat conductivity, density, and specific heat, respectively, Q_i is the heat energy absorbed from the electron beam, V_i is the volume of the ith element, N is the interpolation function array, and * is the transposed operator.

The general FEM form of the thermo-elastic couple equation can be written by

TABLE I. The composition of sample and transducer.

Samples	Composite structure		Transducer 1 Radius (mm) 12.0	Transducer 1 Thickness (mm) 2.54	Transducer 2 Radius (mm) 14.0	Transducer 2 Thickness (mm) 0.86
A	Radius (mm)	5.50	A1		A2	
	Thickness (mm)	1.00				
B	Radius (mm)	4.85	B1		B2	
	Thickness (mm)	0.80				
C	Radius (mm)	4.35	C1		C2	
	Thickness (mm)	0.98				
D	Radius (mm)	3.54	D1		D2	
	Thickness (mm)	0.80				

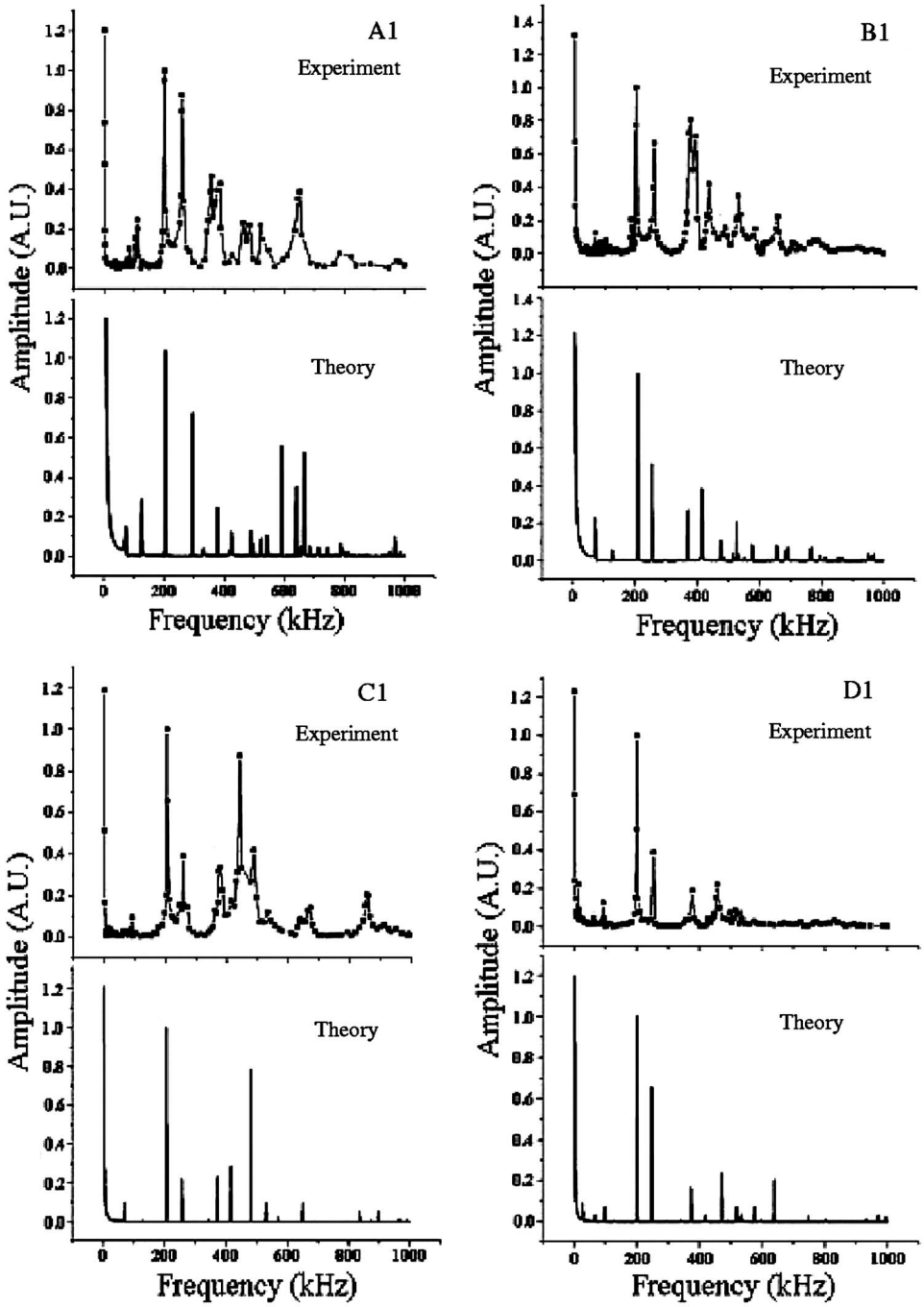


FIG. 4. EA amplitude versus modulation frequency for transducer 1.

$$\begin{bmatrix} [M] & [0] \\ [0] & [0] \end{bmatrix} \begin{Bmatrix} \{\ddot{u}\} \\ \{\ddot{T}\} \end{Bmatrix} + \begin{bmatrix} [D] & [0] \\ [0] & [D^t] \end{bmatrix} \begin{Bmatrix} \{\dot{u}\} \\ \{\dot{T}\} \end{Bmatrix} + \begin{bmatrix} [K] & [0] \\ [0] & [K^t] \end{bmatrix} \begin{Bmatrix} \{u\} \\ \{T\} \end{Bmatrix} = \begin{Bmatrix} \{F\} \\ \{Q\} \end{Bmatrix},$$

$$\dot{u} = \partial u/\partial t, \quad \ddot{u} = \partial^2 u/\partial t^2,$$

$$\dot{T} = \partial T/\partial t, \quad \ddot{T} = \partial^2 T/\partial t^2,$$

$$[K^t] = [K^{tb}] + [K^{tc}], \tag{12}$$

$$\{F\} = \{F^{nd}\} + \{F^{th}\} + \{F^{pr}\} + \{F^{ac}\},$$

$$\{Q\} = \{Q^{nd}\} + \{Q^g\} + \{Q^c\},$$

where M is structural mass matrix; u is displacement vector; T is temperature vector; D is structural damping matrix; K, K^t, K^{tb}, and K^{tc} are the structural stiffness matrix, the total thermal conductivity matrix, the thermal conductivity matrix of material, and the thermal conductivity matrix of surface convection, respectively; F is the force vector; F^{nd} is applied nodal force vector; F^{th} is the thermal strain force vector; F^{pr} is the pressure load vector; F^{ac} is the force vector due to acceleration effects; Q^{nd} is the applied nodal heat flow rate vector; Q^g is the heat generation rate vector for causing other than Joule heating; and Q^c is the convection surface vector.

The general FEM form of the piezoelectric couple equation can be written by

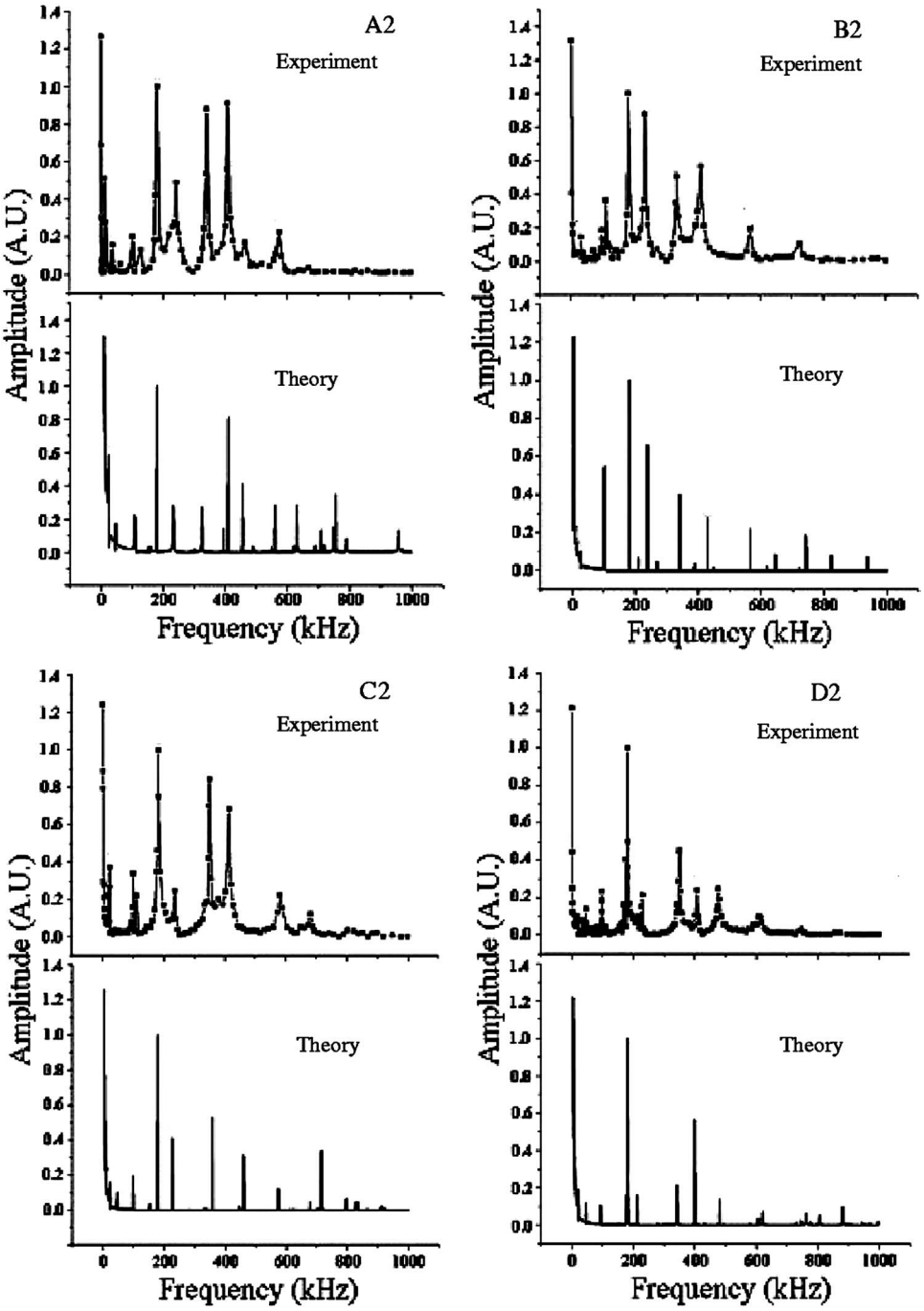


FIG. 5. EA Amplitude versus modulation frequency for transducer 2.

$$\begin{bmatrix} [M] & [0] \\ [0] & [0] \end{bmatrix} \begin{Bmatrix} \{\ddot{u}\} \\ \{\ddot{V}'\} \end{Bmatrix} + \begin{bmatrix} [D] & [0] \\ [0] & [0] \end{bmatrix} \begin{Bmatrix} \{\dot{u}\} \\ \{\dot{V}'\} \end{Bmatrix} + \begin{bmatrix} [K] & [K^z] \\ [K^z]^* & [K^d] \end{bmatrix} \begin{Bmatrix} \{u\} \\ \{V'\} \end{Bmatrix} = \begin{Bmatrix} \{F\} \\ \{L\} \end{Bmatrix}, \tag{13}$$

$$\dot{V}' = \partial V'/\partial t, \quad \ddot{V}' = \partial^2 V'/\partial t^2,$$

$$\{L\} = \{L^{nd}\} + \{L^c\} + \{L^{sc}\},$$

where V' is the output voltage; K, K^z, and K^d are the stiffness coefficient matrix, the piezoelectric coefficient matrix, and the dielectric coefficient matrix of the transducer, respectively; and L, L^{nd}, L^c, and L^{sc} are the load vector, applied nodal charge vector, charge density load vector, and surface charge density load vector, respectively.

Since the detection component is cylindrical symmetric, in order to simplify the calculation, $\frac{1}{8}$ of the detection component can be simulated by the FEM, in which a FEM model is shown in Fig. 2. Using the FEM model, the EA signals can be calculated numerically by Eqs. (10)–(13).

IV. EXPERIMENTAL DETECTION

The piezoelectric detection component is shown in Fig. 3, in which the sample is bonded on the surface of the piezoelectric transducer. In the experiments, in order to verify the theoretical FEM simulation model, copper samples with

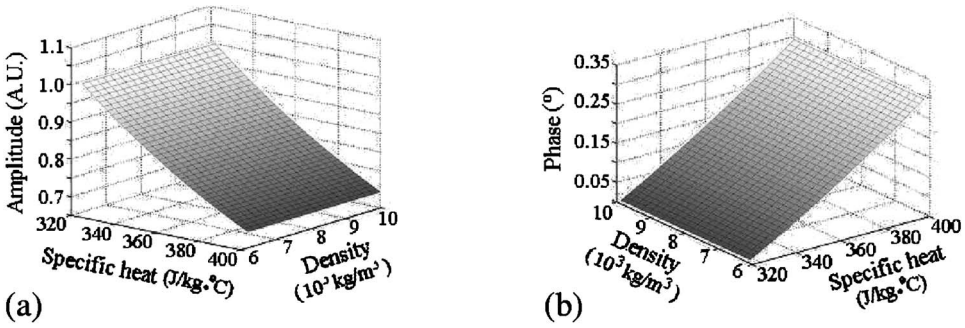


FIG. 6. Dependences of EA signal on the specific heat and density of sample: (a) amplitude and (b) phase.

four different sizes composite with PZT transducers with two different sizes are measured. The different compositions of the samples and transducers are listed in Table I. The distributions of the EA signals versus frequencies obtained by the experimental measurements for the different compositions are shown in Figs. 4 and 5, in which the accelerating voltage and focused radius of the electronic beam are adjusted to 30 KV and 1 μm, respectively.

V. NUMERICAL SIMULATION RESULTS AND DISCUSSIONS

A. EA signals at different frequencies

Based on the experimental conditions, the dependences of the amplitude of the EA signals on the frequency are numerically calculated by using the FEM model, which are also shown in Figs. 4 and 5. In the simulation, the pressures P_1 and P_2 are taken as 10 N/m^2.

Comparing the experimental results with those of the theoretical simulations, it can be seen that the signal peaks observed by the experiments also appear in the simulated curves by the FEM model except a few peaks that appeared in the calculated curves have not been observed in the experiments. From Figs. 4 and 5, it can be illustrated that for most experimental peaks the corresponding frequencies are in good agreement with those of calculated peaks. The frequency deviations between both results are less than 5%, and they are listed in the Appendix. Therefore, the EA signals can be calculated by using the FEM model.

However, except for the strong peaks at around 200 KHz, which are evidently corresponding to the radial resonance frequency of the transducer that appeared in both experimental and theoretical results, the vibration modes of the transducers are very complicated due to the complicated detection structures, i.e., the complicated boundary conditions, which make the mode identification very difficult.

On the other hand, there are a few deviations between the experimental and calculated results including the altitudes and frequencies of the signal peaks. The deviations are probably induced by the following reasons: (1) the material parameters of the sample and transducer may not be adopted precisely in the calculation; (2) the influences of the structures of the detection system, such as the spring, brass electrodes, and other accessories, on the vibrations of the transducer have not been considered; (3) some micro-disbonds between the sample and the transducer may exist in experiment; and (4) the sensitivity of the experimental system needs to be improved to detect very small peaks that appear in the theoretical calculations. In addition, the larger widths of the peaks observed in the experiments, which may be induced by the damping effects of the detection system, are not displayed in the theoretical simulations because the damping effects have not been considered in the theory. Meanwhile, the experimental widening peaks cause several peaks observed in theory to be merged.

B. Influence of the thermal and elastic properties of the samples on EA signals

In order to study the influences of the thermal and elastic characters of the samples on EA signals, the composition system of sample A and transducer 1 is taken as an example to be simulated.

(1) The dependences of the amplitude and phase of the EA signal on the density and specific heat of the sample are shown in Figs. 6(a) and 6(b), respectively. It show that the amplitude of the EA signal decreases and the phase

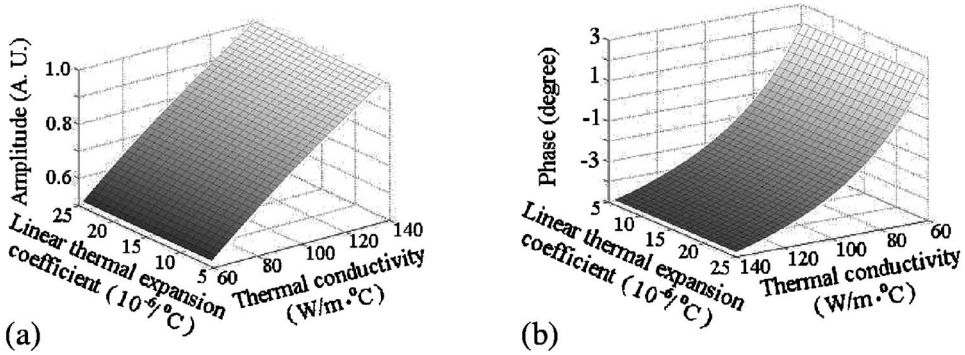


FIG. 7. Dependences of the EA signal on the linear thermal expansion coefficient and thermal conductivity of sample: (a) amplitude and (b) phase.

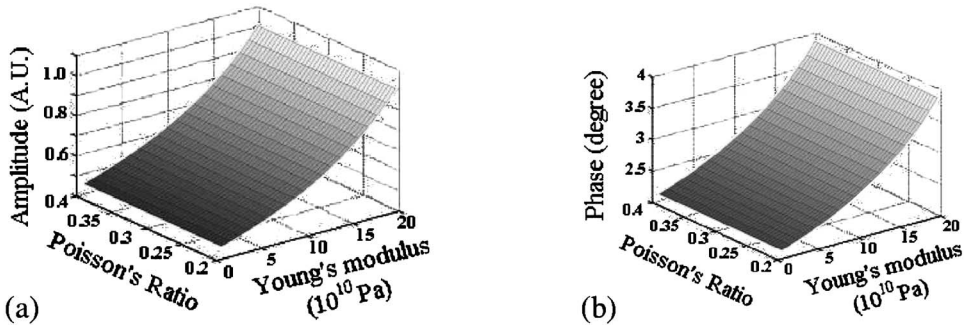


FIG. 8. Dependences of the EA signal on Poisson's ratio and Young's modulus of sample: (a) amplitude and (b) phase.

of the EA signal increases monotonously with the specific heat increase, but there is little influence with the density increase of the sample.

(2) The dependences of the amplitude and phase of the EA signal on the linear thermal expansion coefficient and thermal conductivity of the sample are shown in Figs. 7(a) and 7(b), respectively. It shows that the amplitude and phase of the EA signal increase monotonously with the thermal conductivity increase, and there is little influence with the linear thermal expansion coefficient increase of the sample.

(3) The dependences of the amplitude and phase of the EA signal on Poissons ratios and Young's modules of the sample are shown in Figs. 8(a) and 8(b), respectively, which show that the amplitude and phase of the EA signal increase monotonously with the Young's modules increase, but has influence little on the Poisson's ratios variation of the sample.

(4) Generally, the EA signals change with the variations of the specific heat, thermal conductivity, and Young's modules of the sample sensitively. If the electron beam is focused on the surface of the sample and the thermal wavelength is very small, the EA signal can reveal the local properties of the sample. Therefore, for an inhomogeneous sample, in which the local thermal and/or elastic properties are different from neighboring structures, as the focused electron beam is scanning on the surface of the sample, the EA signal will display different amplitude and phase and then the mechanism of SEAM imaging can be illustrated. On the other hand, it can be pointed out that the properties and parameters of the materials can also be evaluated by EA detections.

VI. CONCLUSIONS

In order to theoretically analyze the EA detections, a 3-D FEM model is used for simulating the EA signals. Comparing the experimental measurements and numerical calculations, several conclusions can be obtained.

(1) In EA detection, the 3-D theoretical relations of the thermal field, elastic field, and piezoelectric signal under complicated boundary conditions are given. Due to the theoretical relations being very complicated, the output signals of EA detection are always difficult to obtain by analytical method. Thus a 3-D FEM model is established to simulate the EA signals.

(2) The experimental measurements of the EA signals in different compositions of the samples and circular PZT transducers are carried out. From the dependences of the amplitude of the EA signals on the modulation frequency, there exist a lot of peaks located at the resonance (harmonic) frequencies and/or their superposition as the frequency lower than 1 MHz.

TABLE II. Frequency of EA signal peak for composite structure with transducer 1.

	Sample											
	A1			B1			C1			D1		
Peak no.	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)
a	82	78	5	69	72	4	81	77	5	16	18	**13**
b	111	116	5	101	116	**13**	205	206	0	62	69	**11**
c	202	206	2	200	208	4	252	267	**6**	98	102	4
d	259	297	**15**	258	289	**11**	379	345	**9**	200	202	1
e	370	378	2	372	350	**6**	442	470	**6**	264	278	5
f	425	424	0	432	416	4	486	482	1	378	373	1
g	484	488	1	476	477	0	528	532	1	457	430	**6**
h	521	524	1	532	527	1	672	683	2	519	529	2
i	651	666	2	581	578	1	862	836	3			
j	784	788	1	654	658	1						
k	974	968	1	694	692	0						
l				770	767	0						

TABLE III. Frequency of EA signal peak for composite structure with transducer 2.

	Sample											
	A2			B2			C2			D2		
Peak no.	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)	f (KHz) Exp.	f (KHz) Cal.	Dev. (%)
a	15	17	**13**	31	29	**7**	19	21	**11**	63	59	**6**
b	36	34	**6**	111	108	3	99	96	3	98	95	3
c	103	108	5	181	182	1	182	179	2	183	166	**9**
d	180	167	**7**	230	240	4	236	228	3	228	213	**7**
e	246	232	**6**	335	343	2	348	337	3	351	344	2
f	341	325	5	412	411	0	413	413	0	407	401	1
g	409	411	0	570	566	1	582	575	1	477	481	1
h	464	456	2	726	725	0	681	693	2	607	607	0
i	575	560	3				803	827	3	748	746	0
j	670	640	4				900	912	1			

(3) In order to compare the theoretical results with those of the experiments, the theoretical simulations according to the experimental conditions are given. They show that the calculated results are in good agreement with the experimental results although there are a few deviations.
(4) The numerical calculations can be used to find the dependences of the EA signals on the thermal and elastic properties of the sample. It shows that the output signals are sensitive to the specific heat, thermal conductivity, and Young's modulus of the sample, even influenced a little by the density, linear expansion coefficient, and Poisson's ratio of the sample. Therefore, the changes of the thermal and/or elastic properties and parameters of the samples can be reflected by the EA signals, and then the mechanism of the SEAM images for inhomogeneous materials can be demonstrated.
(5) The properties and parameters of the samples can be evaluated by the EA detections.

ACKNOWLEDGMENT

This work is supported by National Natural Science Foundation of China, No. 10574073.

APPENDIX

The frequencies of the peaks of experimental and calculated EA signals of the samples consisting of two different transducers are listed in Tables II and III, in which Exp., Cal., and Dev. represent the results of experiment, calculation, and deviation between the experiment and calculation, and bold numbers express the data with the deviation larger than 5%.

[1]E. Brandis and A. Rosencwaig, "Thermal-wave microscopy with electron beams," Appl. Phys. Lett. **37**, 98–100 (1980).
[2]G. S. Cargill III, "Ultrasonic imaging in scanning electron microscopy," Nature (London) **286**, 691–693 (1980).
[3]A. Roscenswaig and R. M. White, "Imaging of dopant regions in silicon with thermal-wave electron microscopy," Appl. Phys. Lett. **38**, 165–167 (1981).
[4]L. J. Balk, D. G. Davies, and N. Kultscher, "Investigation of Si–Fe transformer sheets by Scanning Electron-Acoustic Microscopy (SEAM)," IEEE Trans. Magn. **20**, 1466–1468 (1984).
[5]L. X. Liu and L. J. Balk, "Scanning electron acoustic microscopy for the evaluation of domain structure in $BaTiO_3$ single crystal and ceramics," J. Mater. Sci. **33**, 4543–4549 (1998).
[6]J. H. Cantrell and M. L. Qian, "Scanning electron acoustic microscopy of indentation-induced cracks and residual stresses in ceramics," Appl. Phys. Lett. **57**, 1870–1872 (1990).
[7]Y. Hong, Z. N. Zhang, S. Y. Zhang, Z. Q. Li, and X. J. Shui, "Characterization of residual stress distribution by scanning electron acoustic microscopy," Acoust. Imaging **25**, 273–278 (2000).
[8]L. J. Balk, M. Domnik, and E. Buhm, "Application of scanning electron microscopy for medical research," Proc. SPIE **809**, 219–224 (1987).
[9]L. J. Balk, M. Domnik, K. Niklas, and P. Mestres, "Investigation of human tissues with Scanning Electron Acoustic Microscopy," in *Photoacoustic and Photothermal Phenomena II*, Springer Series in Optical Science (Springer, New York, 1990), Vol. **62**, pp. 428–430.
[10]C. M. Gao, S. Y. Zhang, P. C. Miao, Z. N. Zhang, and P. Li, "Application of scanning electron-acoustic microscopy to biological materials," Int. J. Thermophys. **25**, 857–867 (2004).
[11]K. I. Maslov, T. Kundu, and O. I. Lobkis, "Acoustic microscopy for spherical inclusion characterization," J. Acoust. Soc. Am. **100**, 80–85 (1996).
[12]S. Y. Zhang and L. Chen, "Photoacoustic microscopy and detection of subsurface features of semiconductors," in *Photoacoustic and Photothermal Phenomena in Semiconductors*, edited by A. Mandelis (Elsevier, New York, 1987), pp. 27–52.
[13]J. F. Bresse and A. C. Papadopoulo, "Evolution of the electron acoustic signal as function of doping level in III–V semiconductors," J. Appl. Phys. **64**, 98–102 (1988).
[14]J. F. Bresse, "Electron acoustic signal of metallic layers over a semiconductor substrate," J. Appl. Phys. **71**, 4678–4683 (1992).
[15]A. Roscenswaig and J. Opsal, "Thermal wave imaging with thermoacoustic detection," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **33**, 516–528 (1986).
[16]M. L. Qian and J. H. Cantrell, "Signal generation in scanning electron acoustic microscopy," Mater. Sci. Eng., A **122**, 57–64 (1989).
[17]M. L. Qian and X. M. Wu, "Scanning electron acoustic microscopy of electric domains in ferroelectric materials," J. Mater. Res. **14**, 3096–3101 (1999).
[18]B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973), Vol. **I**, p. 369.
[19]D. G. Davies and A. Howie, "Applications of scanning electron acoustic microscopy," Inst. Phys. Conf. Ser. (1993), No. 68: chapter 12.
[20]X. C. Wang and M. Shao, *Basic Theory and Numerical Method of Finite Element Method* (Tsinghua U. P., Beijing, 2001) (in Chinese).

Cluster control of acoustic potential energy in a structural/acoustic cavity

Nobuo Tanaka[a] and Kozue Kobayashi
Department of Mechanical Engineering, Tokyo Metropolitan Institute of Technology, 6-6 Asahigaoka, Hino-city, Tokyo 191-0065, Japan

(Received 1 February 2005; revised 27 February 2006; accepted 28 February 2006)

This paper is concerned with cluster control of acoustic potential energy in a cluster coupled structural/acoustic cavity using both active noise control (ANC) and active vibration control (AVC). First, fundamental characteristics of a cluster coupled cavity are discussed, intriguing phenomena being confirmed: the coupling is strictly selective; structural modes and acoustic modes interfere with each other *if and only if* they fall into the same cluster, and hence termed *cluster coupling*. It is found that acoustic potential energy is a resultant from each cluster contribution, and thus clustered acoustic potential energy in a cavity may be extracted via *cluster filtering* without causing observation spillover, and may also be controlled independently using *cluster actuation* without causing control spillover. Employing both cluster filtering and cluster actuation, an optimal cluster control law using both ANC and AVC for minimizing clustered acoustic potential energy is derived. Based upon the control law, cluster control may be achieved, enabling the implementation of a multi-SISO system that may significantly mitigate computation load, as opposed to a MIMO system, on a controller. Finally, a numerical simulation is conducted, demonstrating the cluster control effect for suppressing acoustic potential energy in the cluster coupled cavity. [DOI: 10.1121/1.2188815]

PACS number(s): 43.40.−r, 43.40.Vn, 43.50.Ki, 43.40.Dx [KAC] Pages: 2758–2771

I. INTRODUCTION

For the purpose of suppressing noise pollution, much research has been done in the past. Summarizing the active control approaches reported so far, they are classified into two categories; active noise control[1] (ANC) and active vibration control[2,3] (AVC). ANC aims to suppress the noise at sensor locations directly by employing secondary sound sources, while AVC attempts to suppress the structural vibration, the cause of noise radiation, by installing control actuators directly to the target structure. In the active noise control community, noise suppression in an open space seems to be a topic of much research interest, however, noise suppression in a closed space observed in such as aircrafts, automobiles, trains, and so on still needs to be tackled especially for a structure/acoustic coupled cavity.

The term "coupling" indicates the interference between a structural field and an acoustic field of the cavity, resulting in changing the modal frequencies and modal functions of both field systems. Depending on the degree of coupling, coupled cavities can be divided into two categories. A weakly coupled cavity falls into the first category, which is normally used and provides accurate model of the system response if the fluid medium is nondense, and the structure is not "small." Based upon modal coupling theory,[4,5] the cavity dynamics are generally expressed. In this case, eigenfrequencies of both structural field and acoustic field vary. Furthermore, eigenfunctions of acoustic cavity are affected by coupling; however the approach stands on the premise that eigenfunctions of the cavity remain the same as those of an acoustically rigid walled cavity while only the eigenfrequencies of the cavity change. In practice, it is understandable that the eigenfrequencies of the cavity are likely to vary while the eigenfunctions change little. Structural/acoustic systems displaying weak modal coupling characteristics are common amongst sound transmission problems targeted for active noise control: Pan *et al.*[6,7] discussed, analytically and experimentally, the mechanism for controlling noise transmission through a panel of the weakly (modal) coupled cavity; Snyder and Hansen[8,9] showed a theoretical and experimental framework suitable for aiding in the design of active control of sound transmission into the cavity; Snyder and Tanaka[10] investigated the sound transmission control using a shaped modal sensor; Sampath and Balachandran[11,12] studied the performance functions for the interior noise control analytically as well as experimentally; and Kim and Brennan[13] investigated the active control of harmonic sound transmission.

When cavity walls become thin and its size becomes small, such assumption is no longer valid, and this case falls into the second category called a strongly coupled cavity.[14–21] If this is the case, eigenfrequencies and eigenfunctions of structural and acoustic fields change. As a result, complexity in dealing with a cavity model substantially arises: orthogonality of acoustic modal functions is not guaranteed; eigenfunctions and eigenfrequencies of the cavity must be numerically obtained through a complicated procedure even for a simplistic model. With a view to investigating the characteristics of a strongly coupled cavity, an effort has been undertaken: Lyon[14] examined the possibility of estimating the noise reduction in various frequency ranges of the strongly coupled rectangular cavity; Dowell and Voss[15]

[a]Electronic mail: ntanaka@cc.tmit.ac.jp

 0001-4966/2006/119(5)/2758/14/$22.50

studied fundamental characteristics of the cavity or cavity-backed panel; Pretlove[16,17] derived an exact solution of the acoustic velocity potential in a strongly coupled rectangular cavity with one flexible wall; Kihlman[18] investigated the relation between wall vibration and sound pressure; Battacharya and Crocker[19] derived a purely analytical general solution of the wave equation with inhomogeneous boundary conditions; Guy and Bhattacharya[20] studied the influence of a finite cavity backing a finite panel on the sound transmission; and Guy[21] presented a multimodal solution of the airborne sound transmission.

This paper is concerned with cluster control of acoustic potential energy of a cluster coupled cavity when a flexible panel is subjected to external forces. Cluster coupling is tantamount to conventional modal coupling, however this paper adopts the term "cluster" as the authors utilized it in Refs. 22 and 23. First, this paper begins by confirming the fundamental characteristics of the cluster coupled cavity. The coupling is strictly selective; structural modes and the acoustic modes interfere with each other *if and only if* they fall into the same cluster, and hence are termed *cluster coupling*. It is found that acoustic potential energy is a resultant from each cluster contribution, and thus clustered acoustic potential energy in a cavity may be extracted via *cluster filtering*[22,23] without causing observation spillover,[24] and may also be controlled independently using *cluster actuation*[23] without causing control spillover. Furthermore, employing both cluster filtering and cluster actuation, the optimal cluster control law based upon ANC and AVC for minimizing the acoustic potential energy in a cluster coupled enclosure is derived. Using the control law, cluster control may be achieved, enabling the implementation of a multi-SISO system that may significantly mitigate computation load, as opposed to a MIMO system, on a controller. Finally, a numerical simulation is conducted, demonstrating the cluster control effect for suppressing acoustic potential energy in the cluster coupled cavity.

II. A CLUSTER COUPLED STRUCTURAL/ACOUSTIC ENCLOSURE

A. Vibration field system

Consider some generic coupled structural/acoustic enclosure as shown in Fig. 1, subject to excitation by unspecified N_d point disturbance forces and N_c point control forces. Then the equation of motion of a two-dimensional, lossless, distributed paameter structure is expressed as

$$\mathbf{L}[w(\mathbf{x},t)] + m(\mathbf{x})\frac{\partial w^2(\mathbf{x},t)}{\partial t^2} = -p(\mathbf{x},t) + \sum_{i=1}^{N_d} f_{d,i}(t)\,\delta(\mathbf{x}-\mathbf{x}_{d,i}) + \sum_{i=1}^{N_c} f_{c,i}(t)\,\delta(\mathbf{x}-\mathbf{x}_{c,i}), \tag{1}$$

where $\mathbf{L}$, $w(\mathbf{x},t)$, $m(\mathbf{x})$, $p(\mathbf{x},t)$, $f_d(\mathbf{t})$, $f_c(t)$, and δ are the self-adjoint differential operator, deflection of the panel at $\mathbf{x}=(x,y)$, mass density of the panel, sound pressure at $\mathbf{x}$ of a cavity, a disturbance force at $\mathbf{x}_d$, a control force at $\mathbf{x}_c$, and a Dirac delta function, respectively. Note that the panel is subjected to the influence of the cluster coupled circumstances, hence the sound pressure term on the right-hand side of Eq. (1).

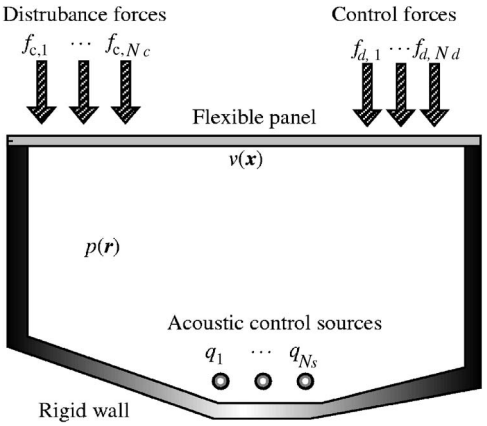


FIG. 1. Enclosure consisting of a flexible wall and rigid walls where active vibration control actuators, disturbance forces, and acoustic control sources are located.

When a disturbance force excites the panel harmonically with the driving frequency ω, the deflection of the panel is then given by

$$w(\mathbf{x},t) = w(\mathbf{x})e^{j\omega t}. \tag{2}$$

For a uniform panel, the equation of motion in Eq. (1) therefore becomes

$$D\nabla^4 w(\mathbf{x}) - \omega^2\rho_h h w(\mathbf{x}) = -p(\mathbf{x}) + \sum_{i=1}^{N_d} f_{d,i}\delta(\mathbf{x}-\mathbf{x}_{d,i}) + \sum_{i=1}^{N_c} f_{c,i}\delta(\mathbf{x}-\mathbf{x}_{c,i}), \tag{3}$$

where D is the bending stiffness of the panel, ρ_h is the density of the structure material, h is the thickness of the panel, and ∇^4 is defined as

$$\nabla^4 = \nabla^2\nabla^2 = \frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2\partial y^2} + \frac{\partial^4}{\partial y^4}. \tag{4}$$

The deflection $w(\mathbf{x})$ of the panel, limiting consideration to M_s structural modes, is then expressed as

$$w(\mathbf{x}) = \sum_{m=1}^{\infty} \varphi_m(\mathbf{x})w_m \approx \boldsymbol{\varphi}^T(\mathbf{x})\mathbf{w}, \tag{5}$$

where the superscript T denotes the transpose of the expression, $\varphi_m(\mathbf{x})$ $(m=1,2,\ldots,M_s)$ is the mth eigenfunction of the target panel, w_m is the associated modal coefficient, $\mathbf{w} \in \mathfrak{C}^{M_s}$ where $\mathfrak{C}^{M_s}$ denotes the set of all ordered M_s tuples of complex numbers, is the structure modal displacement vector, and $\boldsymbol{\varphi}(\mathbf{x}) \in \mathfrak{R}^{M_s}$ where $\mathfrak{R}^{M_s}$ denotes the set of all ordered M_s-tuples of real numbers, is the modal function vector at $\mathbf{x}$ of the panel, and are given by

$$\boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x})\varphi_2(\mathbf{x})\cdots\varphi_M,(\mathbf{x}))^T, \tag{6}$$

and

$$\mathbf{w} = (w_1 w_2, \ldots, w_{M_s})^T, \tag{7}$$

so that the velocity $v(\mathbf{x})$ at some location $\mathbf{x}$ on the enclosing structure yields

$$v(\mathbf{x}) = j\omega w(\mathbf{x}) = \boldsymbol{\varphi}^T(\mathbf{x})\mathbf{v}. \tag{8}$$

Furthermore, the sound pressure $p(\mathbf{r})$ at $\mathbf{r}=(x,y,z)$ in the enclosed space can be expressed in terms of the acoustic mode shape functions of the cavity,

$$p(\mathbf{r}) = \sum_{i=1}^{\infty} p_i \phi_i(\mathbf{r}) \approx \boldsymbol{\phi}^T(\mathbf{r})\mathbf{p}, \tag{9}$$

where p_i is the sound pressure amplitude of the ith acoustic mode and $\phi_i(\mathbf{r})$ is the associated mode shape function at the location $\mathbf{r}$ in space. $\mathbf{p} \in \mathfrak{C}^{M_a}$ and $\boldsymbol{\phi}(\mathbf{r}) \in \mathfrak{R}^{M_a}$ are the acoustic modal amplitude vector and the acoustic modal function vector where they are, respectively, defined as

$$\mathbf{p} = (p_1 p_2, \ldots, p_{M_a})^T, \tag{10}$$

and

$$\boldsymbol{\phi}(\mathbf{r}) = (\phi_1(\mathbf{r})\phi_2(\mathbf{r})\cdots\phi_{M_a}(\mathbf{r}))^T. \tag{11}$$

Note that, as a result of modal orthogonality,

$$\int_V \phi_i(\mathbf{r})\phi_j(\mathbf{r})d\mathbf{r} = V_i\delta_{ij}, \tag{12}$$

where V_i is the volume normalization of the ith acoustic mode and δ_{ij} is the Kronecker delta.

Multiplying Eq. (3) with the ith structure modal function $\varphi_i(\mathbf{x})$ and integrating it over the domain of the plate leads to

$$v_i = \frac{j\omega}{\omega_{s,i}^2 - \omega^2}\left(-\int_s \varphi_i(\mathbf{x})p(\mathbf{x})d\mathbf{x} + \sum_{k=1}^{N_d}\varphi_i(\mathbf{x}_{d,k})f_{d,k} + \sum_{k=1}^{N_c}\varphi_i(\mathbf{x}_{c,k})f_{c,k}\right), \tag{13}$$

where j and $\omega_{s,i}$ are the imaginary unit and the ith structure modal frequency, respectively. Substituting Eq. (9) into the integration term in Eq. (13) leads to

$$-\int_s \varphi_i(\mathbf{x})p(\mathbf{x})d\mathbf{x} = -\int_s \varphi_i(\mathbf{x})\boldsymbol{\phi}^T(\mathbf{x})d\mathbf{x}\mathbf{p} = -\bar{\boldsymbol{\beta}}_i^T\mathbf{p}, \tag{14}$$

where

$$\bar{\boldsymbol{\beta}}_i = (\beta_{1i}\beta_{2i}, \ldots, \beta_{M_a i})^T \tag{15}$$

and where β_{ij} is the modal coupling coefficient[19] for the ith acoustic mode and jth structural mode,

$$\beta_{ij} = \int_s \phi_i(\mathbf{x})\varphi_j(\mathbf{x})d\mathbf{x}. \tag{16}$$

Substituting Eq. (14) into Eq. (13) and after some manipulations of the formulas, the modal velocity vector becomes

$$\mathbf{v} = \tilde{\boldsymbol{\Lambda}}_s(-\mathbf{B}_{\text{couple}}^T\mathbf{p} + \mathbf{Z}_c\mathbf{f}_c + \mathbf{Z}_d\mathbf{f}_d), \tag{17}$$

where

$$\tilde{\boldsymbol{\Lambda}}_s = \begin{pmatrix} \frac{j\omega}{\omega_{s,1}^2-\omega^2} & 0 & \cdots & 0 \\ 0 & \frac{j\omega}{\omega_{s,2}^2-\omega^2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & \frac{j\omega}{\omega_{s,M_s}^2-\omega^2} \end{pmatrix} \in \mathfrak{I}^{M_s\times M_s}, \tag{18}$$

$$\mathbf{B}_{\text{couple}} = \begin{pmatrix} \boldsymbol{\beta}_1^T \\ \boldsymbol{\beta}_2^T \\ \vdots \\ \boldsymbol{\beta}_{Ma}^T \end{pmatrix} = (\bar{\boldsymbol{\beta}}_1\ \bar{\boldsymbol{\beta}}_2\ \bar{\boldsymbol{\beta}}_{Ms}) \in \mathfrak{R}^{M_a\times M_s}, \tag{19}$$

$$\mathbf{Z}_c = \begin{pmatrix} \varphi_1(\mathbf{x}_{c1}) & \varphi_1(\mathbf{x}_{c2}) & \cdots & \varphi_1(\mathbf{x}_{cN_c}) \\ \varphi_2(\mathbf{x}_{c1}) & \varphi_2(\mathbf{x}_{c2}) & \cdots & \varphi_2(\mathbf{x}_{cN_c}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{M_s}(\mathbf{x}_{c1}) & \varphi_{M_s}(\mathbf{x}_{c2}) & \cdots & \varphi_{M_s}(\mathbf{x}_{cN_c}) \end{pmatrix} \in \mathfrak{R}^{M_s\times N_c}, \tag{20}$$

$$\mathbf{Z}_d = \begin{pmatrix} \varphi_1(\mathbf{x}_{d1}) & \varphi_1(\mathbf{x}_{d2}) & \cdots & \varphi_1(\mathbf{x}_{dN_d}) \\ \varphi_2(\mathbf{x}_{d1}) & \varphi_2(\mathbf{x}_{d2}) & \cdots & \varphi_2(\mathbf{x}_{dN_d}) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{M_s}(\mathbf{x}_{d1}) & \varphi_{M_s}(\mathbf{x}_{d2}) & \cdots & \varphi_{M_s}(\mathbf{x}_{dN_d}) \end{pmatrix} \in \mathfrak{R}^{M_s\times N_d}, \tag{21}$$

$$\mathbf{f}_c = (f_{c,1}\ f_{c,2}\ \cdots\ f_{c,N_c})^T \in \mathfrak{C}^{N_c}, \tag{22}$$

$$\mathbf{f}_d = (f_{d,1}\ f_{d,2}\ \cdots\ f_{d,N_d})^T \in \mathfrak{C}^{N_d}, \tag{23}$$

and where $\mathfrak{I}^{M_s\times M_s}$ and $\mathfrak{R}^{M_a\times M_s}$ denote the class of all $M_s \times M_s$ matrices with pure imaginary number elements and all $M_a \times M_s$ matrices with real number elements, respectively.

B. Acoustic field system

Sound pressure $p(\mathbf{r})$ that is a resultant from the surface velocity of the enclosing structure and N_s sound sources for ANC with the volume velocity q_i at $\mathbf{r}_i$ is defined by the Green's function response equation,

$$p(\mathbf{r}) = j\omega\rho_0\int_s \mathrm{G}(\mathbf{x}|\mathbf{r})v(\mathbf{x})d\mathbf{x} + j\omega\rho_0\sum_{i=1}^{N_s}\mathrm{G}(\mathbf{r}_i|\mathbf{r})q_i, \tag{24}$$

where $\mathrm{G}(\mathbf{x}/\mathbf{r})$ is the Green's function of the enclosed space, and is defined as

$$G(\mathbf{x}|\mathbf{r}) = \sum_{i=1}^{\infty} \frac{c_0^2 \phi_i(\mathbf{x}) \phi_i(\mathbf{r})}{V_i(\omega_{a,i}^2 - \omega^2)}, \tag{25}$$

where $\omega_{a,i}$ is the eigenvalue of the ith acoustic mode, c_0 is the speed of sound in the acoustic medium, and the integration in Eq. (24) is over the surface S of the panel. Consequently, the complex amplitude of the ith acoustic mode is found to be

$$p_i = j\omega\rho_0 c_0^2 \int_s \frac{\phi_i(\mathbf{x})}{V_i(\omega_{a,i}^2 - \omega^2)} v(\mathbf{x}) d\mathbf{x} + j\omega\rho_0 c_0^2 \sum_{j=1}^{N_s} \frac{\phi_i(\mathbf{x}_j)}{V_i(\omega_{a,i}^2 - \omega^2)} q_j, \tag{26}$$

where the surface velocity $v(\mathbf{x})$ of the panel may be expanded in terms of the mode shape functions, enabling the amplitude of the ith acoustic mode to be expressed as

$$p_i = \frac{j\omega\rho_0 c_0^2}{V_i(\omega_{a,i}^2 - \omega^2)} (\boldsymbol{\beta}_i^T \mathbf{v} + \bar{\boldsymbol{\phi}}_i^T \mathbf{q}), \tag{27}$$

where

$$\bar{\boldsymbol{\phi}}_i = (\phi_i(\mathbf{r}_1) \phi_i(\mathbf{r}_2) \cdots \phi_i(\mathbf{r}_{N_s}))^T, \tag{28}$$

and

$$\mathbf{q} = (q_1 q_2 \cdots q_{N_s})^T. \tag{29}$$

From Eq. (27), the acoustic modal amplitude vector is given by

$$\mathbf{p} = \tilde{\boldsymbol{\Lambda}}_a (\mathbf{B}_{\text{couple}} \mathbf{v} + \boldsymbol{\Phi} \mathbf{q}), \tag{30}$$

where

$$\tilde{\boldsymbol{\Lambda}}_a = \begin{pmatrix} \frac{j\rho_0 c_0^2 \omega}{V_1(\omega_{a,1}^2 - \omega^2)} & 0 & \cdots & 0 \\ 0 & \frac{j\rho_0 c_0^2 \omega}{V_2(\omega_{a,2}^2 - \omega^2)} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{j\rho_0 c_0^2 \omega}{V_{M_a}(\omega_{a,M_a}^2 - \omega^2)} \end{pmatrix} \in \mathfrak{I}^{M_a \times M_a}, \tag{31}$$

and

$$\boldsymbol{\Phi} = \begin{pmatrix} \bar{\boldsymbol{\phi}}_1^T \\ \bar{\boldsymbol{\phi}}_2^T \\ \vdots \\ \bar{\boldsymbol{\phi}}_{M_a}^T \end{pmatrix} = (\boldsymbol{\phi}(\mathbf{r}_1)\ \boldsymbol{\phi}(r_2)\ \cdots\ \boldsymbol{\phi}(\mathbf{r}_{M_a})) = \begin{pmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \cdots & \phi_1(\mathbf{r}_{N_s}) \\ \phi_2(\mathbf{r}_1) & \phi_2(\mathbf{r}_2) & \cdots & \phi_2(\mathbf{r}_{N_s}) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{M_a}(\mathbf{r}_1) & \phi_{M_a}(\mathbf{r}_2) & \cdots & \phi_{M_a}(\mathbf{r}_{N_s}) \end{pmatrix} \in \mathfrak{R}^{M_a \times N_s}. \tag{32}$$

C. Cluster coupled structural/acoustic system

A cluster coupled structural/acoustic system equation may be obtained by combining both the vibration field equation in Eq. (17) and the acoustic field equation in Eq. (30). By substituting Eq. (30) into Eq. (17), the modal amplitude velocity vector is given by

$$\mathbf{v} = (\mathbf{I}_{M_s} + \tilde{\boldsymbol{\Lambda}}_s \mathbf{B}_{\text{couple}}^T \tilde{\boldsymbol{\Lambda}}_a \mathbf{B}_{\text{couple}})^{-1} \times \tilde{\boldsymbol{\Lambda}}_s (-\mathbf{B}_{\text{couple}}^T \tilde{\boldsymbol{\Lambda}}_a \boldsymbol{\Phi} \mathbf{q} + \mathbf{Z}_c \mathbf{f}_c + \mathbf{Z}_d \mathbf{f}_d), \tag{33}$$

where $\mathbf{I}_{M_s}$, denotes the identity matrix with the order of M_s. Similarly, substituting Eq. (17) into Eq. (30), the acoustic modal amplitude vector yields

$$\mathbf{p} = (\mathbf{I}_{M_a} + \tilde{\boldsymbol{\Lambda}}_a \mathbf{B}_{\text{couple}} \tilde{\boldsymbol{\Lambda}}_s \mathbf{B}_{\text{couple}}^T)^{-1} \times \tilde{\boldsymbol{\Lambda}}_a \{\boldsymbol{\Phi} \mathbf{q} + \mathbf{B}_{\text{couple}} \tilde{\boldsymbol{\Lambda}}_s (\mathbf{Z}_c \mathbf{f}_c + \mathbf{Z}_d \mathbf{f}_d)\}. \tag{34}$$

From Eqs. (33) and (34), it can be found that the properties of a cluster coupled structural/acoustic enclosure are governed by the following characteristic equation:

$$\det(\mathbf{I}_{M_a} + \tilde{\boldsymbol{\Lambda}}_a \mathbf{B}_{\text{couple}} \tilde{\boldsymbol{\Lambda}}_s \mathbf{B}_{\text{couple}}^T) = \det(\mathbf{I}_{M_s} + \tilde{\boldsymbol{\Lambda}}_s \mathbf{B}_{\text{couple}}^T \tilde{\boldsymbol{\Lambda}}_a \mathbf{B}_{\text{couple}}) = 0. \tag{35}$$

Assume that all the structural modes of a rectangular panel are classified into four clusters[22,23]—A_s, B_s, C_s, and D_s. The clustered structure modal function vector in Eq. (6) may then be written as

$$\boldsymbol{\varphi}(\mathbf{x}) = (\boldsymbol{\varphi}_{\mathrm{A}_s}(\mathbf{x})\ \boldsymbol{\varphi}_{\mathrm{B}_s}(\mathbf{x})\ \boldsymbol{\varphi}_{\mathrm{C}_s}(\mathbf{x})\ \boldsymbol{\varphi}_{\mathrm{D}_s}(\mathbf{x}))^T, \tag{36}$$

where $\boldsymbol{\varphi}_{\mathrm{A}_s}(\mathbf{x}) \in \mathfrak{R}^{M_{s\mathrm{A}}}$, $\boldsymbol{\varphi}_{\mathrm{B}_s}(\mathbf{x}) \in \mathfrak{R}^{M_{s\mathrm{B}}}$, $\boldsymbol{\varphi}_{\mathrm{C}_s}(\mathbf{x}) \in \mathfrak{R}^{M_{s\mathrm{C}}}$ and $\boldsymbol{\varphi}_{\mathrm{D}_s}(\mathbf{x}) \in \mathfrak{R}^{M_{s\mathrm{D}}}$. Similarly, when all the acoustic modes of a rectangular cavity are grouped into the clusters: A_a, B_a, C_a, and D_a, the acoustic modal amplitude vector in Eq. (11) is given by

$$\boldsymbol{\phi}(\mathbf{r}) = (\boldsymbol{\phi}_{\mathrm{A}_a}(\mathbf{r})\ \boldsymbol{\phi}_{\mathrm{B}_a}(\mathbf{r})\ \boldsymbol{\phi}_{\mathrm{C}_a}(\mathbf{r})\ \boldsymbol{\phi}_{\mathrm{D}_a}(\mathbf{r}))^T, \tag{37}$$

where $\boldsymbol{\phi}_{\mathrm{A}_a}(\mathbf{r}) \in \mathfrak{R}^{M_{s\mathrm{A}}}$, $\boldsymbol{\phi}_{\mathbf{B}_a}(\mathbf{r}) \in \mathfrak{R}^{M_{s\mathrm{B}}}$, $\boldsymbol{\phi}_{\mathrm{C}_a}(\mathbf{r}) \in \mathfrak{R}^{M_{s\mathrm{C}}}$ and $\boldsymbol{\phi}_{\mathrm{D}_a}(\mathbf{r}) \in \mathfrak{R}^{M_{s\mathrm{D}}}$.c

Note that the acoustic modal indices of the rectangular cavity are described by three indices in the x, y, and z axes, and the two in the x and y axes are used as they meet in the x-y plane of the cavity.

Cluster coupling means that when structural modes and acoustic modes fall into the same common cluster; $\mathrm{A} \equiv \mathrm{A}_s \equiv \mathrm{A}_a$, $\mathrm{B} \equiv \mathrm{B}_s \equiv \mathrm{B}_a$, $\mathrm{C} \equiv \mathrm{C}_s \equiv \mathrm{C}_a$, and $\mathrm{D} \equiv \mathrm{D}_s \equiv \mathrm{D}_a$, the modal pair in each cluster interferes with each other, resulting in changing the eigenvalues of each cluster coupling system independently, as will be discussed shortly.

If this is the case, the coupling coefficient β_{ij} in Eq. (16) with respect to the modal pair becomes nonzero. For example, the odd/odd structural modes do not couple with the acoustic odd/odd or odd/even or even/odd modes, thereby causing the nullification of the modal coupling coefficient. As noted by Pretlove[16,17] and Bhattacharya,[19] the modal coupling or parity-dependent coupling of the cavity modes and the panel modes is so selective that the radiation or coupling may be divided into sub-domains containing only those matched parity modes. Khilman[18] also estimated the parity-dependent coupling coefficient of a panel mode with all the possible cavity modes. Thus the terminology "cluster coupling" used in the work is tantamount to the "modal coupling" or "parity-dependent coupling" used in the work is tantamount to the "modal coupling" or "parity-dependent coupling" used in the former arts; however, in order to maintain the consistency as it is already utilized in the authors' relevant papers,[22,23] this paper employs the label "cluster."

Hence, all the structural/acoustic modal combinations are summarized as follows:

Cluster A: It consists of all the odd/odd structural modes, $\varphi_{\mathrm{odd/odd}}(\mathbf{x})$, and all the even/even acoustic modes, $\phi_{\mathrm{even/even}}(\mathbf{x})$.

Cluster B: It consists of all the odd/even structural modes, $\varphi_{\mathrm{odd/even}}(\mathbf{x})$, and all the even/odd acoustic modes, $\phi_{\mathrm{even/odd}}(\mathbf{x})$.

Cluster C: It consists of all the even/odd structural modes, $\varphi_{\mathrm{even/odd}}(\mathbf{x})$, and all the odd/even acoustic modes, $\phi_{\mathrm{odd/even}}(\mathbf{x})$.

Cluster D: It consists of all the even/even structural modes, $\varphi_{\mathrm{even/even}}(\mathbf{x})$, and all the odd/odd acoustic modes, $\phi_{\mathrm{odd/odd}}(\mathbf{x})$.

Boundary conditions of the structural modes of a rectangular panel satisfying the above-mentioned 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. Thus all the structural/acoustic modes of a rectangular panel/cavity system are classified into four kinds of clusters.

Suppose that all structural modes are orthogonal, and then the clusters consisting of those modes are also orthogonal to each other. Furthermore, taking into consideration the modal orthogonality between the structural modes and acoustic modes, the modal coupling matrix defined in Eq. (19) is written as

$$\mathbf{B}_{\mathrm{couple}} = \begin{pmatrix} \mathbf{B}_{\mathrm{A}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{\mathrm{B}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{B}_{\mathrm{C}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{\mathrm{D}} \end{pmatrix}. \tag{38}$$

Similarly, the matrices $\tilde{\mathbf{\Lambda}}_s$ and $\tilde{\mathbf{\Lambda}}_a$ may also be decoupled into the following:

$$\tilde{\mathbf{\Lambda}}_s = \begin{pmatrix} \tilde{\mathbf{\Lambda}}_{s\mathrm{A}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Lambda}}_{s\mathrm{B}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{\Lambda}}_{s\mathrm{C}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\mathbf{\Lambda}}_{s\mathrm{D}} \end{pmatrix}, \tag{39}$$

and

$$\tilde{\mathbf{\Lambda}}_a = \begin{pmatrix} \tilde{\mathbf{\Lambda}}_{a\mathrm{A}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{\Lambda}}_{a\mathrm{B}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\mathbf{\Lambda}}_{a\mathrm{C}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\mathbf{\Lambda}}_{a\mathrm{D}} \end{pmatrix}. \tag{40}$$

Substituting Eqs. (38)–(40) into Eq. (35), the cluster coupling characteristic equation expands to

$$\begin{aligned} &\det(\mathbf{I}_{M_s} + \tilde{\mathbf{\Lambda}}_a \mathbf{B}_{\mathrm{couple}} \tilde{\mathbf{\Lambda}}_s \mathbf{B}_{\mathrm{couple}}^T) \\ &= \det\begin{pmatrix} \mathbf{\Delta}_{\mathrm{A}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta}_{\mathrm{B}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Delta}_{\mathrm{C}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Delta}_{\mathrm{D}} \end{pmatrix} \\ &= \begin{pmatrix} \det(\mathbf{\Delta}_{\mathrm{A}}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \det(\mathbf{\Delta}_{\mathrm{B}}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \det(\mathbf{\Delta}_{\mathrm{C}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \det(\mathbf{\Delta}_{\mathrm{D}}) \end{pmatrix} = 0, \end{aligned} \tag{41}$$

where

$$\mathbf{\Delta}_{\mathrm{A}} = \mathbf{I}_{\mathrm{A}} + \tilde{\mathbf{\Lambda}}_{a\mathrm{A}} \mathbf{B}_{\mathrm{A}} \tilde{\mathbf{\Lambda}}_{s\mathrm{A}} \mathbf{B}_{\mathrm{A}}^T, \tag{42}$$

$$\mathbf{\Delta}_{\mathrm{B}} = \mathbf{I}_{\mathrm{B}} + \tilde{\mathbf{\Lambda}}_{a\mathrm{B}} \mathbf{B}_{\mathrm{B}} \tilde{\mathbf{\Lambda}}_{s\mathrm{B}} \mathbf{B}_{\mathrm{B}}^T, \tag{43}$$

$$\mathbf{\Delta}_{\mathrm{C}} = \mathbf{I}_{\mathrm{C}} + \tilde{\mathbf{\Lambda}}_{a\mathrm{C}} \mathbf{B}_{\mathrm{C}} \tilde{\mathbf{\Lambda}}_{s\mathrm{C}} \mathbf{B}_{\mathrm{C}}^T, \tag{44}$$

$$\boldsymbol{\Delta}_{\mathrm{D}} = \mathbf{I}_{\mathrm{D}} + \tilde{\boldsymbol{\Lambda}}_{a\mathrm{D}}\mathbf{B}_{\mathrm{D}}\tilde{\boldsymbol{\Lambda}}_{s\mathrm{D}}\mathbf{B}_{\mathrm{D}}^{T}. \tag{45}$$

Moreover, the inverse matrix of $(\mathbf{I}_{M_s} + \tilde{\boldsymbol{\Lambda}}_a\mathbf{B}_{\mathrm{couple}}\tilde{\boldsymbol{\Lambda}}_s\mathbf{B}_{\mathrm{couple}}^{T})$ becomes

$$\boldsymbol{\Delta}_{a,\mathrm{inv}} = (\mathbf{I}_{M_s} + \tilde{\boldsymbol{\Lambda}}_a\mathbf{B}_{\mathrm{couple}}\tilde{\boldsymbol{\Lambda}}_s\mathbf{B}_{\mathrm{couple}}^{T})^{-1} = \begin{pmatrix} \boldsymbol{\Delta}_{\mathrm{A}}^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_{\mathrm{B}}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}_{\mathrm{C}}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Delta}_{\mathrm{D}}^{-1} \end{pmatrix}. \tag{46}$$

Clearly from Eq. (46), the eigenvalues of each cluster are independently determined by the cluster characteristic equation, thus causing no spillover on the other clusters.

III. ACOUSTIC POTENTIAL ENERGY IN A CLUSTER COUPLED CAVITY

A. Acoustic potential energy

Acoustic potential energy E_p in a cavity may be described by

$$E_p = \frac{1}{4\rho_0 c_0^2}\int_V |p(r)|^2 d\mathbf{r}, \tag{47}$$

where V is the volume of the enclosure. By substituting Eq. (9) into Eq. (47), the potential energy E_p becomes

$$E_p = \frac{1}{4\rho_0 c_0^2}\int_V \mathbf{p}^H \boldsymbol{\phi}(r)\boldsymbol{\phi}^T(r)\mathbf{p}\, d\mathbf{r}, \tag{48}$$

where the superscript H represents the Hermitian (conjugate transpose). Furthermore, considering the orthogonality of acoustic modal functions in Eq. (12), the acoustic potential energy E_p is therefore described as

$$E_p = \mathbf{p}^H \boldsymbol{\Lambda}_V \mathbf{p}, \tag{49}$$

where

$$\boldsymbol{\Lambda}_V = \frac{1}{4\rho_0 c_0^2}\begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & V_{M_a} \end{pmatrix} \in \mathfrak{R}^{M_a \times M_a}. \tag{50}$$

As discussed before, the matrix $\boldsymbol{\Lambda}_V$ can further be grouped into the four cluster blocks,

$$\boldsymbol{\Lambda}_V = \begin{pmatrix} \boldsymbol{\Lambda}_{V\mathrm{A}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{V\mathrm{B}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_{V\mathrm{C}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Lambda}_{V\mathrm{D}} \end{pmatrix}. \tag{51}$$

Substituting Eq. (34) into Eq. (49), the acoustic potential energy E_p is given by

$$E_p = \mathbf{q}^H\mathbf{A}_p\mathbf{q} + \mathbf{f}_C^H\mathbf{B}_p^H\mathbf{q} + \mathbf{f}_d^H\mathbf{C}_p^H\mathbf{q} + \mathbf{q}^H\mathbf{B}_p\mathbf{f}_c + \mathbf{f}_c^H\mathbf{D}_p\mathbf{f}_c + \mathbf{f}_d^H\mathbf{E}_p^H\mathbf{f}_c + \mathbf{q}^H\mathbf{C}_p\mathbf{f}_d + \mathbf{f}_c^H\mathbf{E}_p\mathbf{f}_d + E_{p,\mathrm{without}}, \tag{52}$$

where the matrices in Eq. (52) are defined as follows:

$$\mathbf{A}_p = \Phi^T\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a\boldsymbol{\Phi} \in \mathfrak{R}^{N_s\times N_s}, \tag{53}$$

$$\mathbf{B}_p^H = \mathbf{Z}_c^T\tilde{\boldsymbol{\Lambda}}_s^*\mathbf{B}_{\mathrm{couple}}^T\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a\boldsymbol{\Phi} \in \mathfrak{I}^{N_s\times N_c}, \tag{54}$$

$$\mathbf{C}_p^H = \mathbf{Z}_c^T\tilde{\boldsymbol{\Lambda}}_s^*\mathbf{B}_{\mathrm{couple}}^T\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a\boldsymbol{\Phi} \in \mathfrak{I}^{N_s\times N_c}, \tag{55}$$

$$\mathbf{D}_p = \mathbf{Z}_c^T\tilde{\boldsymbol{\Lambda}}_s^*\mathbf{B}_{\mathrm{couple}}^T\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a\mathbf{B}_{\mathrm{couple}}\tilde{\boldsymbol{\Lambda}}_s\mathbf{Z}_c \in \mathfrak{R}^{N_c\times N_c}, \tag{56}$$

$$\mathbf{E}_p = \mathbf{Z}_c^T\tilde{\boldsymbol{\Lambda}}_s^*\mathbf{B}_{\mathrm{couple}}^T\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a\mathbf{B}_{\mathrm{couple}}\tilde{\boldsymbol{\Lambda}}_s\mathbf{Z}_d \in \mathfrak{R}^{N_c\times N_d}, \tag{57}$$

$$E_{p,\mathrm{without}} = \mathbf{Z}_d^T\tilde{\boldsymbol{\Lambda}}_s^*\mathbf{B}_{\mathrm{couple}}^T\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a\mathbf{B}_{\mathrm{couple}}\tilde{\boldsymbol{\Lambda}}_s\mathbf{Z}_d \in \mathfrak{R}^{1\times 1}, \tag{58}$$

and where the asterisk and $E_{p,\mathrm{without}}$ denote complex conjugate and the acoustic potential energy in the cavity without control, respectively.

Here the common term $\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a$ in Eqs. (53)–(58) further expands to

$$\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a = \begin{pmatrix} \tilde{\boldsymbol{\Lambda}}_{\mathrm{A}}^*\boldsymbol{\Delta}_{\mathrm{A}}^{T,-1}\boldsymbol{\Lambda}_{V,\mathrm{A}}\boldsymbol{\Delta}_{\mathrm{A}}^{-1}\tilde{\boldsymbol{\Lambda}}_{\mathrm{A}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\boldsymbol{\Lambda}}_{\mathrm{B}}^*\boldsymbol{\Delta}_{\mathrm{B}}^{T,-1}\boldsymbol{\Lambda}_{V,\mathrm{B}}\boldsymbol{\Delta}_{\mathrm{B}}^{-1}\tilde{\boldsymbol{\Lambda}}_{\mathrm{B}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \tilde{\boldsymbol{\Lambda}}_{\mathrm{C}}^*\boldsymbol{\Delta}_{\mathrm{C}}^{T,-1}\boldsymbol{\Lambda}_{V,\mathrm{C}}\boldsymbol{\Delta}_{\mathrm{C}}^{-1}\tilde{\boldsymbol{\Lambda}}_{\mathrm{C}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \tilde{\boldsymbol{\Lambda}}_{\mathrm{D}}^*\boldsymbol{\Delta}_{\mathrm{D}}^{T,-1}\boldsymbol{\Lambda}_{V,\mathrm{D}}\boldsymbol{\Delta}_{\mathrm{D}}^{-1}\tilde{\boldsymbol{\Lambda}}_{\mathrm{D}} \end{pmatrix}. \tag{59}$$

Thus, due to the structure of the matrix $\tilde{\boldsymbol{\Lambda}}_a^*\boldsymbol{\Delta}_{a,\mathrm{inv}}^T\boldsymbol{\Lambda}_V\boldsymbol{\Delta}_{a,\mathrm{inv}}\tilde{\boldsymbol{\Lambda}}_a$, acoustic potential energy is decomposed into the four clusters, so that the matrices $\mathbf{A}_p$, $\mathbf{B}_p$, $\mathbf{C}_p$, $\mathbf{D}_p$, and $\mathbf{E}_p$ are partitioned into four cluster matrices, and hence the acoustic potential energy may be expressed as the sum of the acoustic potential energy in each cluster,

$$E_p = E_A + E_B + E_C + E_D. \tag{60}$$

B. Cluster filtering of acoustic potential energy

Employing cluster filtering,[22,23] a way to estimate the acoustic potential energy is presented. Assume that the acoustic point sensors at $\mathbf{r}_{s1}-\mathbf{r}_{s4}$ are situated such that the following descriptions are satisfied:

$$\mathbf{r}_{s1} = (x_s, y_s, z_s), \tag{61}$$

$$\mathbf{r}_{s2} = (L_x - x_s, y_s, z_s), \tag{62}$$

$$\mathbf{r}_{s3} = (L_x - x_s, L_x - y_s, z_s), \tag{63}$$

$$\mathbf{r}_{s4} = (x_s, L_x - y_s, z_s), \tag{64}$$

where

$$x_s \ll L_x, y_s \ll L_y, z_s \ll L_z. \tag{65}$$

By combining the four sound sensor outputs (sound pressure output) into the sound pressure vector $\mathbf{p}_c$, we have

$$\mathbf{p}_c = \begin{pmatrix} p(\mathbf{r}_{s1}) \\ p(\mathbf{r}_{s2}) \\ p(\mathbf{r}_{s3}) \\ p(\mathbf{r}_{s4}) \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix}. \tag{66}$$

Introducing the Hadamard matrix $\mathbf{T}$, the sound pressure vector $\mathbf{p}_c$ is transformed into

$$\tilde{\mathbf{p}} = \begin{pmatrix} \tilde{p}_A \\ \tilde{p}_B \\ \tilde{p}_C \\ \tilde{p}_D \end{pmatrix} = \mathbf{T}\mathbf{p}_c, \tag{67}$$

where

$$\mathbf{T} = (\mathbf{t}_A\ \mathbf{t}_B\ \mathbf{t}_C\ \mathbf{t}_D) = \begin{pmatrix} \mathbf{t}_A^T \\ \mathbf{t}_B^T \\ \mathbf{t}_C^T \\ \mathbf{t}_D^T \end{pmatrix} \tag{68}$$

and where

$$\mathbf{t}_A = (1\ 1\ 1\ 1)^T, \tag{69}$$

$$\mathbf{t}_B = (1\ 1\ -1\ -1)^T, \tag{70}$$

$$\mathbf{t}_C = (1\ -1\ -1\ 1)^T, \tag{71}$$

$$\mathbf{t}_D = (1\ -1\ 1\ -1)^T. \tag{72}$$

Taking into consideration the properties of acoustic eigenfunction, the elements of $\tilde{\mathbf{p}}$ in Eq. (67) are expressed as

$$\tilde{p}_A = \sum_{i=1}^{\infty} p_{i,\text{even/even}} \phi_{i,\text{even/even}} \approx \boldsymbol{\phi}_A^T \mathbf{p}_A, \tag{73}$$

$$\tilde{p}_B = \sum_{i=1}^{\infty} p_{i,\text{even/odd}} \phi_{i,\text{even/odd}} \approx \boldsymbol{\phi}_B^T \mathbf{p}_B, \tag{74}$$

$$\tilde{p}_C = \sum_{i=1}^{\infty} p_{i,\text{odd/even}} \phi_{i,\text{odd/even}} \approx \boldsymbol{\phi}_C^T \mathbf{p}_C, \tag{75}$$

$$\tilde{p}_D = \sum_{i=1}^{\infty} p_{i,\text{odd/odd}} \phi_{i,\text{odd/odd}} \approx \boldsymbol{\phi}_D^T \mathbf{p}_D. \tag{76}$$

Clearly from Eqs. (73)–(76), $\tilde{p}_A$, $\tilde{p}_B$, $\tilde{p}_C$, and $\tilde{p}_D$ express the summation of all the sound pressure belonging to clusters A, B, C, and D, respectively. Thus, it may extract each cluster attributes by introducing the cluster filtering processing.

Therefore, estimated acoustic potential energy of each cluster may be given by

$$\hat{E}_A = \frac{V}{4\rho_0 c_0^2} |\tilde{p}_A|^2, \tag{77}$$

$$\hat{E}_B = \frac{V}{4\rho_0 c_0^2} |\tilde{p}_B|^2, \tag{78}$$

$$\hat{E}_C = \frac{V}{4\rho_0 c_0^2} |\tilde{p}_C|^2, \tag{79}$$

$$\hat{E}_D = \frac{V}{4\rho_0 c_0^2} |\tilde{p}_D|^2. \tag{80}$$

C. Structural cluster actuation of acoustic potential energy

Suppose that the vibration control forces $f_{c1}-f_{c4}$, the amplitude of which is unity, are located at $\mathbf{x}_{c1}-\mathbf{x}_{c4}$ such that the following descriptions are satisfied:

$$\mathbf{x}_{c1} = (x_c, y_c), \tag{81}$$

$$\mathbf{x}_{c2} = (L_x - x_c, y_c), \tag{82}$$

$$\mathbf{x}_{c3} = (L_x - x_c, L_y - y_c), \tag{83}$$

$$\mathbf{x}_{c4} = (x_c, L_y - y_c), \tag{84}$$

where

$$x_c \ll L_x, L_c \ll L_y. \tag{85}$$

Incorporating the four control forces into the control force vector $\mathbf{f}$, we have

$$\mathbf{f} = \begin{pmatrix} f_{c1}(\mathbf{x}_{c1}) \\ f_{c2}(\mathbf{x}_{c2}) \\ f_{c3}(\mathbf{x}_{c3}) \\ f_{c4}(\mathbf{x}_{c4}) \end{pmatrix} = \begin{pmatrix} \delta(\mathbf{x} - \mathbf{x}_{c1}) \\ \delta(\mathbf{x} - \mathbf{x}_{c2}) \\ \delta(\mathbf{x} - \mathbf{x}_{c3}) \\ \delta(\mathbf{x} - \mathbf{x}_{c4}) \end{pmatrix}. \tag{86}$$

When driving these four control forces in phase with the amplitude f_A, the resultant force is then given by

$$\tilde{f}_{\mathrm{A}} = f_{\mathrm{A}}\{f_{c1}(\mathbf{x}_{c1}) + f_{c2}(\mathbf{x}_{c2}) + f_{c3}(\mathbf{x}_{c3}) + f_{c4}(\mathbf{x}_{c4})\}$$
$$= f_{\mathrm{A}}\{\delta(\mathbf{x}-\mathbf{x}_{c1}) + \delta(\mathbf{x}-\mathbf{x}_{c2}) + \delta(\mathbf{x}-\mathbf{x}_{c3}) + \delta(\mathbf{x}-\mathbf{x}_{c4})\}. \quad (87)$$

Then, as a result of applying $\tilde{f}_{\mathrm{A}}$ to excite the panel, the velocity modal amplitude v_i of the panel in Eq. (13) yields

$$v_i^{\mathrm{A}} = \frac{j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{A}}\{\varphi_i(\mathbf{x}_{c1}) + \varphi_i(\mathbf{x}_{c2}) + \varphi_i(\mathbf{x}_{c3}) + \varphi_i(\mathbf{x}_{c4})\}$$
$$= \frac{4j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{A}}\varphi_i^{\mathrm{A}}(\mathbf{x}_{c1}), \quad (88)$$

where the superscript A denotes a whole collection of the cluster "A" elements. Apparently from Eq. (88), only the odd/odd structural modes belonging to the cluster "A" are excited as a result of applying the transformed force $\tilde{f}_{\mathrm{A}}$ in Eq. (87). Note that all the odd/odd structural modes interfere with all the even/even acoustic modes as both of them belong to the same cluster "A," and hence exciting the panel via $\tilde{f}_{\mathrm{A}}$ leads to the direct activation of even/even acoustic modes without causing spillover.

In the same way, exciting the panel with the four point forces with an appropriate phase arrangement as described in the following permits one to activate the following cluster: For cluster B (odd/even structural modes),

$$\tilde{f}_{\mathrm{B}} = f_{\mathrm{B}}\{f_{c1}(\mathbf{x}_{c1}) + f_{c2}(\mathbf{x}_{c2}) - f_{c3}(\mathbf{x}_{c3}) - f_{c4}(\mathbf{x}_{c4})\}$$
$$= f_{\mathrm{B}}\{\delta(\mathbf{x}-\mathbf{x}_{c1}) + \delta(\mathbf{x}-\mathbf{x}_{c2}) - \delta(\mathbf{x}-\mathbf{x}_{c3}) - \delta(\mathbf{x}-\mathbf{x}_{c4})\} \quad (89)$$

then

$$v_i^{\mathrm{B}} = \frac{j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{A}}\{\varphi_i(\mathbf{x}_{c1}) + \varphi_i(\mathbf{x}_{c2}) - \varphi_i(\mathbf{x}_{c3}) - \varphi_i(\mathbf{x}_{c4})\}$$
$$= \frac{4j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{B}}\varphi_i^{\mathrm{B}}(\mathbf{x}_{c1}). \quad (90)$$

For cluster C (even/odd structural modes),

$$\tilde{f}_{\mathrm{C}} = f_{\mathrm{C}}\{f_{c1}(\mathbf{x}_{c1}) - f_{c2}(\mathbf{x}_{c2}) - f_{c3}(\mathbf{x}_{c3}) + f_{c4}(\mathbf{x}_{c4})\}$$
$$= f_{\mathrm{C}}\{\delta(\mathbf{x}-\mathbf{x}_{c1}) - \delta(\mathbf{x}-\mathbf{x}_{c2}) - \delta(\mathbf{x}-\mathbf{x}_{c3}) + \delta(\mathbf{x}-\mathbf{x}_{c4})\}, \quad (91)$$

then

$$v_i^{\mathrm{C}} = \frac{j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{C}}\{\varphi_i(\mathbf{x}_{c1}) - \varphi_i(\mathbf{x}_{c2}) - \varphi_i(\mathbf{x}_{c3}) + \varphi_i(\mathbf{x}_{c4})\}$$
$$= \frac{4j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{C}}\varphi_i^{\mathrm{C}}(\mathbf{x}_{c1}). \quad (92)$$

For cluster D (even/even structural modes),

$$\tilde{f}_{\mathrm{D}} = f_{\mathrm{D}}\{f_{c1}(\mathbf{x}_{c1}) - f_{c2}(\mathbf{x}_{c2}) + f_{c3}(\mathbf{x}_{c3}) - f_{c4}(\mathbf{x}_{c4})\}$$
$$= f_{\mathrm{D}}\{\delta(\mathbf{x}-\mathbf{x}_{c1}) - \delta(\mathbf{x}-\mathbf{x}_{c2}) + \delta(\mathbf{x}-\mathbf{x}_{c3}) - \delta(\mathbf{x}-\mathbf{x}_{c4})\}, \quad (93)$$

then

$$v_i^{\mathrm{D}} = \frac{j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{D}}\{\varphi_i(\mathbf{x}_{c1}) - \varphi_i(\mathbf{x}_{c2}) + \varphi_i(\mathbf{x}_{c3}) - \varphi_i(\mathbf{x}_{c4})\}$$
$$= \frac{4j\omega}{\omega_{s,i}^2-\omega^2} f_{\mathrm{D}}\varphi_i^{\mathrm{D}}(\mathbf{x}_{c1}). \quad (94)$$

Combining the transformed forces into the cluster force vector $\tilde{\mathbf{f}}$, we have

$$\tilde{\mathbf{f}} = \begin{pmatrix} \tilde{f}_{\mathrm{A}} \\ \tilde{f}_{\mathrm{B}} \\ \tilde{f}_{\mathrm{C}} \\ \tilde{f}_{\mathrm{D}} \end{pmatrix} = \mathbf{FTf}, \quad (95)$$

where $\mathbf{F}$ is the gain matrix and defined as

$$\mathbf{F} = \begin{pmatrix} f_{\mathrm{A}} & 0 & 0 & 0 \\ 0 & f_{\mathrm{B}} & 0 & 0 \\ 0 & 0 & f_{\mathrm{C}} & 0 \\ 0 & 0 & 0 & f_{\mathrm{D}} \end{pmatrix}. \quad (96)$$

Thus, the control force vector $\mathbf{f}$ is transformed into the cluster force vector $\tilde{\mathbf{f}}$ using the Hadamard matrix $\mathbf{T}$ assigned with a gain matrix $\mathbf{F}$, and hence termed a structural cluster actuation.

D. Acoustic cluster actuation of acoustic potential energy

Place four acoustic sources q_1-q_4 with the unity amplitude at $\mathbf{r}_{a1}-\mathbf{r}_{a4}$ such that the following descriptions are satisfied:

$$\mathbf{r}_{a1} = (x_a, y_a, z_a), \quad (97)$$

$$\mathbf{r}_{a2} = (L_x - x_a, y_a, z_a), \quad (98)$$

$$\mathbf{r}_{a3} = (L_x - x_a, L_x - y_a, z_a), \quad (99)$$

$$\mathbf{r}_{a4} = (x_a, L_x - y_a, z_a), \quad (100)$$

where

$$x_a \ll L_x, y_a \ll L_y, z_a \ll L_z. \quad (101)$$

Incorporating the four point forces into the sound source vector $\mathbf{q}$, we have

$$\mathbf{q} = \begin{pmatrix} q(\mathbf{r}_{a1}) \\ q(\mathbf{r}_{a2}) \\ q(\mathbf{r}_{a3}) \\ q(\mathbf{r}_{a4}) \end{pmatrix} = \begin{pmatrix} \delta(\mathbf{r}-\mathbf{r}_{a1}) \\ \delta(\mathbf{r}-\mathbf{r}_{a2}) \\ \delta(\mathbf{r}-\mathbf{r}_{a3}) \\ \delta(\mathbf{r}-\mathbf{r}_{a4}) \end{pmatrix}. \quad (102)$$

When driving these four sound sources in phase with the amplitude q_{A}, the resultant sound source is then given by

$$\tilde{q}_{\mathrm{A}} = q_{\mathrm{A}}\{q(\mathbf{r}_{a1}) + q(\mathbf{r}_{a2}) + q(\mathbf{r}_{a3}) + q(\mathbf{r}_{a4})\}$$
$$= q_{\mathrm{A}}\{\delta(\mathbf{r}-\mathbf{r}_{s1}) + \delta(\mathbf{r}-\mathbf{r}_{s2}) + \delta(\mathbf{r}-\mathbf{r}_{s3}) + \delta(\mathbf{r}-\mathbf{r}_{s4})\}, \quad (103)$$

and the sound pressure modal amplitude p_i in Eq. (26), be-

cause of the characteristics of the acoustic modal function, yields

$$p_i^{\mathrm{A}} = \frac{j\omega\rho_0 c_0^2}{V_i(\omega_{a,i}^2 - \omega^2)} q_{\mathrm{A}}\{\phi_i(\mathbf{r}_{s1}) + \phi_i(\mathbf{r}_{s2}) + \phi_i(\mathbf{r}_{s3}) + \phi_i(\mathbf{r}_{s4})\} = \frac{4j\omega\rho_0 c_0^2}{V_i(\omega_{a,i}^2 - \omega^2)} q_{\mathrm{A}}\phi_i^{\mathrm{A}}(\mathbf{r}_{s1}). \quad (104)$$

Apparently from Eq. (104), only the even/even acoustic modes belonging to the cluster "A" are excited as a result of applying the transformed sound source $\tilde{q}_{\mathrm{A}}$ in Eq. (103) whereas the other acoustic modes are not.

In the same way, driving the four point sound sources with an appropriate phase arrangement as described in the following permits one to excite the following a acoustic cluster: For cluster B (even/odd acoustic modes),

$$\tilde{q}_{\mathrm{B}} = q_{\mathrm{B}}\{q(\mathbf{r}_{a1}) + q(\mathbf{r}_{a2}) - q(\mathbf{r}_{a3}) - q(\mathbf{r}_{a4})\} = q_{\mathrm{B}}\{\delta(\mathbf{r}-\mathbf{r}_{a1}) + \delta(\mathbf{r}-\mathbf{r}_{a2}) - \delta(\mathbf{r}-\mathbf{r}_{a3}) - \delta(\mathbf{r}-\mathbf{r}_{a4})\}. \quad (105)$$

For cluster C (odd/even acoustic mode),

$$\tilde{q}_{\mathrm{C}} = q_{\mathrm{C}}\{q(\mathbf{r}_{a1}) - q(\mathbf{r}_{a2}) - q(\mathbf{r}_{a3}) + q(\mathbf{r}_{a4})\} = q_{\mathrm{C}}\{\delta(\mathbf{r}-\mathbf{r}_{a1}) - \delta(\mathbf{r}-\mathbf{r}_{a2}) - \delta(\mathbf{r}-\mathbf{r}_{a3}) + \delta(\mathbf{r}-\mathbf{r}_{a4})\}. \quad (106)$$

For cluster D (odd/odd acoustic mode),

$$\tilde{q}_{\mathrm{D}} = q_{\mathrm{D}}\{q(\mathbf{r}_{a1}) - q(\mathbf{r}_{a2}) + q(\mathbf{r}_{a3}) - q(\mathbf{r}_{a4})\} = q_{D}\{\delta(\mathbf{r}-\mathbf{r}_{a1}) - \delta(\mathbf{r}-\mathbf{r}_{a2}) + \delta(\mathbf{r}-\mathbf{r}_{a3}) - \delta(\mathbf{r}-\mathbf{r}_{a4})\}. \quad (107)$$

Thus, combining the transformed sound sources into the cluster sound vector $\tilde{\mathbf{q}}$ leads to

$$\tilde{\mathbf{q}} = \begin{pmatrix} \tilde{q}_{\mathrm{A}} \\ \tilde{q}_{\mathrm{B}} \\ \tilde{q}_{\mathrm{C}} \\ \tilde{q}_{\mathrm{D}} \end{pmatrix} = \mathbf{QTq}, \quad (108)$$

where $\mathbf{Q}$ denotes the gain matrix, and is defined as

$$\mathbf{Q} = \begin{pmatrix} q_{\mathrm{A}} & 0 & 0 & 0 \\ 0 & q_{\mathrm{B}} & 0 & 0 \\ 0 & 0 & q_{\mathrm{C}} & 0 \\ 0 & 0 & 0 & q_{\mathrm{D}} \end{pmatrix}. \quad (109)$$

Here the sound source vector $\mathbf{q}$ is transformed via the Hadamard matrix into the cluster sound source vector $\tilde{\mathbf{q}}$, and hence termed an acoustic cluster actuation.

IV. CLUSTER CONTROL OF ACOUSTIC POTENTIAL ENERGY IN A CLUSTER COUPLED CAVITY

A. Optimal control law for minimizing clustered acoustic potential energy

Now, sound pressure modal amplitude vector in Eq. (34) is partitioned into the four cluster vectors as

$$\mathbf{p} = \begin{pmatrix} \mathbf{p}_{\mathrm{A}} \\ \mathbf{p}_{\mathrm{B}} \\ \mathbf{p}_{\mathrm{C}} \\ \mathbf{p}_{\mathrm{D}} \end{pmatrix}. \quad (110)$$

Substituting Eq. (110) into Eq. (34), the clustered sound pressure modal amplitude vector may then be written as

$$\mathbf{p}_{\mathrm{A}} = \boldsymbol{\Delta}_{\mathrm{A}}^{-1}\tilde{\boldsymbol{\Lambda}}_{a\mathrm{A}}\{q_{\mathrm{A}}\boldsymbol{\Phi}_{\mathrm{A}}\mathbf{t}_{\mathrm{A}} + \mathbf{B}_{\mathrm{A}}\tilde{\boldsymbol{\Lambda}}_{s\mathrm{A}}(f_{\mathrm{A}}\bar{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{t}_{\mathrm{A}} + \hat{\boldsymbol{\Phi}}_{\mathrm{A}}\mathbf{f}_d)\}, \quad (111)$$

$$\mathbf{p}_{\mathrm{B}} = \boldsymbol{\Delta}_{\mathrm{B}}^{-1}\tilde{\boldsymbol{\Lambda}}_{a\mathrm{B}}\{q_{\mathrm{B}}\boldsymbol{\Phi}_{\mathrm{B}}\mathbf{t}_{\mathrm{B}} + \mathbf{B}_{\mathrm{B}}\tilde{\boldsymbol{\Lambda}}_{s\mathrm{B}}(f_{\mathrm{B}}\bar{\boldsymbol{\Phi}}_{\mathrm{B}}\mathbf{t}_{\mathrm{B}} + \hat{\boldsymbol{\Phi}}_{\mathrm{B}}\mathbf{f}_d)\}, \quad (112)$$

$$\mathbf{p}_{\mathrm{C}} = \boldsymbol{\Delta}_{\mathrm{C}}^{-1}\tilde{\boldsymbol{\Lambda}}_{a\mathrm{C}}\{q_{\mathrm{C}}\boldsymbol{\Phi}_{\mathrm{C}}\mathbf{t}_{\mathrm{C}} + \mathbf{B}_{\mathrm{C}}\tilde{\boldsymbol{\Lambda}}_{s\mathrm{B}}(f_{\mathrm{C}}\bar{\boldsymbol{\Phi}}_{\mathrm{C}}\mathbf{t}_{\mathrm{C}} + \hat{\boldsymbol{\Phi}}_{\mathrm{C}}\mathbf{f}_d)\}, \quad (113)$$

$$\mathbf{p}_{\mathrm{D}} = \boldsymbol{\Delta}_{\mathrm{D}}^{-1}\tilde{\boldsymbol{\Lambda}}_{a\mathrm{D}}\{q_{\mathrm{D}}\boldsymbol{\Phi}_{\mathrm{D}}\mathbf{t}_{\mathrm{D}} + \mathbf{B}_{\mathrm{D}}\tilde{\boldsymbol{\Lambda}}_{s\mathrm{D}}(f_{\mathrm{D}}\bar{\boldsymbol{\Phi}}_{\mathrm{D}}\mathbf{t}_{\mathrm{D}} + \hat{\boldsymbol{\Phi}}_{\mathrm{D}}\mathbf{f}_d)\} \quad (114)$$

where $\boldsymbol{\Phi}_{\mathrm{A}}$–$\boldsymbol{\Phi}_{\mathrm{D}}$, $\bar{\boldsymbol{\Phi}}_{\mathrm{A}}$–$\bar{\boldsymbol{\Phi}}_{\mathrm{D}}$, and $\hat{\boldsymbol{\Phi}}_{\mathrm{A}}$–$\hat{\boldsymbol{\Phi}}_{\mathrm{D}}$ are the partitioned sub-matrices of $\boldsymbol{\Phi}$ in Eq. (32), $\mathbf{Z}_c$ in Eq. (20), and $\mathbf{Z}_d$ in Eq. (21), respectively, and hence

$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_{\mathrm{A}} \\ \boldsymbol{\Phi}_{\mathrm{B}} \\ \boldsymbol{\Phi}_{\mathrm{C}} \\ \boldsymbol{\Phi}_{\mathrm{D}} \end{pmatrix}, \quad (115)$$

$$\mathbf{Z}_c = \begin{pmatrix} \bar{\boldsymbol{\Phi}}_{\mathrm{A}} \\ \bar{\boldsymbol{\Phi}}_{\mathrm{B}} \\ \bar{\boldsymbol{\Phi}}_{\mathrm{C}} \\ \bar{\boldsymbol{\Phi}}_{\mathrm{D}} \end{pmatrix}, \quad (116)$$

$$\mathbf{Z}_d = \begin{pmatrix} \hat{\boldsymbol{\Phi}}_{\mathrm{A}} \\ \hat{\boldsymbol{\Phi}}_{\mathrm{B}} \\ \hat{\boldsymbol{\Phi}}_{\mathrm{C}} \\ \hat{\boldsymbol{\Phi}}_{\mathrm{D}} \end{pmatrix}. \quad (117)$$

As such, the clustered acoustic potential energy is expressed as

$$\hat{E}_{\mathrm{A}} = \frac{V}{4\rho_0 c_0^2}|\tilde{p}_{\mathrm{A}}|^2 = \frac{4V}{\rho_0 c_0^2}\mathbf{p}_{\mathrm{A}}^H\boldsymbol{\Psi}_{\mathrm{A}}(\mathbf{r}_{s1})\mathbf{p}_{\mathrm{A}}, \quad (118)$$

$$\hat{E}_{\mathrm{B}} = \frac{V}{4\rho_0 c_0^2}|\tilde{p}_{\mathrm{B}}|^2 = \frac{4V}{\rho_0 c_0^2}\mathbf{p}_{\mathrm{B}}^H\boldsymbol{\Psi}_{\mathrm{B}}(\mathbf{r}_{s1})\mathbf{p}_{\mathrm{B}}, \quad (119)$$

$$\hat{E}_{\mathrm{C}} = \frac{V}{4\rho_0 c_0^2}|\tilde{p}_{\mathrm{C}}|^2 = \frac{4V}{\rho_0 c_0^2}\mathbf{p}_{\mathrm{C}}^H\boldsymbol{\Psi}_{\mathrm{C}}(\mathbf{r}_{s1})\mathbf{p}_{\mathrm{C}}, \quad (120)$$

$$\hat{E}_{\mathrm{D}} = \frac{V}{4\rho_0 c_0^2}|\tilde{p}_{\mathrm{D}}|^2 = \frac{4V}{\rho_0 c_0^2}\mathbf{p}_{\mathrm{D}}^H\boldsymbol{\Psi}_{\mathrm{D}}(\mathbf{r}_{s1})\mathbf{p}_{\mathrm{D}}, \quad (121)$$

where

$$\boldsymbol{\Psi}_i(\mathbf{r}_{s1}) = \boldsymbol{\phi}_i(\mathbf{r}_{s1})\boldsymbol{\phi}_i^T(\mathbf{r}_{s1}) \quad (i = \mathrm{A,B,C,D}). \quad (122)$$

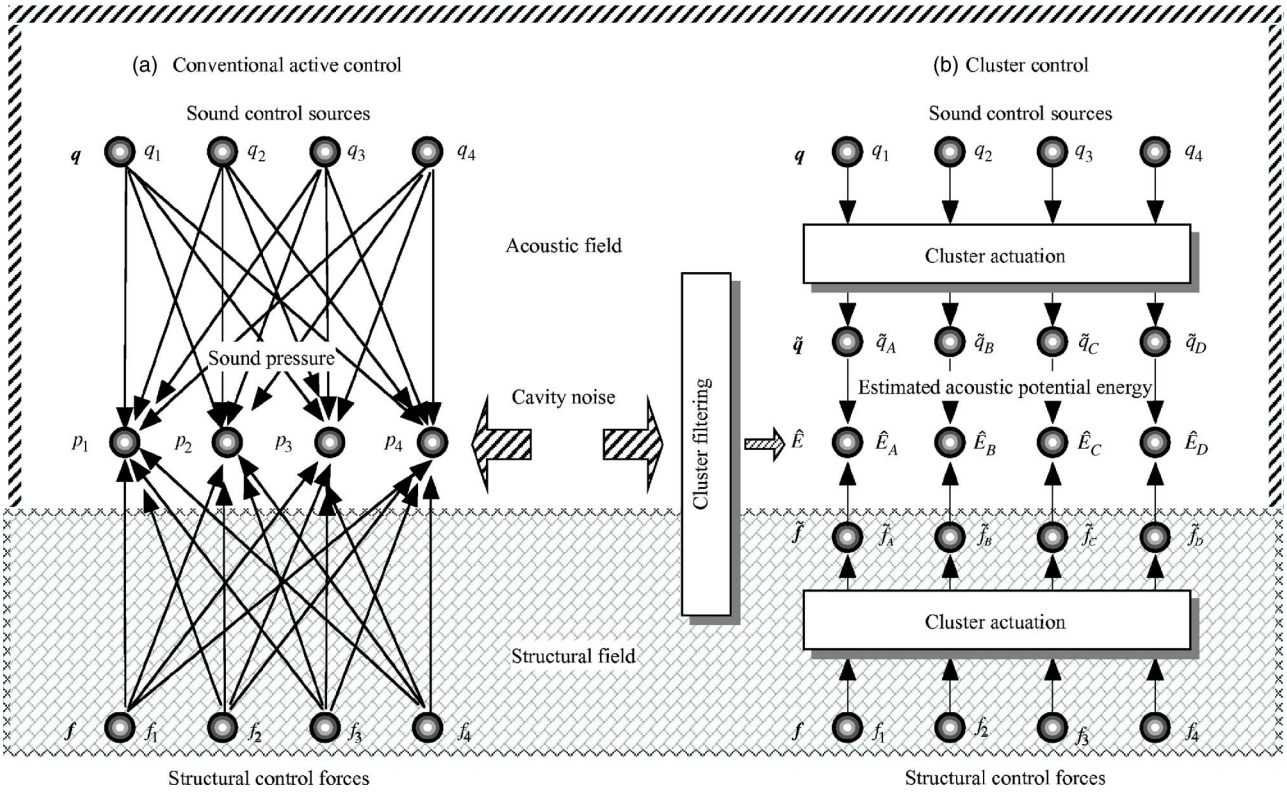


FIG. 2. Schematic signal flow diagram of (a) a conventional active control and (b) cluster control with cluster filtering and cluster actuation.

By substituting Eqs. (111)–(114) into Eqs. (118)–(121), the potential energy further expands to

$$\hat{E}_i = a_i q_i^2 + 2b_i q_i f_i + 2c_i q_i + d_i f_i^2 + 2e_i f_i + f_0 \quad (i = \mathrm{A,B,C,D}), \tag{123}$$

where

$$a_i = \mathbf{t}_i^T \boldsymbol{\Phi}_i^T (\boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai})^H \boldsymbol{\Psi}_i \boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai} \boldsymbol{\Phi}_i \mathbf{t}_i, \tag{124}$$

$$b_i = \mathbf{t}_i^T \boldsymbol{\Phi}_i^T (\boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai})^H \boldsymbol{\Psi}_i \boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai} \bar{\boldsymbol{\Phi}}_i \mathbf{t}_i, \tag{125}$$

$$c_i = \mathbf{t}_i^T \boldsymbol{\Phi}_i^T (\boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai})^H \boldsymbol{\Psi}_i \boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai} \hat{\boldsymbol{\Phi}}_i \mathbf{f}_d, \tag{126}$$

$$d_i = \mathbf{t}_i^T \bar{\boldsymbol{\Phi}}_i^T (\boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai})^H \boldsymbol{\Psi}_i \boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai} \bar{\boldsymbol{\Phi}}_i \mathbf{t}_i, \tag{127}$$

$$e_i = \mathbf{t}_i^T \bar{\boldsymbol{\Phi}}_i^T (\boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai})^H \boldsymbol{\Psi}_i \boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai} \hat{\boldsymbol{\Phi}}_i \mathbf{f}_d, \tag{128}$$

$$f_0 = \mathbf{f}_d^H \hat{\boldsymbol{\Phi}}_i^T (\boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai})^H \boldsymbol{\Psi}_i \boldsymbol{\Delta}_i^{-1} \tilde{\boldsymbol{\Lambda}}_{ai} \hat{\boldsymbol{\Phi}}_i \mathbf{f}_d. \tag{129}$$

Taking the derivative of Eq. (123) with respect to q_i and f_i, respectively, and setting them to zero, we have

$$\frac{\partial \hat{E}_i}{\partial q_i} = 2a_i q_i + 2b_i f_i + 2c_i = 0, \tag{130}$$

$$\frac{\partial \hat{E}_i}{\partial f_i} = 2b_i q_i + 2d_i f_i + 2e_i = 0. \tag{131}$$

Solving Eqs. (130) and (131) simultaneously, the optimal control law for minimizing the clustered acoustic potential energy is given by

$$q_i = -\frac{c_i d_i - b_i e_i}{a_i d_i - b_i^2}, \tag{132}$$

$$f_i = -\frac{a_i e - b_i c_i}{a_i d_i - b_i^2}. \tag{133}$$

B. Cluster control system arrangement

With a view to achieving global noise suppression, it is required to increase the number of sensors and actuators. 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. Figure 2(a) illustrates the schematic diagram of a conventional active control system with eight inputs (four sound control sources and four structural control forces) and four outputs (four sound pressures). As seen from the figure, control signals intersect each other, a complicated control system being constructed. On the contrary, as shown in Fig. 2(b), cluster control enables the independent control of each cluster without crosstalk, thereby substantially reducing computational burden. More importantly, the cluster filtering outputs are no longer the sound pressure *per se* at sensor locations but the

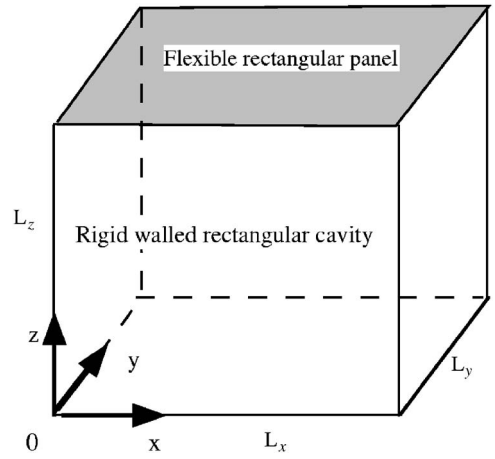


FIG. 3. A rectangular cavity with a top simply supported panel and rigid walls.

set of all acoustic modes with the same attributes of each cluster, which is responsible for the total acoustic potential energy. Thus, the suppression of the cluster filtering outputs via cluster actuation leads to the suppression of the global noise attenuation. Furthermore, an eight input four output control system is regarded as 8 SISO (single input single output) systems in Fig. 2(b) as opposed to MIMO (multi-input multi-output) in Fig. 2(a), thus reducing significantly the computation load. Even in the worst case when the cluster control fails to be implemented, the cluster control system turns into a conventional MIMO system.

V. NUMERICAL SIMULATION

In an effort to investigate the fundamental characteristics of cluster coupling as well as the cluster control effect for suppressing the acoustic potential energy in a cluster coupled structural/acoustic enclosure, an orthorhombic enclosure, as shown in Fig. 3, with the dimensions of 0.917 m $\times$0.707 m$\times$0.6 m is employed. Among the six panels encompassing the enclosure, the top panel is assumed to be flexible whereas the other five are perfectly rigid. All four edges of the top steel panel of thickness h are simply supported, the panel being subjected to a point disturbance force of 1 N at $\mathbf{r}_d=(0.2,0.2,0.6\ \text{m})$. Four cluster filtering sensors for estimating acoustic potential energy are placed with the origin at $\mathbf{r}_{s1}=(0.1,0.1,0\ \text{m})$, four cluster control actuators for AVC actuators and four cluster sound sources for ANC are located with the base at $\mathbf{r}_{c1}=(0.1,0.1,0.6\ \text{m})$ and $\mathbf{r}_{a1}=(0.15\ \text{m},0.15\ \text{m},0)$, respectively. Table I lists the structural and acoustic modes grouped into the four clusters A, B, C, and D in the frequency range up to 200 Hz.

Consider the cluster filtering for extracting the designated acoustic potential energy into an appropriate cluster. Figure 4(a) shows the frequency characteristics of acoustic potential energy E_p of the cavity with the thickness of the panel $h=1$ mm and estimated acoustic potential energy $\hat{E}_\text{A}$. Aside from approximately 20 dB discrepancy between the two curves, these dynamical properties look significantly similar, implying the validity of the cluster filtering for estimating the acoustic potential energy in a cluster coupled cavity using four point sound sensors as will be discussed shortly. Illustrated in Fig. 4(b) are the cluster filtering outputs for cluster from A through D, each property being depicted separately in Fig. 4(c). It is interesting to note that the degenerate resonance, for instance, at 22.1 Hz, consisting of the (1,1) and (2,1) modes, is successfully filtered out in each cluster. Thus, as a result of applying cluster filtering, four clusters are extracted perfectly without causing observation spillover.

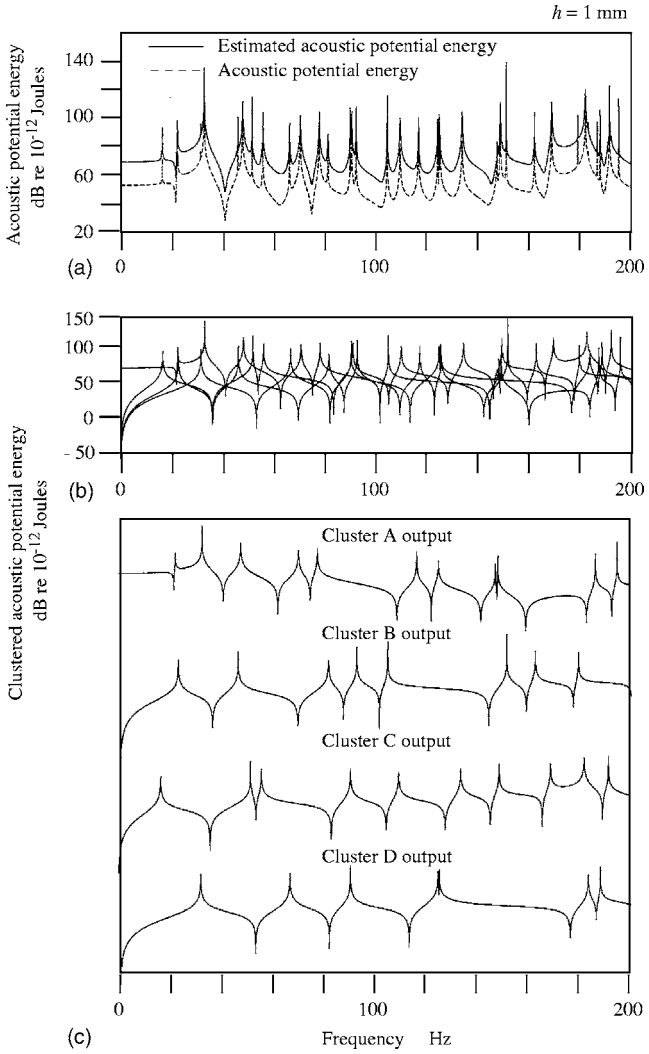


FIG. 4. Acoustic potential energy in a cluster coupled cavity with the panel of $h=1$ mm; (a) acoustic potential energy E_p and estimated acoustic potential energy $\hat{E}_p$; (b) cluster filtering outputs; (c) components of each cluster output.

Figure 5 depicts the cluster control effect for suppressing the acoustic potential energy belonging to cluster A in the cavity with the panel thickness $h=1$ mm using both AVC and ANC. The control laws derived in Eqs. (132) and (133) are designed so as to minimize the estimated acoustic potential energy $\hat{E}_\text{A}$ consisting of all odd/odd structural modes and even/even acoustic modes in the cavity. Figures 5(b) and 5(c) show the estimated acoustic potential energy $\hat{E}_p$ and the estimated acoustic potential energy reduction achieved by minimizing the contribution of $\hat{E}_\text{A}$ to total acoustic potential

TABLE I. Clustered structural and acoustic modes and their modal indices of a weekly coupled rectangular cavity with a simply supported top panel of dimensions of 0.917 m $\times$0.707 m $\times$0.6 m (h=1 mm).

Cluster A		Cluster B		Cluster C		Cluster D	
Modal indices	Frequency (Hz)	Modal indices	Frequency (Hz)	Modal indices	Frequency (Hz)	Modal indices	Frequency (Hz)
0 0 0	0	1 2	22.511	2 1	16.55	2 2	31.249
1 1	7.8121	3 2	45.811	4 1	51.5	4 2	66.198
3 1	31.112	1 4	81.307	2 3	55.747	2 4	90.044
1 3	47.009	5 2	92.411	4 3	90.697	6 2	124.45
3 3	70.309	3 4	104.61	6 1	109.75	4 4	124.99
5 1	77.712	5 4	151.21	2 5	134.14	6 4	183.24
5 3	116.91	7 2	162.31	6 3	148.95	2 6	188.04
1 5	125.4	1 6	179.3	4 5	169.09		
7 1	147.61			1 0 0	180.73		
3 5	148.7			8 1	191.3		
7 3	186.81						
5 5	195.3						

energy in the cavity. As seen from Fig. 5(c), only the resonant modes in terms of cluster A are suppressed whereas the other modes remain unaffected.

Figure 5(a) illustrates the plot of E_p obtained when minimizing the estimated potential energy $\hat{E}_A$. Comparing the curve with that of the estimated potential energy $\hat{E}_p$ in Fig. 5(b), there seems little discrepancy between both of them, implying that the control law in terms of $\hat{E}_p$ is valid even for suppressing E_p. The reason for this is clear if the frequency properties of E_p and $\hat{E}_p$ in Fig. 4(a) are considered; the curve of $\hat{E}_p$ roughly overlaps that of E_p as a result of raising the base line of E_p, and thus E_p is expressed approximately by multiplying $\hat{E}_p$ with some constant value. If this is the case, the derivative of $\hat{E}_p$ with respect to control forces becomes the same as that of E_p, and hence the control law for minimizing $\hat{E}_p$ is commensurate with that for minimizing E_p. Fig-

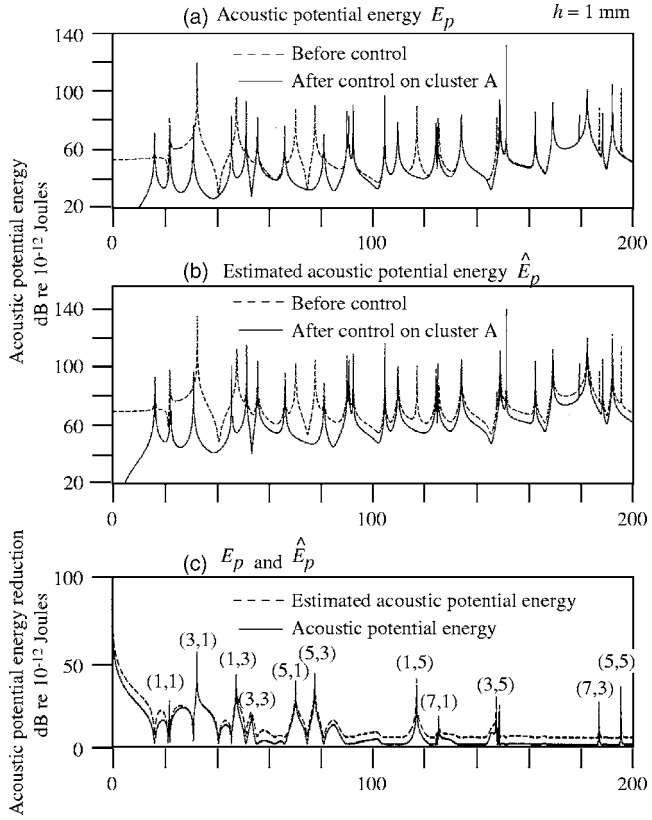


FIG. 5. Cluster control effect for minimizing cluster A of a cluster coupled cavity with the panel of h=1 mm; (a) E_p before and after cluster control on cluster A; (b) $\hat{E}_p$ before and after cluster control on cluster A; (c) reduction in E_p and $\hat{E}_p$.

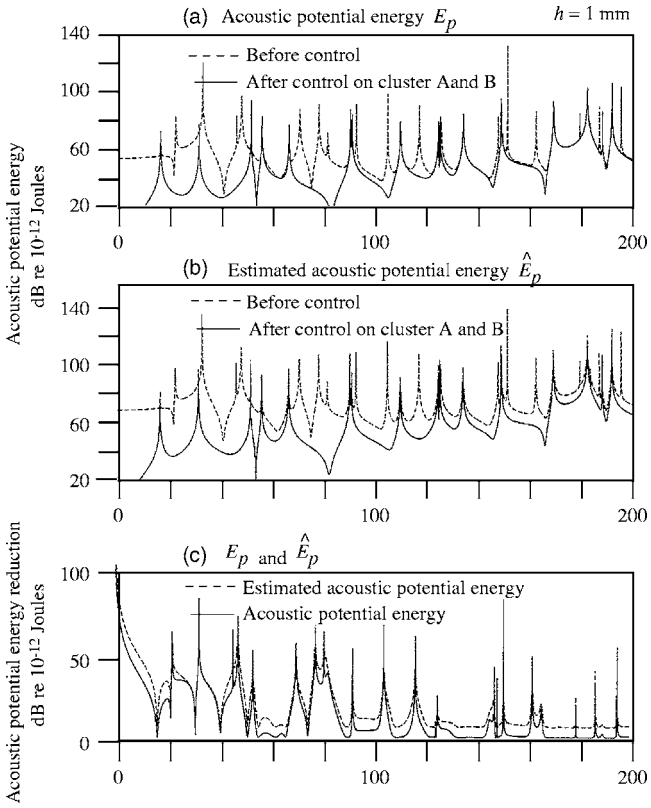


FIG. 6. Cluster control effect for minimizing clusters A and B of a cluster coupled cavity with the panel of h=1 mm; (a) E_p before and after cluster control on clusters A and B; (b) $\hat{E}_p$ before and after cluster control on clusters A and B; (c) reduction in E_p and $\hat{E}_p$.

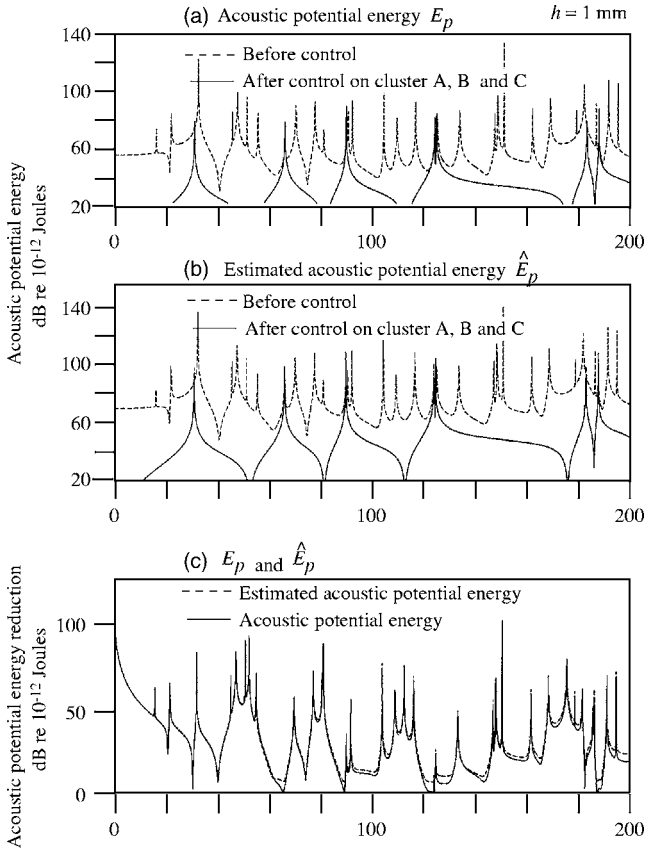


FIG. 7. Cluster control effect for minimizing clusters A, B, and C of a cluster coupled cavity with the panel of $h=1$ mm; (a) E_p before and after cluster control on clusters A, B, and C; (b) $\hat{E}_p$ before and after cluster control on clusters A, B, and C; (c) reduction in E_p and $\hat{E}_p$.

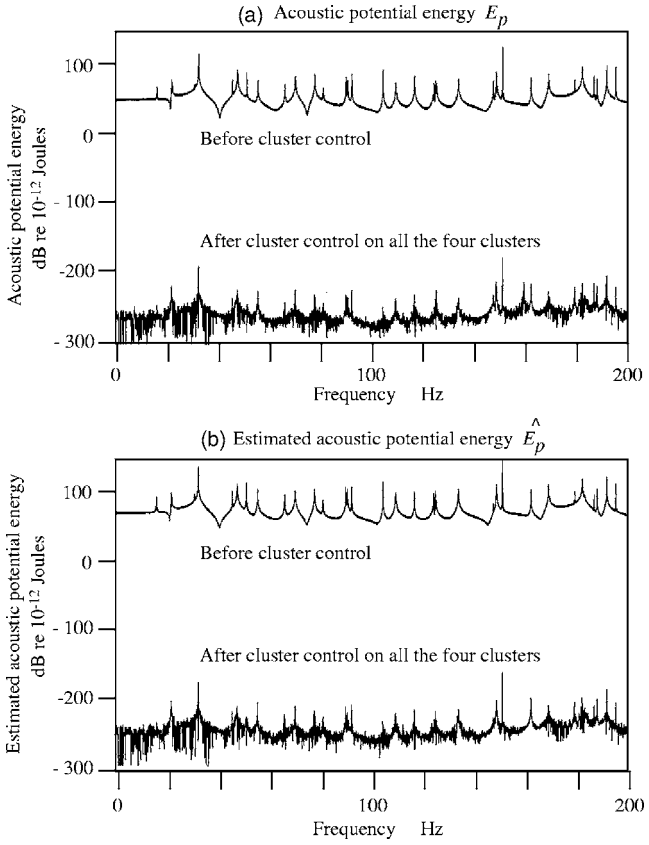


FIG. 8. Cluster control effect for minimizing clusters A, B, C, and D of a cluster coupled cavity with the panel of $h=1$ mm; (a) E_p before and after cluster control on all the four clusters; (b) $\hat{E}_p$ before and after cluster control on all the four clusters.

ure 5(c) shows the acoustic potential energy reduction in E_p and $\hat{E}_p$, demonstrating a small amount of inconsistency between the two because of $\hat{E}_p$ still being the approximation of acoustic potential energy. To evaluate quantitatively the control effect for suppressing the acoustic potential energy, an overall level based upon acoustic potential energy reduction in Fig. 5(c) is introduced, the results being as follows: 9.6 and 12.9 dB for the reduction of E_p and $\hat{E}_p$, respectively, via cluster control on the cluster A.

Depicted in Figs. 6(a) and 6(b) are, respectively, the curves of E_p and $\hat{E}_p$ obtained when minimizing both $\hat{E}_\mathrm{A}$ and $\hat{E}_\mathrm{B}$ in Eqs. (118) and (119). Comparing the plots with those in Fig. 5, it is found that the number of resonant peaks decreases since the attribute of E_p after the cluster control is governed only by that of clusters C and D. In contrast, acoustic potential energy reduction in Fig. 6(c) consists of the suppressed clusters, A and B, with the result that an overall level indicating the control effect further improves; 23.4 dB for the reduction of E_p and 28.6 dB in $\hat{E}_p$.

In addition to the performance index in the above, $\hat{E}_C$ is further included, hence the minimization of $\hat{E}_\mathrm{A}$, $\hat{E}_\mathrm{B}$, and $\hat{E}_\mathrm{C}$. Illustrated in Figs. 7(a) and 7(b) are the acoustic potential energy before and after the cluster control. The acoustic potential energy E_p in this case is dominated by cluster D only, so that the resonant peaks appearing in the figures refer to those of cluster D, whereas the number of resonant peaks in the acoustic potential energy reduction in Fig. 7(c) further increases, resulting in enhancement of control effect. Observe that the curve in Fig. 7(a) coincides with that of the bottom picture in Fig. 4(c) that shows the frequency characteristics of cluster D extracted by cluster filtering. Therefore, the acoustic potential energy in the cluster coupled cavity after control is now dictated by the attribute of cluster D. An overall value quantitatively indicating control effect for suppressing acoustic potential energy is found to be 41.4 dB in E_p and 44.8 dB in $\hat{E}_p$.

Note that acoustic potential energy may be grouped into four clusters, each cluster possessing the same properties in common, and hence the suppression of all four clusters leads to the suppression of total acoustic potential energy in the cavity. Figure 8 shows the control effect obtained by suppressing all the estimated acoustic potential energy; $\hat{E}_\mathrm{A}$, $\hat{E}_\mathrm{B}$, $\hat{E}_\mathrm{C}$, and $\hat{E}_\mathrm{D}$ using both AVC and ANC. Figures 8(a) and 8(b) show, respectively, the frequency properties of E_p and $\hat{E}_p$ before and after the cluster control. As is seen from the figures, control effect for suppressing the acoustic potential energy in the cluster coupled cavity is outstanding. It may be of little significance to depict the frequency characteristics, the gain of which is under −200 dB; however it is still helpful to grasp the control effect.

VI. CONCLUSIONS

For the purpose of minimizing acoustic potential energy in a cluster coupled structural/acoustic cavity, cluster control comprising both cluster filtering and cluster actuation has been presented. Acoustic potential energy in a cluster coupled cavity was found to be expressed as a resultant from each cluster contribution, and hence cluster potential energy may be controlled independently without causing control spillover. This fact enables one to implement a multi-SISO system whereby computation load may significantly be alleviated. More important, the cluster filtering outputs are no longer the sound pressure *per se* at sensor locations but the set of all acoustic modes with the same attributes of each cluster, which is responsible for the total acoustic potential energy. Thus, the suppression of the cluster filtering outputs via cluster actuation leads to the suppression of the global noise attenuation. Even in the worst case when the cluster control fails to be implemented, the cluster control system turns into a conventional MIMO system. Furthermore, cluster control law to suppress clustered potential energy using ANC and AVC has also been derived. Based upon the cluster control law, a numerical simulation was carried out, demonstrating the validity of cluster control for suppressing acoustic potential energy in a cluster coupled cavity.

ACKNOWLEDGMENTS

Support for the work from the Nissan Motor Corporation, and Tokyo Metropolitan Institute of Technology is gratefully acknowledged.

[1]See, for instance, P. A. Nelson, A. R. D. Curtis, S. J. Elliot, and A. J. Bullmore, "The minimum power output of free field point sources and the active control of sound," J. Sound Vib. **116**, 397–414 (1987).
[2]See, for instance, S. D. Snyder, N. Tanaka, and Y. Kikushima, "The use of optimally shaped piezo-electric film sensors in the active control of free field structural radiation. 1. Feedforward control," Trans. ASME, J. Vib. Acoust. **117**, 311–322 (1995).
[3]See, for instance, S. D. Snyder, N. Tanaka, and Y. Kikushima, "The use of optimally shaped piezo-electric film sensors in the active control of free field structural radiation. 2. Feedback control," Trans. ASME, J. Vib. Acoust. **118**, 112–121 (1996).
[4]J. Pan and D. A. Bies, "The effect of fluid-structural coupling on sound waves in an enclosure-theoretical part," J. Acoust. Soc. Am. **87**, 691–707 (1990).
[5]L. D. Pope, "On the transmission of sound through finite closed shells: Statistical energy analysis, modal coupling and non-resonant transmission," J. Acoust. Soc. Am. **50**, 1004–1018 (1971).
[6]J. Pan, C. H. Hansen, and D. A. Bies, "Active control of noise transmission through a panel into a cavity. I. Analytical study," J. Acoust. Soc. Am. **87**, 2098–2108 (1990).
[7]J. Pan and C. H. Hansen, "Active control of noise transmission through a panel into a cavity: Experimental study," J. Acoust. Soc. Am. **90**, 1488–1492 (1991).
[8]S. D. Snyder and C. H. Hansen, "The design of systems to actively control periodic sound transmission into enclosed spaces 1. Analytical models," J. Sound Vib. **170**, 433–449 (1994).
[9]S. D. Snyder and C. H. Hansen, "The design of systems to actively control periodic sound transmission into enclosed spaces. 2. Mechanisms and trends," J. Sound Vib. **170**, 451–472 (1994).
[10]S. D. Snyder and N. Tanaka, "On feedforward active control of sound and vibration using vibration error signals," J. Acoust. Soc. Am. **94**, 2181–2193 (1993).
[11]A. Sampath and B. Balachandran, "Studies on performance functions of interior noise control," Smart Mater. Struct. **6**, 315–332 (1997).
[12]A. Sampath and B. Balachandran, "Active control of multiple tones in an enclosure," J. Acoust. Soc. Am. **106**, 211–225 (1999).
[13]S. M. Kim and M. J. Brennan, "Active control of harmonic sound transmission into an acoustic enclosure using both structural and acoustic actuators," J. Acoust. Soc. Am. **107**, 2523–2534 (2000).
[14]R. H. Lyon, "Noise reduction of rectangular enclosures with one flexible wall," J. Acoust. Soc. Am. **35**, 1791–1797 (1963).
[15]E. H. Dowell and H. M. Voss, "The effect of a cavity on panel vibration," AIAA J. **1**, 476–477 (1963).
[16]A. J. Pretlove, "Free vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **2**, 197–209 (1965).
[17]A. J. Pretlove, "Forced vibrations of a rectangular panel backed by a closed rectangular cavity," J. Sound Vib. **3**, 252–261 (1966).
[18]T. Kihlman, "Sound radiation into a rectangular room. Application to airborne sound transmission in buildings," Acustica **18**, 11–20 (1967).
[19]M. C. Bhattacharya and M. J. Crocker, "Forced vibration of a panel and radiation of sound into a room," Acustica **22**, 275–294 (1969/1970).
[20]R. W. Guy and M. C. Bhattacharya, "The transmission of sound through a cavity-backed finite plate," J. Sound Vib. **27**, 207–223 (1973).
[21]R. W. Guy, "The steady state transmission of sound at normal and oblique incidence through a thin panel backed by a rectangular room—a multimodal analysis," Acustica **43**, 295–304 (1979).
[22]N. Tanaka and S. D. Snyder, "Cluster control of a distributed-parameter planar structure—Middle authority control," J. Acoust. Soc. Am. **112**, 2798–2807 (2002).
[23]N. Tanaka, R. Fukuda, and C. H. Hansen, "Acoustic cluster control of noise radiated from a planar structure," J. Acoust. Soc. Am. **117**, 3686–3694 (2005).
[24]M. J. Balas, "Active control of flexible systems," J. Optim. Theory Appl. **25**, 415–436 (1978).

Aircraft noise effects on sleep: Application of the results of a large polysomnographic field study[a)]

Mathias Basner[b)] and Alexander Samel
German Aerospace Center (DLR), Institute of Aerospace Medicine, 51170 Cologne, Germany

Ullrich Isermann
German Aerospace Center (DLR), Institute of Aerodynamics and Flow Technology, Bunsenstr. 10, 37073 Goettingen, Germany

(Received 3 November 2005; accepted 11 February 2006)

The Institute of Aerospace Medicine at the German Aerospace Center (DLR) investigated the influence of nocturnal aircraft noise on sleep in polysomnographic laboratory and field studies between 1999 and 2004. The results of the field studies were used by the Regional Council of Leipzig (Germany) for the establishment of a noise protection plan in the official approval process for the expansion of Leipzig/Halle airport. Methods and results of the DLR field study are described in detail. Special attention is given to the dose-response relationship between the maximum sound pressure level of an aircraft noise event and the probability to wake up, which was used to establish noise protection zones directly related to the effects of noise on sleep. These protection zones differ qualitatively and quantitatively from zones that are solely based on acoustical criteria. The noise protection plan for Leipzig/Halle airport is presented and substantiated: (1) on average, there should be less than one additional awakening induced by aircraft noise, (2) awakenings recalled in the morning should be avoided as much as possible, and (3) aircraft noise should interfere as little as possible with the process of falling asleep again. Issues concerned with the representativeness of the study sample are discussed. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184247]

PACS number(s): 43.50.Qp, 43.50.Sr, 43.50.Rq, 43.50.Lj [BSF] Pages: 2772–2784

I. INTRODUCTION

Between 1999 and 2004, the DLR–Institute of Aerospace Medicine (IAM) in Cologne, Germany, performed extensive laboratory and field studies on the effects of aircraft noise on sleep, mood, and performance in the DLR/HGF project "Quiet Air Traffic."[2–4] The Regional Council of Leipzig (RCL) asked the IAM to propose a concept for the protection of airport residents against the adverse effects of nocturnal aircraft noise on sleep based on the findings of these studies.

Leipzig/Halle airport is planned to be extended to an international freight hub with air traffic predominantly occurring during the night. In order to be able to handle the prognosticated traffic volumes, the southern runway will be turned and extended to a length of 3600 m. Together with the northern runway, this independent parallel runway system will allow for simultaneous takeoffs and landings on both runways. The traffic volume is predicted with 81 000 aircraft movements during the six busiest months in the year 2015.[1] Of these, 45 600 will take place during the day between 6:00 and 22:00 and 35 400 will occur during the night between 22:00 and 6:00. Thus, a large part of the aircraft movements will take place during the night. This situation distinguishes Leipzig/Halle airport from most other airports worldwide.

After an extensive course of consideration the RCL decided to develop a plan for the protection of airport residents against the adverse effects of nocturnal aircraft noise primarily based on the results of the DLR field study,[2] and therefore on the newest available scientific data. On 4 November 2004 the noise protection plan was presented for approval.[1] A few days later, DHL decided to move its international cargo hub from Brussels to Leipzig.

In this publication, methodological aspects of the DLR field study will be reported. The most important findings on the effects of nocturnal aircraft noise on sleep in general and on the probability of noise-induced awakenings in particular will be presented. Based on these results, it will be shown that noise protection plans based on number above threshold (NAT) and/or L_{eq} criteria are not suitable for an adequate description of the effects of nocturnal aircraft noise on sleep. Finally, a noise protection plan based on the findings of the DLR field study will be presented and substantiated.

II. METHODS

The DLR–Institute of Aerospace Medicine investigated the influence of nocturnal aircraft noise on human sleep, mood, and performance in a laboratory and a field study. The concepts of the noise protection plan for Leipzig/Halle airport are mainly based on the results of the field study. Therefore, study design and methods used in the field study will be briefly described. For a detailed description the reader is

[a)]Portions of this work were published in M. Basner, U. Isermann, and A. Samel, "Die Ergebnisse der DLR-Studie und ihre Umsetzung in einer lärmmedizinischen Beurteilung für ein Nachtschutzkonzept (The application of the DLR-study for a medical evaluation of a protective concept on adverse effects of nocturnal aircraft noise)," Z. Lärmbekämpfung **52**, 109–123 (2005).

[b)]Electronic mail: mathias.basner@dlr.de

 0001-4966/2006/119(5)/2772/13/$22.50

asked to refer to the executive summary of the study.[2] The field study was conducted between September 2001 and November 2002 with 64 residents of Cologne-Bonn airport, which is one of the German airports with the highest nighttime traffic densities and mainly used for freight traffic during the night. Subjects were investigated for nine consecutive nights, starting on Mondays.

Participants, selected in a multi-level process, were between 19 and 61 years old (average: 38 years). Fifty-six percent of the participants were female. Subjects had to be free of intrinsic sleep disorders and had to have normal hearing thresholds according to age. A detailed description of the selection process can be found in the report DLR-FB 2004/7E.[2] Consequences of the selection process for the representativeness of the sample are discussed in detail in Sec. V F. The study protocol was approved by the ethics committee of the Medical Board of the district North Rhine. Subjects were instructed according to the Helsinki declaration, participated voluntarily, and were free to discontinue their participation at any time without explanation.

The electroencephalogram (brain current diagram, EEG), the electrooculogram (eye movements, EOG), the electromyogram (muscle tension, EMG), the electrocardiogram (ECG), respiratory movements, finger pulse amplitude, position in bed, and actigraphy were sampled continuously during the night. With the EEG, EOG, and EMG signals (also called polysomnography), sleep can be classified into different sleep stages.[5]

Wake is differentiated from sleep. Sleep itself is classified in REM sleep, with its typical rapid eye movements, and nonREM sleep. NonREM sleep can be further divided in the four sleep stages S1, S2, S3, and S4. Because of high arousal thresholds,[6] stages S3 and S4 are also called "deep sleep." Deep sleep as well as REM sleep are known to be very important for the restorative power of sleep.[7] Wake and stage S1, on the other hand, do not seem to contribute to recuperation, or only very little.[8]

Historically, each night is usually divided into 30-s epochs. A trained scorer then assigns one of the sleep stages or "awake" to each of the epochs. Since reliable procedures for an automatic sleep stage analysis do not yet exist, sampling and analysis of polysomnographic data are sumptuous, and therefore have only been applied in studies with relatively small sample sizes (see Sec. V F). However, only polysomnography allows the assessment of structural aspects of sleep. Studies using actigraphy try to draw conclusions on sleep quality and quantity based on movements of the wrist of one arm, and are therefore obviously inferior to polysomnography according to the informational value. With 64 subjects and 576 subject nights, the DLR study is the largest polysomnographic field study on the effects of nocturnal aircraft noise so far.

In the field study, sound pressure levels (SPL) and the actual sounds were recorded inside the bedroom (at the sleeper's ear) and outside (2 m in front of the window) with class-1 sound level meters. All events (e.g., aircraft noise, road traffic noise, snoring, etc.) were identified by a human scorer. The beginning and the end of each event were marked. The simultaneous recording of acoustical and electrophysiological signals allowed for an event-correlated analysis with a maximum resolution of 125 ms.

Aircraft noise is intermittent noise. An event-correlated analysis establishes a direct temporal association between the occurrence of an aircraft noise event (ANE) and the reaction of the investigated subject to the ANE. This is only possible because of the synchronous sampling of electrophysiological and acoustical signals. Variables like the nocturnal secretion rate of stress hormones, the annoyance of study subjects asked for in the morning by questionnaires, or the amounts of the different sleep stages are represented in a single datum, which summarizes the effects of all nocturnal ANEs. These integrative measures are unsuitable for an event-correlated analysis, because the connection to single noise events cannot be made.

The reactions of sleeping humans to aircraft noise are nonspecific, since they may also be observed during natural sleep otherwise undisturbed by external stimuli. Hence, reactions observed during an ANE cannot be differentiated from spontaneous reactions according to electrophysiological criteria. Therefore, it is necessary to use an event-correlated analysis to distinguish spontaneous reactions from reactions observed during an ANE. Furthermore, spontaneous reactions occur irregularly. Therefore, if there is a reaction during an ANE, it is important to ask how often this reaction would have taken place spontaneously anyway, i.e., without the influence of aircraft noise. In epidemiology the term *attributable risk* is often used in this context. The probability of a reaction induced by aircraft noise is calculated as

$$P_{\text{induced}} = P_{\text{ANE}} - P_{\text{spontaneous}}. \tag{1}$$

As the physiological reactions may not immediately start after the beginning of an ANE, a certain time interval is screened for reactions of the sleeper. This time interval is called a "noise window." With a size of three epochs (90 s) after the beginning of an ANE, the length of the noise window was chosen to maximize the probability of reactions induced by aircraft noise [P_{induced} in Eq. (1)].

Several potential indicators for noise-induced sleep disturbances have been identified and proposed in the past. Brief EEG and EMG activations are called **arousals**.[9] Because of their short duration they are not classified as stage "awake" according to the rules of Rechtschaffen and Kales.[5] **Awakenings** are longer arousals, defined as EEG and EMG activations that last for at least 15 s and therefore lead to a classification of the sleep stage as "awake." **Sleep stage changes** are defined as transitions from one sleep stage to a different sleep stage. In the context of noise effects research, commonly only those sleep stage changes leading to a lighter sleep are considered, e.g., changes from deep sleep stage S4 to the light sleep stage S2.

Polysomnographic studies conducted in the past predominantly used awakenings as the primary indicator of sleep disturbances induced by environmental noise.[10,11] Because of the following reasons awakenings are appropriate indicators for sleep disturbances induced by environmental noise:

(i) The awakening is the **strongest form of activation**

of the sleeping organism. The consequences for the restorative functions of sleep are accordingly severe.

(ii) Awakenings are relatively **specific**, i.e., the frequency of spontaneous awakenings is relatively low compared to other indicators. In the 112 baseline nights of the experimental group in the laboratory study, on average about 24 spontaneous awakenings were observed.[4,2] Spontaneous sleep stage changes were seen more than twice as often (on average about 52 changes per night). Mathur and Douglas[12] investigated the spontaneous onset of EEG arousals according to ASDA criteria.[9] They found on average about 21 arousals per hour of sleep. If the mean sleep period time (SPT) of 411.5 min of the noise-free baseline nights of the laboratory study is taken as a basis, this value corresponds to about 144 spontaneous EEG arousals per night.

(iii) In contrast to arousal, awakenings are usually accompanied by prolonged and unimodal increases in heart frequency.[13] We observed in our own investigations that the amplitude and/or the frequency of heart frequency accelerations are relatively low if there is no simultaneous awakening. But especially the regular occurrence of these nocturnal **vegetative reactions** seems to be a possible cause for the development of high blood pressure and the associated diseases of the cardiovascular system (myocardial infarction, stroke).[14] The degree of vegetative reactions accompanied by sleep stage changes or short arousals alone is low compared to reactions associated with awakenings.

(iv) The majority of awakenings last for exactly one epoch (15 to 45 s) and, therefore, are too short to be remembered on the next day. On the other hand, single awakenings may last longer and, therefore, be associated with the occurrence of waking consciousness. As a consequence, these longer awakenings may be recalled on the next day. In this case, they will also dominate the subjective assessment of sleep quality and quantity on the next day. Sleep stage changes and arousals will not be remembered on the next day as they do not lead to the occurrence of waking consciousness.

Sleep stage S1 does not contribute or only little contributes to the recuperative value of sleep. On the contrary, increased fractions of sleep stage S1 were identified as typical effects of sleep fragmentation in the past.[8] Hence, in this analysis not only changes to stage awake were regarded as relevant sleep stage changes, but also changes to sleep stage S1. This preventive measure increased the fraction of reactions associated with ANEs without significantly lowering the specifity of the proposed indicator, which is also called the sleep fragmentation index (SFI). Other authors also prefer to use this indicator in noise effects research.[15] The SFI was shown to correlate highly with the arousal index following ASDA criteria.[16,14] Therefore, in this publication the term "awakenings" implicitly means transitions from sleep stages REM, S4, S3, or S2 to the sleep stages S1 or awake.

Awakening probability does not solely depend on the maximum SPL of the ANE. On the one hand, other acoustical characteristics of the noise event (spectral content, duration, etc.) play an important role. On the other hand, situative and individual factors moderate the reactions to aircraft noise.[2] Therefore, in order to assess the influence of the maximum SPL, the other moderating factors have to be controlled for, which is called adjustment. Since an awakening represents a dichotomous dependent variable (yes/no), logistic regression was used for the analyses. As every subject was exposed to multiple ANEs, the observed reactions within one subject were not independent. Hence, random effects logistic regression was used, which is able to handle clustered data.[17]

Environmental conditions are less controlled in field studies compared to laboratory studies. The emergence of an ANE from the background noise level was identified as an important factor for the incidence of noise-induced awakenings. Therefore, in the field the background noise level was estimated for the minute preceding each ANE. The $L_{AS,eq}$ varied between 16.4 und 58.3 dB with a median of 27.1 dB.

Other noises originating from inside or outside the bedroom may occur during an ANE or between two ANEs. They were identified in the field study. An ANE contributed only to the final analysis if the following conditions were met: (1) In the minute before or during the ANE currently analyzed, only noises that were caused by the subject (except snoring) or by another ANE were allowed. Here, eliminating data with other ANEs in the minute before the start of the ANE currently analyzed could have led to a systematic underestimation of awakening probabilities in times of high air traffic. (2) Noises produced by the subject during the ANE currently analyzed were explicitly not discarded from the analysis, as they could have been caused by a reaction to the ANE. For each ANE, every of the other investigated nights was checked for ANEs in the same period according to the elapsed time after sleep onset. If there was no ANE, this period was used for the estimation of spontaneous awakening probability.

III. RESULTS

In total, 61 of 64 subjects contributed to the final analysis with 483 subject nights, in which 15 556 ANEs were recorded. The data of three subjects had to be discarded because of constant snoring (two subjects) or an intrinsic sleep disorder (one subject). The first night was not analyzed because of the so-called first-night effect.[18] In total, 10 658 ANEs met the inclusion criteria (see above) and contributed to the regression analyses.

Table I summarizes the results of a multivariable random effects logistic regression model (software Egret, Version 2.0.31, Cytel Corp.). The model contains the maximum A-weighted SPL ($L_{AS,max}$) and the background noise level in the minute preceding the ANE ($L_{eq_}1$min) as well as their interaction term $L_{AS,max} \times L_{eq_}1$min as statistically significant variables. Additionally, the sleep stage prior to the occur-

TABLE I. Results of a random effects logistic regression based on 61 subjects, 483 subject nights, and 10 658 ANEs. −2 log L=6659.8 with 10 650 degrees of freedom.

	Coefficient	Standard error	p
Intercept	−7.0734	0.8816	<0.001
$L_{AS,max}$	0.0946	0.0185	<0.001
L_{eq-}1min	0.1319	0.0327	<0.001
$L_{AS,max} \times L_{eq-}$1min	−0.0027	0.0007	<0.001
Elapsed sleep time	0.0006	0.0002	<0.001
Prior stage S3 and stage S4	−0.3205	0.1161	0.0058
Prior REM	0.4195	0.0733	<0.001
Random subject effect	0.3395	0.0540	

rence of an ANE (indicator variables prior stage S3 and stage S4 and prior REM) as well as elapsed sleep time are incorporated as statistically significant moderators in the model.

Awakening probability increases with maximum SPL $L_{AS,max}$ of the ANE, with background noise level $L_{eq_}$1min as well as with elapsed sleep time (positive coefficients). Awakening probability is lower from deep sleep (stages S3 and S4) and higher from REM sleep compared to stage S2. Nevertheless, stage S2 constitutes the most vulnerable sleep stage according to noise-induced awakenings, as the probability of spontaneous awakenings was much higher from REM sleep than from stage S2 sleep [see Eq. (1)]. The statistically significant interaction of maximum SPL $L_{AS,max}$ and background noise level $L_{eq_}$1min corroborates the importance of the emergence of an ANE from the background noise level.

Figure 1 illustrates the relationship between the maximum SPL of an ANE and the percentage awakened based on results of the regression model presented in Table I (black line). The background noise level was assumed constant with 27.1 dB (median). For preventive reasons, the sleep stage prior to the ANE was assumed to be stage S2 in all cases, i.e., the most sensitive sleep stage. Likewise, elapsed sleep time was set to the middle of the more sensitive second half of the night (epoch 601, about 5 h after sleep onset in the field study).

The highest SPL measured in the field inside the bedroom was 73.2 dB. Spontaneous changes to awake or stage S1 occurred with a probability of 8.6% (dashed line). A threshold value of about 33 dB was found in the field study, i.e., awakening probability increased only for ANEs with maximum SPL above 33 dB compared to spontaneous awakening probability (see Fig. 1). This threshold was only 6 dB above the background noise level, which seems physiologically plausible: First noise-induced awakenings should be observed once the human auditory system is able to differentiate the ANE from the background noise. Nevertheless, it must be pointed out that the awakening probability just above the threshold is accordingly very low. Only 2 of 1000 people exposed to an ANE with a maximum SPL of 34 dB will show a noise-induced awakening. Due to the large number of subjects and ANEs, the precision of the point estimate is very high, i.e., the width of the 95% confidence interval is very low (3.1% at 39 dB and 10.5% at 73.2 dB).

As already mentioned, awakenings are not specific for aircraft noise, as they also occur spontaneously. The probability of noise-induced awakenings is calculated by sub-

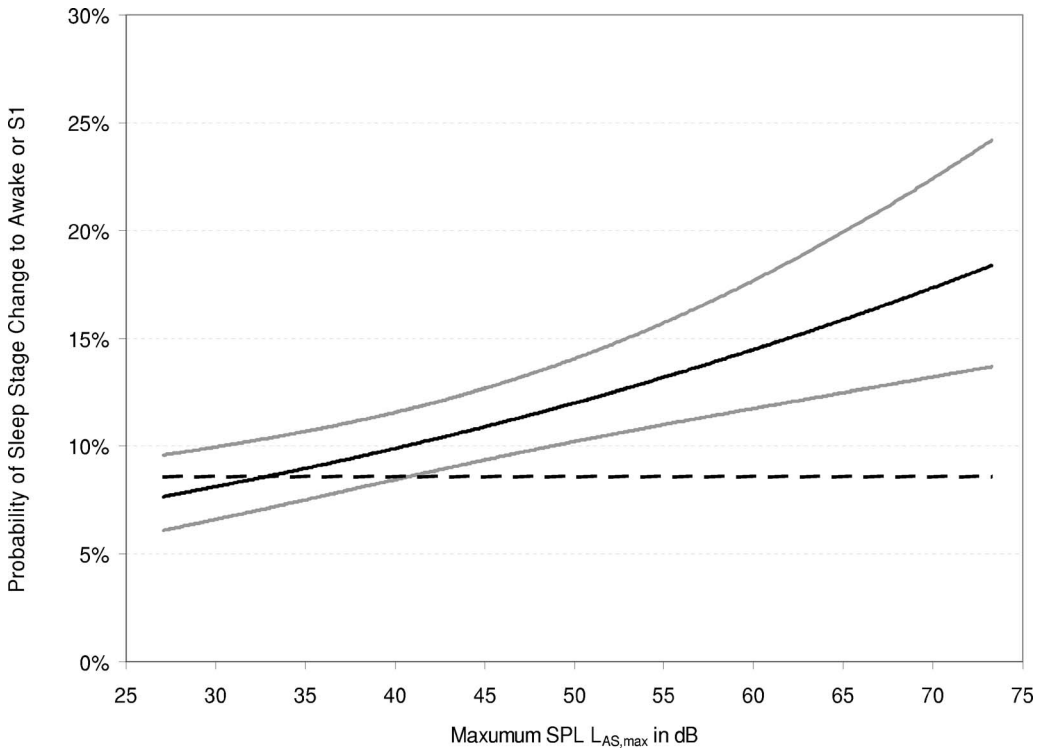


FIG. 1. Probability of sleep stage change to stage S1 or awake depending on maximum SPL $L_{AS,max}$ based on the regression results from Table I. Assumptions: Background noise level L_{eq-}1min=27.1 dB constant (median), prior sleep stage = stage 2, elapsed sleep time = 601 epochs (middle of second half of the night). Point estimates (black line), 95% confidence limits (gray lines), and spontaneous reaction probabilities (dashed line) are shown.

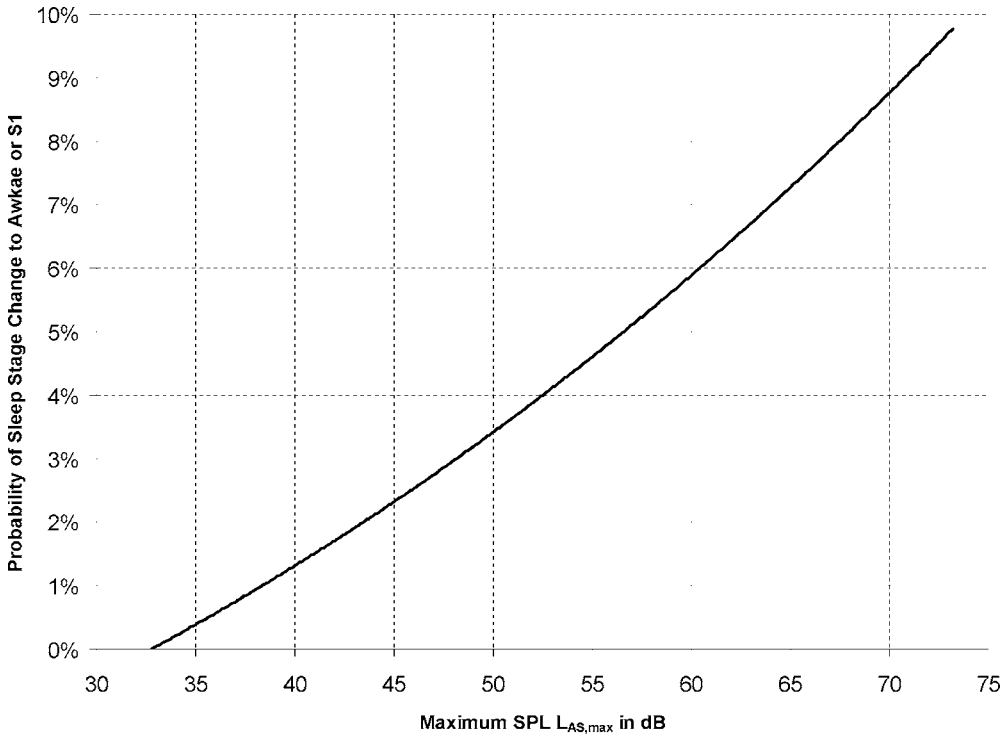


FIG. 2. Probability of aircraft-noise-induced awakenings depending on maximum SPL of ANEs. First reactions occur above maximum SPLs of 32.7 dB. This threshold exceeds the assumed background noise level of 27.1 dB only by 5.6 dB.

tracting spontaneous awakening probability (dashed line in Fig. 1) from awakening probability observed under the influence of aircraft noise (black line in Fig. 1) as indicated in Eq. (1). Aircraft-noise-induced awakening probability depending on the maximum SPL of an ANE is shown in Fig. 2.

The regression line can be approximated with a second-degree polynomial between 32.7 and 73.2 dB. Awakening probability in % is calculated as

$$P_{\mathrm{AWR}} = 1.894 \times 10^{-3} L_{\mathrm{AS,max}}^2 + 4.008 \times 10^{-2} L_{\mathrm{AS,max}} - 3.3243. \quad (2)$$

The awakening probabilities calculated by the polynomial deviate less than 0.1% from the original regression line within the specified interval.

Both the number and the duration of aircraft-noise-induced awakenings play an important role for the evaluation of the effects of aircraft noise on sleep, because the probability of a recalled awakening in the morning increases with the awakening duration. Thus, the results of the DLR laboratory study showed that awakening duration increased with the maximum SPL of an ANE (see Fig. 3).

Awakenings induced by ANEs with maximum SPLs of 65 dB or lower were relatively short. After 1.5 min, descriptively no difference in the percentage of subjects having fallen asleep again compared to spontaneous awakenings was observed. In contrast to that, awakenings induced by ANEs with maximum SPLs of 70 dB or higher were markedly longer than spontaneous awakenings.

IV. DISCUSSION OF NOISE PROTECTION STRATEGIES

In Germany, the most recent proposal for the protection against aircraft noise effects on sleep is based on a combination of *number above threshold* (NAT) and equivalent continuous sound level (L_{eq}) criteria ("Beschluss zur Novelle des Fluglärmgesetzes" from 25 May 2005). Pros and cons of NAT and L_{eq} criteria will be briefly discussed here based on the findings of the DLR field study. Both criteria are calculated from acoustical parameters (maximum SPL, time integrated SPL, or noise duration).

A. Number above threshold (NAT) criteria

NAT criteria are based on the assumption that below a defined threshold value no or only negligible effects of aircraft noise on sleep can be found. It was shown in Sec. III that first noise-induced awakenings can be expected if the maximum SPL exceeds 33 dB. Current proposals recommend limit values for NAT criteria between 52 and 55 dB, i.e., markedly above the threshold found in the DLR field study. Awakenings induced by ANEs with maximum SPLs between 33 dB and the proposed limit value are therefore not taken into account by the corresponding NAT criterion. Theoretically, an arbitrary number of ANEs with maximum SPLs below the NAT limit value are permitted without violating the NAT criterion, but simultaneously inducing relevant sleep disturbances.

NAT criteria also limit the number of ANEs above the threshold value, but without a definition of how much this threshold value may be exceeded by single ANEs. For example, a NAT criterion of 4×52 dB states that a maximum

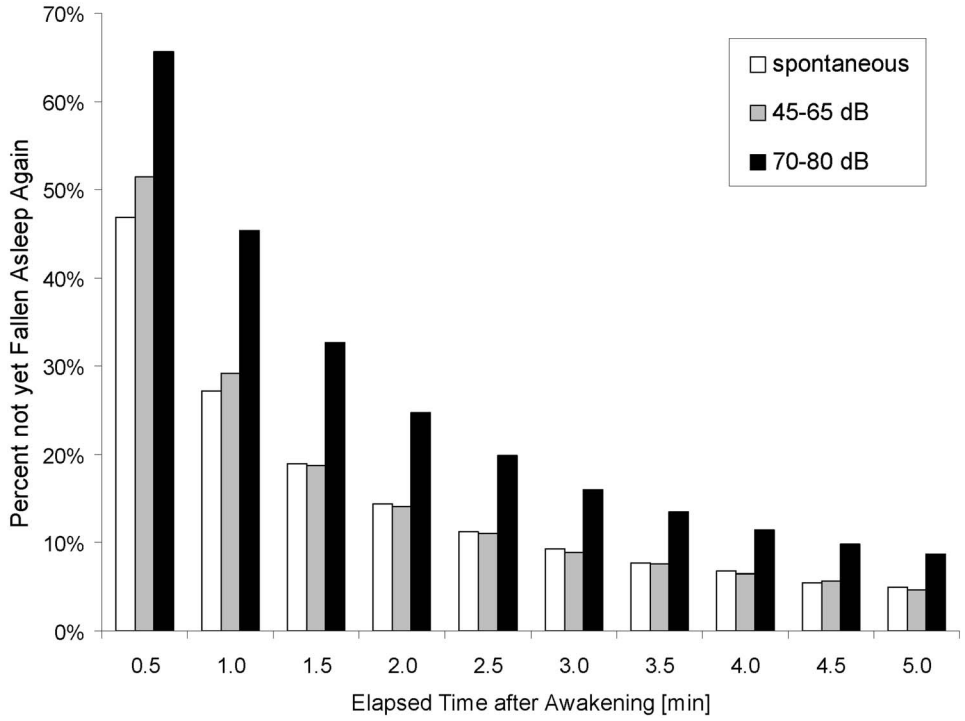


FIG. 3. Duration of noise-induced awakenings compared to spontaneous awakenings. Parameter is maximum SPL.

SPL of 52 dB may be exceeded no more than four times. Therefore, the criterion is neither violated by four ANEs with maximum SPLs of 53 dB nor by four ANEs with maximum SPLs of 73 dB. Calculations based on the dose-response relationship established in the DLR field study expect that 16 of 100 airport residents will be woken up by four events with 53 dB, whereas 39 of 100 residents, i.e., more than twice as many, will be woken up by four events with 73 dB (see Fig. 2).

B. L_{Aeq} criteria

Reducing the number of ANEs by 50% without changing the aircraft types means that the energy equivalent continuous sound level L_{Aeq} will decrease by 3 dB. Criteria solely depending on L_{Aeq} therefore implicitly assume that the effects of aircraft noise on sleep are simultaneously diminished by 50%, e.g., that the number of awakenings induced by aircraft noise is halved. Figure 4 demonstrates that this is not true. Following the epidemiologic concept of *numbers needed to harm*, it shows, depending on the maximum SPL of single ANEs, how many ANEs are needed to induce one additional awakening on average, where independent events were assumed.

If the maximum SPL of single ANEs is reduced by 3 dB from 72 to 69 dB, the permitted number of ANEs inducing one additional awakening may not be doubled but only in-

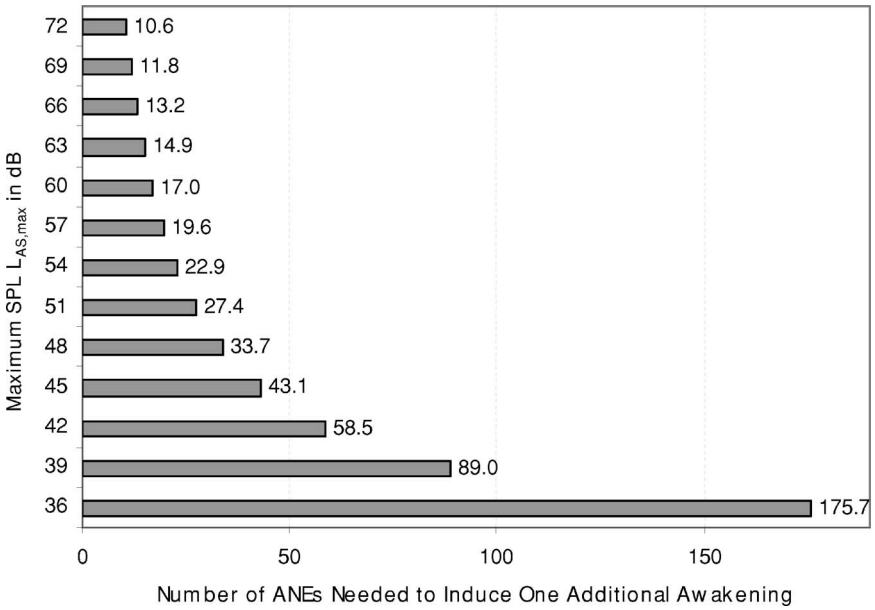


FIG. 4. Number of ANEs needed to induce one additional awakening on average and depending on the maximum SPL. Results are based on the dose-response relationship found in the field study (see Table I).

creased by 11% from 10.6 to 11.8 movements. The allowable change in the number of ANEs following reductions in maximum SPL of 3 dB increases continuously from 11% (decrease from 72 to 69 dB) to 97% (decrease from 39 to 36 dB), i.e., the number of ANEs may be nearly doubled only very close to the threshold value of 33 dB.

C. Combinations of NAT and L_{Aeq} criteria

Potentially, there are two main advantages of combining NAT and L_{Aeq} criteria:

(i) The L_{Aeq} is mainly influenced by events with high SPLs. If ANEs are much louder than the limit value of the NAT criterion, the L_{Aeq} criterion will be violated quickly.

(ii) The number of ANEs with maximum SPLs between the physiological threshold (33 dB) and the limit value of the NAT criterion cannot be increased at will without violating the L_{Aeq} criterion. Especially in the case of high traffic volumes, the L_{Aeq} criterion will dominate the combined criterion, although there is a strong dependence on the actual limit values for both criteria.

Nevertheless, the problems associated with each of the criteria are not completely solved by the combination of both criteria. There are several constellations concerning the number and the maximum SPL of ANEs with maximum SPLs between the physiological threshold and the limit value of the NAT criterion that violate neither the L_{Aeq} criterion nor the NAT criterion, but lead to a relevant number of noise-induced awakenings nevertheless. A publication of the Health Council of the Netherlands states that, given a certain L_{night}, the least favorable situation regarding a direct biological effect occurs if maximum SPLs of single ANEs are about 5 dB higher than the physiological effect threshold.[19] Of course, this noise pattern constitutes an unrealistic *worst case scenario*.

D. A physiological noise effects criterion

As mentioned above, combinations of the acoustical NAT and L_{Aeq} criteria dominate current proposals of concepts for the protection of sleep against adverse effects of aircraft noise. Practically, noise protection zones around airports are represented by noise contours, i.e., curves on which a certain noise descriptor has a constant value. Such contours are usually estimated by calculations rather than by measurements. They are derived from the maximum SPL as well as the duration and number of ANEs, which again are based upon the airport's traffic description. This description, which can be a forecast, specifies the number and types of aircraft operating at the airport during a well defined time period. Until now, noise contours around German airports were solely based on acoustical criteria, i.e., areas where a certain L_{Aeq} is exceeded (e.g., 50 dB outside) or a certain maximum SPL is exceeded too often (e.g., 6×75 dB outside). By establishing these purely acoustical contours, it was implicitly assumed that aircraft-noise-induced sleep disturbances are acceptable outside the contours without additional sound insulation measures.

The dose-response relationship established in the DLR field study (see Fig. 2) can be combined with the estimation of immission values in order to explicitly specify the effects of aircraft noise on sleep around airports. The average number of aircraft-noise-induced awakenings at a certain location in the airport environment is calculated from the distribution of A-weighted maximum SPLs $n(L_{AS,max})$ at this location:

$$N_{AWR} = \int_{-\infty}^{\infty} f_{AWR}(L_{AS,max}) n(L_{AS,max})\, dL. \tag{3a}$$

The function f_{AWR} follows from Eq. (2) as

$$f_{AWR}(L_{AS,max}) = \max(1.894 \times 10^{-5} L_{AS,max}^2 + 4.008 \times 10^{-4} L_{AS,max} - 3.3243 \times 10^{-2}; 0). \tag{4}$$

The max-function assures that there are no negative contributions of maximum SPLs below the threshold of 33 dB. The assumption that the function f_{AWR} is still valid above the range of 73 dB is arbitrary. In practice, there are no problems associated with this assumption, as maximum SPLs of this magnitude inside the bedroom only occur in highly exposed areas close to the airport.

These equations can easily be implemented in any calculation procedure capable of providing distributions of maximum SPLs (e.g., the German AzB procedure, which was used for the calculations in this publication). In practice level distributions are realized by SPL classes of a certain width rather than by a distribution function $n(L_{As,max})$. In that case, Eq. (3a) migrates to the following equation:

$$N_{AWR} = \sum_i f_{AWR}(L_{AS,max,i}) n(L_{AS,max,i}). \tag{3b}$$

The summation has to be performed over all level classes denoted by the index i. It is likely that there will be an influence of the class width. In order to minimize this effect, the calculations performed for this investigation were carried out with a class width of 0.2 dB. Additionally, normally distributed maximum SPLs with a standard deviation of 3 dB were assumed instead of the discrete maximum SPL values provided by the AzB algorithm for the particular aircraft categories. This is currently also a common approach for the calculation of NAT contours. There is some potential to improve the implementation of the N_{AWR}-calculation scheme into existing aircraft noise calculation tools. Such optimizations are currently the subject of further investigations.

With the method described above, the number of aircraft-noise-induced awakenings can be predicted for each location around the airport. Hence, the need for protective measures against the adverse effects of aircraft noise can be quantified explicitly and precisely. This is illustrated in Fig. 5 for Frankfurt airport. Two areas based on L_{Aeq} criteria are compared with three areas outside of which less than one, two, or three additional noise-induced awakenings are expected.

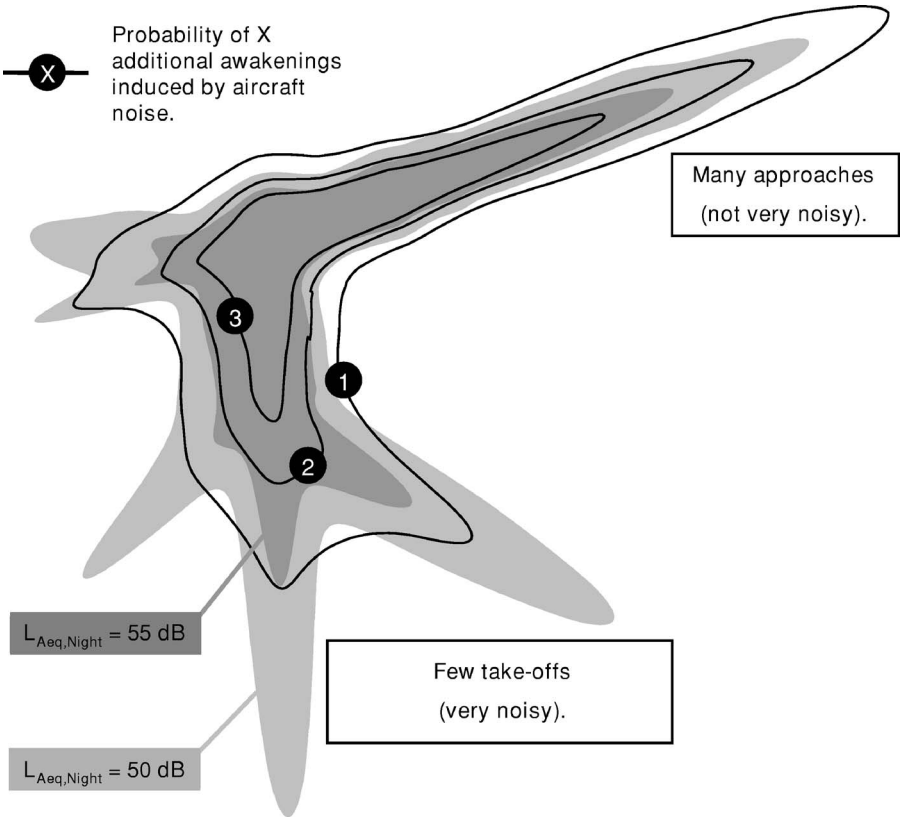


FIG. 5. Prognosis of noise effects for Frankfurt airport: on average one, two, or three additional awakenings induced by aircraft noise (black lines), L_{Aeq}=55 dB (dark gray), and L_{Aeq} =50 dB (light gray). Calculations are based on 25 000 nocturnal aircraft movements in the busiest six months of the year.

Apparently, there are qualitative differences: On the one hand, the contours for additional awakenings extend into areas with many but relatively quiet fly-overs (approaching aircrafts). On the other hand, in areas with few but relatively loud fly-overs (departures) they are not as pronounced as the L_{Aeq} contours. This illustrates the fact that L_{Aeq} criteria are not suitable for an adequate description of the effects of nocturnal aircraft noise on sleep.

The introduction of a physiological measure for sleep disturbances (i.e., the probability of awakenings) reflects a more medical- and health-related position than acoustical dimensions can do. The combination of a physiological reaction and an acoustical event, represented in a dose-response curve, and the calculation of acoustical immissions at a given location in the vicinity of airports provides a powerful tool for the protection of the affected population. Therefore, an easily applicable concept for the protection against adverse effects of nocturnal aircraft noise on sleep has been developed.

V. CONCEPT FOR THE PROTECTION AGAINST ADVERSE EFFECTS OF NOCTURNAL AIRCRAFT NOISE

The DLR concept for the protection against adverse effects of nocturnal aircraft noise on sleep will be presented and substantiated. Potential restrictions of the concept will also be discussed.

A. Objectives of the concept

The adequate protection of people affected by nocturnal aircraft noise has to be the main objective of a protective concept in order to prevent negative health consequences. Changes in sleep structure that may lead to a nonrestorative sleep are the primary effects of nocturnal aircraft noise. Sleepiness and impaired mental capacities are two of the possible immediate consequences. Furthermore, annoyance may be induced by consciously perceived noise events during the night. It is also being discussed whether repeatedly (over years) occurring noise-induced sleep disturbances may lead to other health impairments, such as an increased risk for high blood pressure or myocardial infarction.[20–22] If established, these noise impacts on health would be of major societal importance. However, in practice it is very difficult to substantiate a causal link between noise and long-term health effects, as many different and well-proven risk factors lead to the same diseases and induction periods are usually very long. Until now, there is no study corroborating this causal link for nocturnal aircraft noise.[19]

In order to overcome this dilemma, the DLR concept is based on two assumptions:

(1) Because of biological plausibility, it is hypothesized that a causal link between noise-induced sleep disturbances and long-term health effects exists. Vice versa, long-term health effects can be prevented with a high probability if noise-induced sleep disturbances are minimized.
(2) It is assumed that humans—like any organism—represent an adaptive system, which is able to compensate for certain strains without negative effects for the organism. Hence, it is not necessary to eliminate strains completely. It is simultaneously assumed that there are very sensitive subjects, who fail to compen-

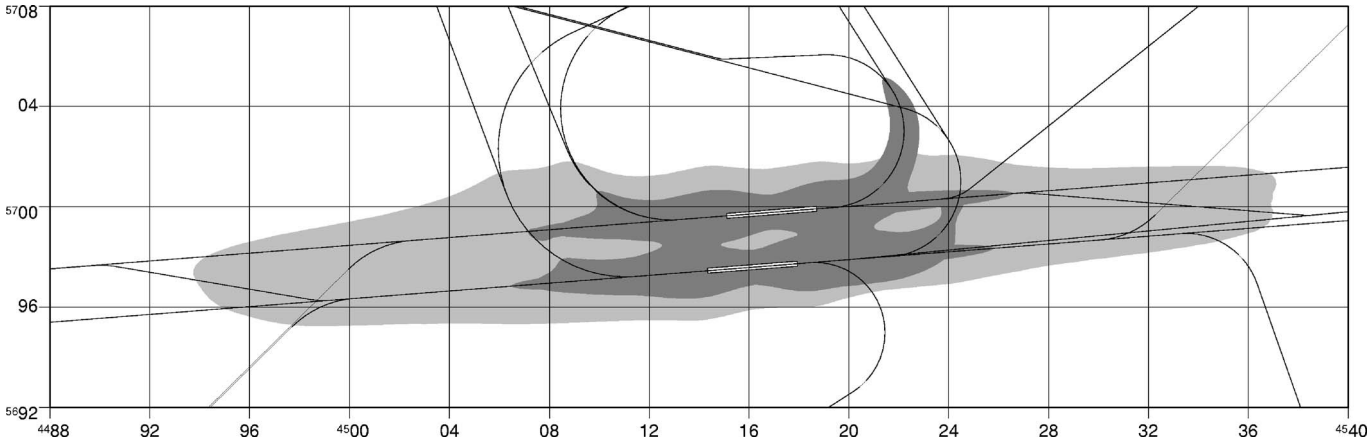


FIG. 6. Noise protection zone for Leipzig/Halle airport (traffic prognosis for 2015) consisting of the combination of two areas: (1) area outside of which less than one additional awakening induced by aircraft noise is expected on average (light gray) and (2) area outside of which maximum SPLs of 80 dB or higher (measured outside) occur less than once (dark gray, envelope of two contours calculated separately for 100% flight movements in each direction).

sate for small strains, as well as there are very robust subjects, who endure strong strains without negative consequences.

From a medical point of view, sound insulation should only be used if all other measures fail. Methods for the reduction of noise emissions are an active research field (e.g., silent engines, noise reduced takeoff and landing procedures, etc.). They should be consequently and quickly applied in practice.

B. Description of the concept

The DLR concept is based on three objectives, which reflect three highly correlated dimensions of sleep:

(1) On average, there should be less than one additional awakening induced by aircraft noise. Here, awakenings are defined as an electrophysiological phenomenon classified according to the rules of Rechtschaffen *et al.*[5]
(2) Awakenings recalled in the morning should be prevented as much as possible.
(3) There should be no relevant impairment of the process of falling asleep again.

Figure 6 illustrates the proposed noise protection zone for Leipzig/Halle airport for the night (22:00 until 06:00), based on a traffic prognosis for 2015. Two contours are combined: Outside of the light gray area on average less than one additional awakening induced by aircraft noise is expected. This contour is based on the expected, average distribution of flight movements on the two operation directions. Outside of the dark gray area, maximum SPLs of 80 dB or higher (measured outside) occur less than once. This contour is the envelope of two contours estimated for a 100% distribution of flight movements in both operating directions. This leads to an overestimation of effects, which was intended as awakenings recalled in the morning are regarded as especially severe sleep disturbances. The three columns of the DLR concept will be discussed in detail in Secs. V C to V E.

C. Less than one additional awakening induced by aircraft noise on average (first criterion)

Self-evidently, humans either wake up or they do not, i.e., on the individual level and within one night noninteger values for the number of additional awakenings do not make sense. However, they do if one refers to more than one night or to more than one subject (i.e., to averages). The criterion "on average less than one awakening per night" would be violated if a subject is woken up by aircraft noise 365 times in one year. However, it would just not be violated if the subject is woken up by aircraft noise 364 times in one year. When interpreting these numbers it has to be kept in mind that about 24 spontaneous awakenings can be expected per night on average and therefore about 8760 spontaneous awakenings can be expected per year.[2]

If and how often a subject is actually woken up by aircraft noise depends on the amount of air traffic in the special night, other situative and individual factors, as well as on chance. Therefore, in single nights it is possible that a subject is woken up more than once, e.g., two times. If the criterion should not be violated it has to be guaranteed that the subject is not woken up by aircraft noise in one other night, thus compensating for the two awakenings. The same is true for an even higher number of noise-induced awakenings: If a subject is woken up four times by aircraft noise in one night, this has to be compensated for by three nights with no additional awakenings, otherwise the criterion would be violated in the long run.

A Monte Carlo Markov chain (MCMC) simulation was used to calculate how the numbers of noise-induced awakenings per night are distributed over the 365 nights of one year: Maximum SPLs were randomly drawn from the maximum SPL distribution found in the DLR field study. For each SPL, awakening probabilities were calculated according to the dose-response relationship shown in Fig. 2. With a random number generator and based on the derived awakening probability it was determined whether the simulated human subject woke up or did not wake up induced by aircraft noise. This procedure was repeated and awakening probabilities were summed until the criterion of one additional aircraft-

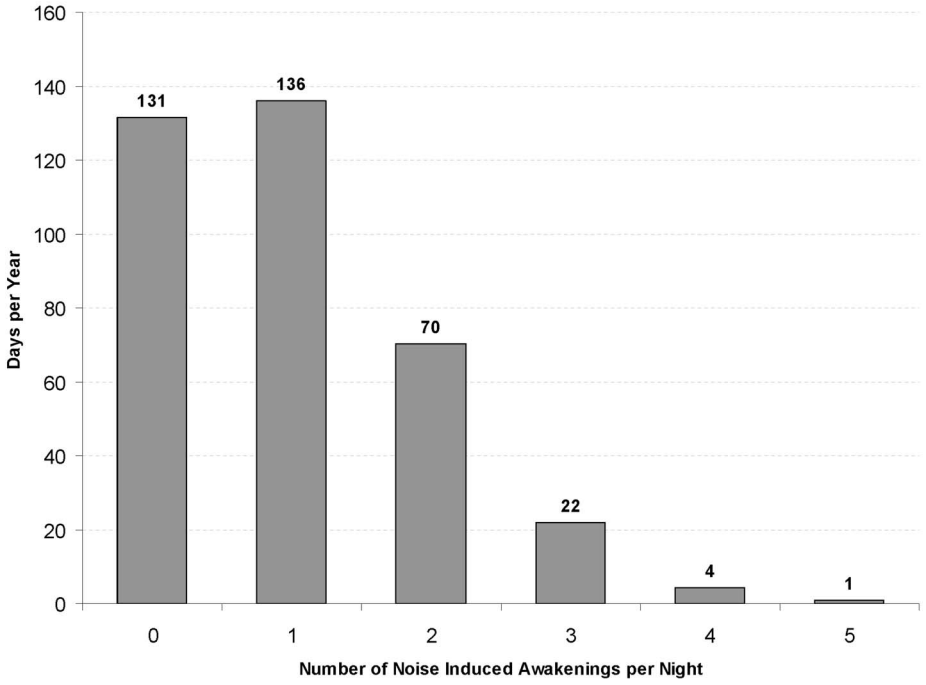


FIG. 7. Distribution of aircraft-noise-induced awakenings over the 365 days of one year if the criterion "on average less than one noise-induced awakening" is just not violated.

noise-induced awakening was just violated. This noise event was the last event counted, unless the sum of awakening probabilities up to the previous noise event, that just did not violate the criterion, was closer to the limit value in absolute terms. In that case, the previous noise event was the last event counted. With this method, 1 000 000 noise nights were simulated.

The expected distribution of aircraft noise-induced awakenings over the 365 days of one year is shown in Fig. 7.

In 131 nights there is no, in 136 nights there is one, in 70 nights there are two, and in 22 nights there are three additional awakenings. Four or five noise-induced awakenings per night occur extremely seldomly, and six or more awakenings practically do not occur.

These reactions are electrophysiological awakenings based on the definition by Rechtschaffen and Kales.[5] They are usually too short to be remembered on the next day (see above). From a preventive point of view, however, the number of electrophysiological awakenings induced by aircraft noise should be restricted as much as possible,[23] although the impact of electrophysiological awakenings on health, quality of life, and psychological outcomes remains a matter of scientific debate and therefore uncertain.[24] As electrophysiological awakenings go along with vegetative arousal reactions (e.g., increases in heart frequency and blood pressure), it is at least biologically plausible that repeatedly occurring noise-induced awakenings over years may impact health.[13] It is currently not known how many noise-induced awakenings are tolerable without leading to middle- or long-term impairments of well-being and health. With the high number of spontaneous awakenings and the high variability in several nights of the same person in mind, it is not deemed necessary from a medical point of view to completely avoid additional awakenings induced by aircraft noise. It is rather assumed that impacts of aircraft noise on health can be excluded in areas where less than one additional awakening is expected to be induced by aircraft noise on average.

D. Recalled awakening (second criterion)

The risk of recalled awakening increases with the duration of the awakening. Recalled awakenings are correlated with the subjective evaluation of sleep quality and sleep quantity: The higher the number of recalled awakenings, the worse is the evaluation of sleep quality (field study: $r_{\text{Spearman}}=-0.316$) and sleep quantity (field study: $r_{\text{Spearman}}=-0.269$). Additionally, ANEs occurring in the sleep period influence the assessment of annoyance only when they are perceived consciously by airport residents, and longer awakenings are a prerequisite for regaining consciousness.[19]

Recalled awakenings not only fragment sleep: They go along with psychological disadvantages as well and therefore constitute a major sleep disturbance. Psychosomatic disorders cannot be excluded if recalled awakenings are induced over longer time periods. Therefore, special attention has to be drawn to recalled awakenings in the process of evaluating the impacts of aircraft noise on sleep. From a medical point of view, recalled awakenings induced by aircraft noise should be prevented as much as possible.

The first criterion limits the number of noise-induced awakenings irrespective of the duration of the awakenings. Therefore, the number of recalled awakenings is limited as well. Most of the spontaneous awakenings are too short to be remembered on the next day. Analyses of the laboratory study showed that the duration of noise-induced awakenings increases with the maximum SPL of ANEs. Relevant differences compared to spontaneous awakenings were observed for maximum SPLs of more than 65 dB (see Fig. 3).

For this reason, maximum SPLs of more than 65 dB should be avoided in the bedroom. For a partly opened window with an assumed difference in SPLs of 15 dB between inside and outside, the 1×80 $dB_{outside}$ contour of Fig. 6 (dark gray) assures that outside this area maximum SPLs of 65 dB are exceeded less than once inside the bedroom on average. As recalled awakenings should be avoided as much as possible, this contour is based on a 100% flight movement in one direction estimation, i.e., a worst case.

E. Falling asleep again (third criterion)

The problem of falling asleep again is practically not considered in the literature of noise effects on sleep, disregarding the fact that about 7% of the sleep period is spent awake.[13] ANEs can prevent the sleeper from falling asleep again in these situations, and therefore have a negative impact on sleep structure.[2]

The traffic prognosis for Leipzig/Halle airport in 2015 forecasts two very busy periods during the night caused by freight handling. Between 0:00 and 1:30 up to 60 approaches per hour and between 4:00 and 5:30 up to 50 starts per hour are expected. The short time period of 1 to 1.5 min between two noise events in these peak hours leads to an increased risk of preventing the affected population from falling asleep again. If a subject already regained consciousness, annoyance reactions may result from consciously perceived noise events. Indeed, many airport residents complain about ANEs in early morning hours. At this time of the night, sleep pressure and awakening thresholds are low: Falling asleep again is difficult anyhow and aggravated by aircraft noise events.

Extensive analyses based on the data of the DLR field study were performed to assess the impact of aircraft noise on falling asleep again. The analyses are complicated by the fact that people who are prevented from falling asleep again stay awake and may be repeatedly prevented from falling asleep again by additional ANEs. In statistical terms, probabilities are no longer independent, but they are conditional on what happened in the past.

The results of these analyses will be presented elsewhere. They are based on a Markov state transition model which differentiates between two states only: awake and sleep. In the model, transitions between these two states depend on maximum SPL $L_{AS,max}$ of the ANE, elapsed sleep time, the current state (awake/sleep), and the elapsed time spent in the same sleep stage and estimated with autoregressive logistic regression based on the data of the field study. The Markov model was used to predict the number of awakenings, the duration of wake periods, the number of awakenings recalled in the morning, and the percentage of highly annoyed subjects. The results indicated that maximum SPLs of ANEs in the second half of the night should receive a malus of 1.4 dB, i.e., they should be artificially elevated by 1.4 dB, in order to assure an undisturbed process of falling asleep again similarly in all regions around Leipzig/Halle airport.

TABLE II. Relevant polysomnographic field studies on the effects of aircraft noise on sleep. In the study of Hume *et al.*[15] SPLs were measured outside the bedroom only.

Study	No. of subjects	No. of subject nights	Age range (years)
Basner *et al.*[2] (2004)	64	576	19–61
Hume *et al.*[15] (2003)	46	178	20–70
Flindell *et al.*[11] (2000)	18	90	30–40
Ehrenstein *et al.*[26] (1982)	3	30–45	8–10
Vallet *et al.*[25] (1980)	40	160	20–55

F. Transferring study results to the population level

Polysomnography is the only method that allows us to draw conclusions on structural aspects of sleep. At the same time, data acquisition and analysis are cumbersome, time consuming, and therefore expensive. Hence, polysomnographic studies on the impact of aircraft noise on sleep are scarce and were usually based upon small samples. These studies differed considerably in study design and the methods applied, thus complicating comparisons or meta-analyses between them. As there is considerable intersubject variability in noise sensitivity, small studies may per chance investigate only very sensitive or only very insensitive subjects according to aircraft-noise-induced changes in sleep structure. Hence, transferring the results of small studies to the population level is not possible or limited.

The sample size of the DLR field study is compared to other relevant field studies using polysomnography in Table II.

Vallet[25] investigated the effects of aircraft noise on sleep in the field on 40 subjects aged 20 to 55 years (160 subject nights). Flindell *et al.*[11] examined 18 subjects aged 30 to 40 years on five consecutive nights (90 subject nights). The authors call their study a pilot study for a potential extension of the Ollerhead *et al.* study.[10] Ehrenstein[26] investigated three children aged 8 to 10 years in the vicinity of Munich airport. As there was no air traffic after 21:30, only ANEs between 20:00 and 21:30 could be analyzed. Ollerhead *et al.*[10] performed an actigraphic study in the vicinity of several British airports. Polysomnography was additionally performed on 46 of the 178 subjects. The results of the polysomnographic data were published by Hume *et al.*[15] in 2003. In this study, the SPL was measured only outside of the dwellings, restricting the validity of the results. None of the studies investigated nonhealthy subjects.

In the DLR field study, 64 subjects were studied for 576 subject nights, resulting in the largest polysomnographic study with identical methodological approach so far. Nevertheless, the study does not claim representativeness for the whole population. It is impossible to be representative for a whole population in a study with huge methodological expenses for a single subject like the DLR study. Additionally, some inclusion criteria had to be met in order to be eligible for study participation, leading to a higher internal validity of the results. This is a prerequisite for external validation, but also it restricts it to some extent.[2]

Therefore, the results of the field study were not trans-

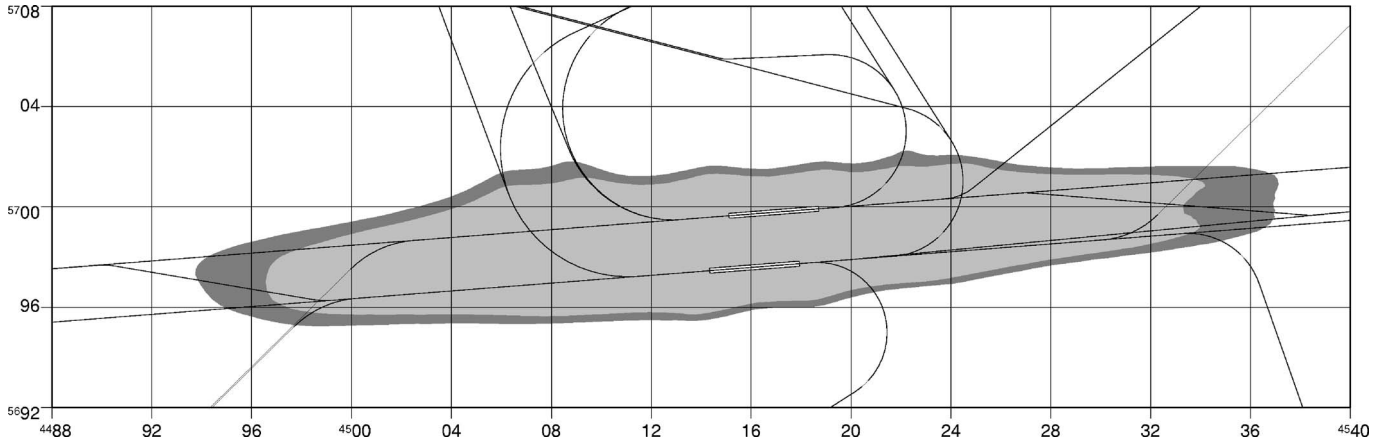


FIG. 8. Areas outside of which less than one awakening is additionally induced by aircraft noise. Comparison of preventive approach (sleep stage S2, middle of second half of the night, outer dark gray area) with approach based on actual data found in the field study (inner light gray area). By the preventive approach the noise protection zone is increased by 28% (=43 km^2).

ferred 1:1 to the population living in the vicinity of Leipzig/Halle airport. Instead, several preventive measures were taken in order to protect those parts of the population that were not represented in the DLR field study and that are more sensitive to aircraft noise at the same time. Some of these measures shall be briefly summarized:

(i) Subjects assessing themselves as sensitive to and annoyed by aircraft noise were included preferably into the study. Seventy-five percent of study subjects assessed themselves as moderately, strongly, or very strongly annoyed, compared to 15% of a representative German survey.[27]

(ii) Not only awakenings but also sleep stage changes to stage S1 were counted as relevant noise-induced sleep disturbances, increasing the probability of reactions to aircraft noise.

(iii) For the calculation of the dose-response curve based on the regression results it was assumed that the sleeper spent the whole night in the most sensitive sleep stage S2 and in the middle of the more sensitive second half of the night. In reality, an average night contains only about 50% of sleep stage S2. Hence, the dose-response curve is shifted to higher probabilities compared to calculations where the actual sleep stage distribution is used. Because of this measure alone the noise protection zone increases from 156 km^2 by 28% to 199 km^2 (see Fig. 8).

(iv) Subjects with illnesses leading to a lower noise sensitivity (e.g., hypakusis, hypersomnolence) were excluded from study participation.

(v) The calculations for the noise protection zone were based on the six busiest months of the year according to air traffic.

(vi) Sound insulation was increased by 3 dB for sensitive institutions (e.g., hospitals) and individuals with relevant diseases accompanied by a higher noise sensitivity.

(vii) The proposal of allowing only one additional awakening induced by aircraft noise makes sense in terms of preventive medicine. It has to be taken into account that on average 24 spontaneous awakenings can be observed in an otherwise undisturbed night anyway.

VI. CONCLUSIONS

The DLR-Institute of Aerospace Medicine investigated the influence of aircraft noise on sleep, mood, and performance in an extensive polysomnographic field study between 1999 and 2004 as part of the DLR/HGF-project "Quiet Air Traffic." The dose-response relationship developed in this study was used to establish a concept for the protection of subjects against the adverse effects of nocturnal aircraft noise on sleep. Advantages of this new concept compared to conventional NAT criteria, L_{eq} criteria, or combinations of NAT and L_{eq} criteria were discussed.

It is planned to extend Leipzig/Halle airport to a freight hub with an independent parallel runway system. Major parts of air traffic will take place during the night. The Regional Council of Leipzig decided to use the results of the DLR field study for a new noise protection concept at Leipzig/Halle airport. This concept culminates in the three propositions presented and discussed in Sec. V and reflects three correlated dimensions of sleep: There should be on average less than one additional awakening induced by aircraft noise, noise-induced awakenings recalled in the morning should be prevented as much as possible, and no relevant impairments of the process of falling asleep again should occur. These three provisions have been proposed in order to consider the special conditions under which Leipzig/Halle airport will operate: (1) construction of a second independent runway, (2) settlement of a night cargo hub for a big service integrator, (3) heavy air traffic during night including peak hours with up to 60 movements per hour, and (4) practically no nocturnal air traffic in the present. These circumstances necessitate a special concept for the protection of the affected population against the adverse effects of nocturnal aircraft noise on sleep.

With the decision for the implementation of the results of the DLR field study, fresh ground was broken, as noise

protection zones solely depended on acoustical criteria so far. The noise protection zone for nocturnal air traffic proposed by DLR exceeds the one of the current law amendment under discussion, which should come in force in 2011, by 60 km^2, and will be correspondingly accompanied by additional financial burdens for the airport resulting from the installation of sound insulation.

Shortly after the publication of the official documents of the approval process for the extension of Leipzig/Halle airport in November 2004, the integrator DHL decided to move its European freight hub from Brussels to Leipzig/Halle. In the long run, this could lead to several thousands of new jobs in this region. Despite the very conservative approach taken in constructing the noise protection zones, some residents living in the vicinity of Leipzig/Halle airport were still not satisfied with the concept: They sued in order to prevent the start of construction at the airport. The Federal Administrative Court rebutted this legal action in May 2005 in the first instance, and the construction measures started without delay in August 2005.

ACKNOWLEDGMENTS

We would like to thank all subjects who agreed to participate in our studies as well as the investigators from the DLR flight physiology department for their extraordinary effort in sampling and analyzing the data.

[1]Regierungspräsidium Leipzig (Leipzig Regional Council), "Planfeststellungsbeschluss für das Vorhaben Ausbau des Verkehrsflughafens Leipzig/Halle Start-/Landebahn Süd mit Vorfeld" (official documents for the approval process of Leipzig/Halle airport, available on the internet http://www.rpl.sachsen.de/de/internet/service/planung/flughafen/startseite.htm last visited 11/03/2005), Leipzig, Germany, Az.: 14-0513.20-10/14 (2004).

[2]M. Basner, H. Buess, D. Elmenhorst, A. Gerlich, N. Luks, H. Maaß, L. Mawet, E. W. Müller, U. Müller, G. Plath, J. Quehl, A. Samel, M. Schulze, M. Vejvoda, and J. Wenzel, "Effects of nocturnal aircraft noise (Volume 1): Executive summary," German Aerospace Center (DLR), Cologne, Germany, FB2004-07/E, ISSN 1434-8454 (2004).

[3]M. Basner and A. Samel, "Effects of nocturnal aircraft noise on sleep structure," Somnologie **9**, 84–95 (2005).

[4]A. Samel, M. Basner, H. Maaß, U. Müller, J. Quehl, and J. Wenzel, "Wirkungen des Nachtfluglärms auf den Schlaf: Ergebnisse aus dem Projekt 'Leiser Flugverkehr' (Effects of nocturnal aircraft noise on sleep: Results from the 'Quiet Air Traffic' project)," Umweltmed. Forsch. Prax. **10**, 89–104 (2005).

[5]A. Rechtschaffen, A. Kales, R. J. Berger, W. C. Dement, A. Jacobsen, L. C. Johnson, M. Jouvet, L. J. Monroe, I. Oswald, H. P. Roffwarg, B. Roth, and R. D. Walter, "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects," Public Health Service, U.S. Government, Printing Office, Washington, DC (1968).

[6]H. L. Williams, J. T. Hammack, R. L. Daly, W. C. Dement, and A. Lubin, "Responses to auditory stimulation, sleep loss and the EEG stages of sleep," Electroencephalogr. Clin. Neurophysiol. **16**, 269–279 (1964).

[7]M. H. Bonnet, "Performance and sleepiness following moderate sleep disruption and slow wave sleep deprivation," Physiol. Behav. **37**, 915–918 (1986).

[8]N. J. Wesensten, T. J. Balkin, and G. Belenky, "Does sleep fragmentation impact recuperation? A review and reanalysis," J. Sleep Res. **8**, 237–245 (1999).

[9]M. Bonnet, D. W. Carley, M. A. Carskadon, P. Easton, C. Guilleminault, R. Harper, B. Hayes, M. Hirshkowitz, P. Y. Ktonas, S. Keenan, M. Pressman, T. Roehrs, J. Smith, J. K. Walsh, S. Weber, and P. R. Westbrook, "EEG arousals: Scoring rules and examples. A preliminary report from the Sleep Disorders Atlas Task Force of the American Sleep Disorders Association," Sleep **15**, 173–184 (1992).

[10]J. B. Ollerhead, C. J. Jones, R. E. Cadoux, A. Woodley, B. J. Atkinson, J. A. Horne, F. L. Pankhurst, L. A. Reyner, K. Hume, F. Van, A. Watson, I. D. Diamond, P. Egger, D. Holmes, and J. McKean, "Report of a field study of aircraft noise and sleep disturbance," Department of Transport, London, United Kingdom (1992).

[11]I. Flindell, A. Bullmore, K. Robertson, N. Wright, C. Turner, C. Birch, M. Jiggins, B. Berry, M. Davison, and M. Dix, "Aircraft noise and sleep 1999 UK Trial Methodology Study," ISRV Consultancy Services, Southampton, United Kingdom, Ref. 6131 R01 (2000).

[12]R. Mathur and N. J. Douglas, "Frequency of EEG arousals from nocturnal sleep in normal subjects," Sleep **18**, 330–333 (1995).

[13]E. Sforza, F. Chapotot, S. Lavoie, F. Roche, R. Pigeau, and A. Buguet, "Heart rate activation during spontaneous arousals from sleep: Effect of sleep deprivation," Clin. Neurophysiol. **115**, 2442–2451 (2004).

[14]M. J. Morrell, L. Finn, H. Kim, P. E. Peppard, M. S. Badr, and T. Young, "Sleep fragmentation, awake blood pressure, and sleep-disordered breathing in a population-based study," Am. J. Respir. Crit. Care Med. **162**, 2091–2096 (2000).

[15]K. Hume, F. Van, and A. Watson, "Effects of aircraft noise on sleep: EEG-based measurements," Manchester Metropolitan University, Manchester, United Kingdom (2003).

[16]J. Haba-Rubio, V. Ibanez, and E. Sforza, "An alternative measure of sleep fragmentation in clinical practice: The sleep fragmentation index," Sleep Med. **5**, 577–581 (2004).

[17]D. H. Hosmer and S. Lemeshow, *Applied Logistic Regression*, 2nd ed. (Wiley, New York, 2000).

[18]H. W. Agnew, Jr., W. B. Webb, and R. L. Williams, "The first night effect: An EEG study of sleep," Psychophysiology **2**, 263–266 (1966).

[19]Health Council of the Netherlands, 'The influence of night time noise on sleep and health," The Hague: Health Council of The Netherlands, Publication No. 2004/14E, ISBN 90-5549-550-6 (2004).

[20]W. Babisch, "Traffic noise and cardiovascular disease: Epidemiological review and synthesis," Noise Health **2**, 9–32 (2000).

[21]S. Morrell, R. Taylor, and D. Lyle, "A review of health effects of aircraft noise," Aust. N. Z. J. Public Health **21**, 221–236 (1997).

[22]W. Babisch, B. Beule, M. Schust, N. Kersten, and H. Ising, "Traffic noise and risk of myocardial infarction," Epidemiology **16**, 33–40 (2005).

[23]T. Akerstedt, M. Billiard, M. Bonnet, G. Ficca, L. Garma, M. Mariotti, P. Salzarulo, and H. Schulz, "Awakening from sleep," Sleep Med. Rev. **6**, 267–286 (2002).

[24]R. Cluydts, "Comparing the effects of sleep loss after experimental sleep deprivation and in clinical patients," Sleep Med. Rev. **7**, 293–295 (2003).

[25]M. Vallet, J. M. Gagneux, and F. Simonet, "Effects of aircraft noise on sleep: An in situ experiment,"ASHA Report No. 10, 391–396 Rockville, Maryland (1980).

[26]W. Ehrenstein, M. Schuster, and W. Müller-Limroth, "Felduntersuchungen über Wirkungen von Lärm auf den schlafenden Menschen (Field studies on the effects of noise on the sleeping human)," Abschlussbericht des Umweltbundesamtes 1–101, Dessau, Germany (1982).

[27]J. Ortscheid and H. Wende, "Lärmbelästigung in Deutschland—Ergebnisse einer repräsentativen Umfrage (Annoyance caused by noise in Germany)," Z. Lärmbekämpfung **49**, 41–45 (2002).

Acoustic impedance of perforations in contact with fibrous material

Iljae Lee and Ahmet Selamet[a)]
Department of Mechanical Engineering, and The Center for Automotive Research, The Ohio State University, 930 Kinnear Road, Columbus, Ohio 43212

Norman T. Huff
Owens Corning, 4600 Humboldt Drive, Novi, Michigan 48377

(Received 27 September 2005; revised 22 February 2006; accepted 24 February 2006)

Acoustic impedance of perforations in contact with fibrous material, as well as air alone, is experimentally determined in the absence of mean flow. Different porosities (2.1, 8.4, 13.6, and 25.2 %) and hole diameters (0.249 and 0.498 cm) are applied for perforations, along with two fiber filling densities (100 and 200 kg/m^3) to illustrate the effect of variations in such parameters on the perforation impedance. A modified impedance tube setup is developed for the measurement of the acoustic impedance of perforations in contact with fibrous material, and the air alone. The complex wave-number and characteristic impedance of the fibrous material are also experimentally obtained in this study to utilize such properties for the calculation of acoustic impedance of perforations in contact with the material. The experimental results show that both resistance and reactance of the perforations in contact with air decrease as the porosity increases. It is also shown that the fibrous material significantly increases both the resistance and reactance. For a given filling density, both resistance and reactance decrease as the porosity increases, except for the specific samples with highest porosity (25.2%). © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188354]

PACS number(s): 43.50.Gf, 43.58.Bh [DKW] Pages: 2785–2797

I. INTRODUCTION

Perforated ducts are often used in reactive silencers to control the acoustic performance, guide the flow and reduce the flow losses, and influence the sound quality through suppression of flow noise from abrupt cross-sectional area changes. In dissipative silencers, perforated ducts are also utilized to protect the absorbing material, while modifying the acoustic characteristics by adding mainly reactance to the impedance. The predictions for both types of silencers rely on the acoustic impedance of perforations. This impedance is dependent on the perforation geometry (porosity, shape and size of holes, wall thickness, and distance between holes), the flow field (grazing and through flow), sound pressure level, and the medium in contact with the perforation. The complexity of the perforation impedance confines the analytical approaches only to simple perforations, and experimental methods are often used to obtain the impedance. While the acoustic characteristics of perforations in contact with air alone are relatively well established primarily by experimental work (refer, for example, to Sullivan and Crocker, 1978), only limited studies on perforations facing absorbing material are available.

The acoustic impedance of perforations in contact with air alone has been extensively investigated, for example, by Ingard (1953), Sullivan and Crocker (1978), Melling (1973), Dickey *et al.* (2000, 2001), Dean (1974), Rao and Munjal (1986), and Cummings (1986). An empirical formulation of the perforation impedance by Sullivan and Crocker (1978) has been widely used in the linear regime in the absence of mean flow. Melling (1973) and Dickey *et al.* (2000) considered the impact of high amplitudes on the perforation impedance. In the nonlinear regime, the resistance and reactance of the perforation impedance are shown to increase and decrease, respectively, with the acoustic velocity. The effect of grazing mean flow on the impedance was studied by Dean (1974), Rao and Munjal (1986), Cummings (1986), and Dickey *et al.* (2001). Grazing flow is generally observed to decrease reactance and increase resistance of the perforation impedance.

Studies have also been conducted on the acoustic impedance of perforations facing fibrous material in the absence of mean flow, which may be categorized as analytical, experimental, and semiempirical. The analytical work has considered the effect of fibrous material on the perforation impedance. For example, Bolt (1947) considered the mass reactance of perforations facing the absorbing material. He concluded that the addition of perforations to the absorbing material generally increased the absorption coefficient at low frequencies, while decreasing it at high frequencies. Later, Ingard and Bolt (1951) illustrated the effectiveness of perforations at low frequencies through an increase of the resistance. They concluded that a combination of absorbing material and perforations formed a Helmholtz resonator. Ingard (1954) emphasized the increase of resistance by the addition of perforations and the effect of an air gap between the perforations and absorbing material. His analysis showed that the air gap between the perforations and facing material nar-

[a)]Electronic mail: Selamet.1@osu.edu

rowed the effective frequency range of absorption coefficients. The experimental work on the acoustic characteristics of perforations combined with an absorbent is confined to determining the surface acoustic impedance or the absorption coefficient, rather than the perforation impedance. Having considered the perforations facing absorbing material as a Helmholtz resonator, one of the earlier works by Callaway and Ramer (1952) showed that the air gap between the perforations and absorbing material increased the absorption coefficient for a very low duct porosity (1.23%). Later, Davern (1977) experimentally illustrated the effect of porosity and wall thickness of the perforations, absorbing material density, air space, and the impervious layer between perforations and absorbing material on the absorption coefficient. The semiempirical studies first measured the perforation impedance in the absence of fibrous material, then empirically incorporated the acoustic properties of the material to account for their effect on the perforation impedance. Kirby and Cummings (1998) presented a semiempirical impedance formulation for the perforation backed by absorbing material in the presence of grazing mean flow. They concluded that the absorbing material increased the perforation impedance and the results were strongly dependent on the density of absorbing material immediately adjacent to perforations. They also argued that the interaction among holes could be neglected and thus the formulation for single hole could be used for a multihole perforation. Selamet *et al.* (2001) adapted the concept of Kirby and Cummings by modifying the formulation of Sullivan and Crocker (1978) for the perforations facing air-fibrous material. They illustrated the significant impact of perforation impedance on the transmission loss of dissipative silencers particularly with low porosity. The foregoing studies with the absorbent, however, are limited to developing expressions for the acoustic impedance of perforations by intuitive modifications of the available no-fiber relationships. Hence, the predictive studies on the acoustic performance of dissipative silencers have thus far used the perforation impedance with air, or its alternatives with such heuristic modifications.

The present study experimentally determines the acoustic impedance of perforations facing air-air (air on both sides) and air-fibrous material (air on one side and fibrous material on the other side) directly in the absence of mean flow. First, the measurement methodology developed in this study is applied to the perforations facing air-air, and the results are compared with the available literature. Then, the impedance of perforation in contact with air-fibrous material of $\rho_f = 100$ and 200 kg/m^3 is also experimentally acquired. Since accurate acoustic properties of the fibrous material are necessary for the measurement of acoustic impedance of perforations in contact with the material, the characteristic impedance and wave number of the material are also experimentally obtained in this work.

Following this Introduction, Sec. II presents the definitions of perforation impedance and existing formulations and Sec. III describes the experimental setups and data reduction. Section IV gives the results of acoustic properties of fibrous material and acoustic impedance of perforations followed by the concluding remarks in Sec. V.

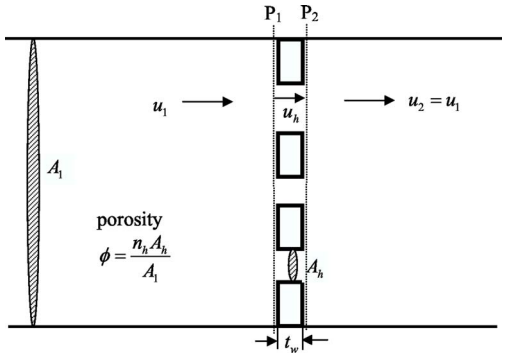


FIG. 1. A perforated plate in a duct.

II. DEFINITIONS OF PERFORATION IMPEDANCE AND EXISTING FORMULATIONS

A. Definitions of perforation impedance

The acoustic impedance is the ratio of the pressure difference across perforations to the particle velocity (Fig. 1). Two available definitions of the impedance differ by the way the particle velocity is specified. One uses the particle velocity u_h at the hole in Fig. 1, and the other the velocity u_1 or u_2 in the duct. These definitions and the relationship in between are introduced next.

The specific acoustic impedance of a hole is defined as

$$Z_h = \frac{p_1 - p_2}{u_h} = R_s + i\omega L_s = R_s + i\omega\rho_0 \ell_{\text{eff}} = R_s + ik_0\rho_0 c_0 \ell_{\text{eff}}, \tag{1}$$

where p_1 and p_2 are the acoustic pressures before and after the perforated plate, respectively, as shown in Fig. 1; ρ_0 the air density; c_0 the speed of sound in air; k_0 the wave number in air; R_s the specific resistance; L_s the specific inertance; and

$$\ell_{\text{eff}} = t_w + \alpha d_h \tag{2}$$

the effective length of the hole, where t_w is the wall thickness, d_h the hole diameter, and α the end correction coefficient. The coefficient α is dependent on the hole geometry and the characteristics of medium in contact with the hole. The nondimensional acoustic impedance of a hole is then given by

$$\zeta_h = \frac{p_1 - p_2}{Z_0 u_h} = R + ik_0(t_w + \alpha d_h), \tag{3}$$

where $Z_0 = \rho_0 c_0$ is the characteristic impedance of air and R is the dimensionless resistance.

The nondimensional acoustic impedance of a perforated plate can be defined as

$$\zeta_p = \frac{p_1 - p_2}{Z_0 u_1}, \tag{4}$$

where the particle velocity u_1 in the duct is related to u_h through conservation of mass by

$$u_1 = \frac{n_h A_h}{A_1} u_h = \phi u_h, \tag{5}$$

n_h being the number of holes, A_h the area of a hole, and ϕ the porosity. Equations (3)–(5) yield the relationships between the impedance of a hole and a plate as

$$\zeta_p = \frac{p_1 - p_2}{Z_0 u_h \phi} = \frac{\zeta_h}{\phi} = \frac{R + ik_0(t_w + \alpha d_h)}{\phi}. \tag{6}$$

While the impedance of a hole is representative of the acoustic characteristics of an isolated single perforation, it is the impedance of a plate typically used in the analyses, for example, in transmission loss predictions. The impedance of a plate ζ_p is usually measured first, the impedance of a hole ζ_h is then calculated using Eq. (6).

B. Existing formulations

The empirical relationship of Sullivan and Crocker (1978) given by

$$\zeta_p = \frac{0.006 + ik_0(t_w + 0.75 d_h)}{\phi} \tag{7}$$

has been widely used for the perforation impedance with air-air contact. This expression was obtained using a perforation with ϕ=4.2%. In Eq. (7), the coefficient 0.75 is associated with the end correction of the perforation. While the theoretical maximum value is $8/3\pi(\cong 0.85)$ for a single hole with an infinite flange, the coefficient may be smaller for multiholes due to the interaction among holes. Melling (1973) graphically illustrated such interaction between two holes.

Kirby and Cummings (1998) suggested a semiempirical acoustic impedance $\tilde{\zeta}_p$ for perforated plates facing air-fibrous material by modifying ζ_p that faces air-air as

$$\tilde{\zeta}_p = \zeta_p + \frac{ik_0 0.425\left(-1 + \frac{\tilde{Z}}{Z_0}\frac{\tilde{k}}{k_0}\right) d_h}{\phi}. \tag{8}$$

In Eq. (8), half of the end correction for a perforation facing air-air (the first term inside the parentheses) is combined with a second term to account for the effect of the fibrous material. However, the use of Eq. (8) for dissipative silencers leads to erroneous transmission loss predictions possibly due, according to Kirby (2001), to an overestimated perforation impedance. Such overestimation may be attributed to the ignored interaction among holes. Selamet *et al.* (2001) adapted the concept by applying Eq. (7) to the perforations facing air-fibrous material as follows:

$$\tilde{\zeta}_p = \frac{0.006 + ik_0\left[t_w + \frac{0.75}{2}\left(1 + \frac{\tilde{Z}}{Z_0}\frac{\tilde{k}}{k_0}\right) d_h\right]}{\phi}. \tag{9}$$

In Eq. (9), the effect of absorbent is considered with the complex characteristic impedance $\tilde{Z}$ and wave-number $\tilde{k}$ of the material, thus both real and imaginary parts of the impedance can be modified in Eq. (9). Since it is obtained by modifying Eq. (7) based on ϕ=4.2%, Eq. (9) is expected to be valid near this particular porosity.

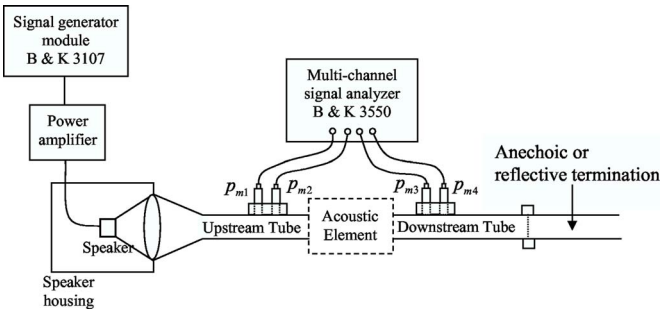


FIG. 2. Schematic of the impedance tube setup in the absence of mean flow.

III. EXPERIMENTAL SETUP AND DATA REDUCTION

An impedance tube setup shown in Fig. 2 (Refer, for example, to Dickey *et al.*, 2000 or Selamet *et al.*, 1994, for details) is modified for the measurements of characteristic impedance and wave number of the absorbing material, and the acoustic impedance of perforations in the absence of mean flow, particularly in contact with the material. The resulting setups and the data reduction for such measurements are described next.

A. Measurement of characteristic impedance and wave number of fibrous material

A transfer function method with two different termination conditions (Fig. 3) in conjunction with the impedance tube of Fig. 2 is developed to acquire experimentally the complex characteristic impedance and wave number of the fibrous material in the absence of mean flow. A sample of absorbing material is mounted inside the main duct bounded by wire screens with a porosity higher than 75% defining the locations a and b in Fig. 3. Transfer functions among the microphones and two measurements with distinctly different terminations are required in the present method to determine the complex characteristic impedance and wave number of the absorbing material in the frequency domain. Anechoic and fully reflective boundaries are applied as two different termination conditions.

The transfer matrix, which expresses the relationships between acoustic pressure and velocities at the surfaces a and b of the sample, is given for the first termination in terms of the complex characteristic impedance $\tilde{Z}$, the wave-number $\tilde{k}$, and the length of the sample ℓ_f as

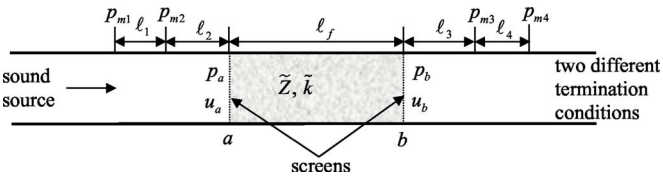


FIG. 3. Schematic of the experimental setup for obtaining the acoustic properties of absorbing material in the absence of mean flow.

$$\begin{bmatrix} p_a \\ u_a \end{bmatrix} = \begin{bmatrix} \cos(\tilde{k}\ell_f) & i\tilde{Z}\sin(\tilde{k}\ell_f) \\ i\dfrac{\sin(\tilde{k}\ell_f)}{\tilde{Z}} & \cos(\tilde{k}\ell_f) \end{bmatrix} \begin{bmatrix} p_b \\ u_b \end{bmatrix}, \tag{10}$$

and for the second termination (designated by prime)

$$\begin{bmatrix} p'_a \\ u'_a \end{bmatrix} = \begin{bmatrix} \cos(\tilde{k}\ell_f) & i\tilde{Z}\sin(\tilde{k}\ell_f) \\ i\dfrac{\sin(\tilde{k}\ell_f)}{\tilde{Z}} & \cos(\tilde{k}\ell_f) \end{bmatrix} \begin{bmatrix} p'_b \\ u'_b \end{bmatrix}. \tag{11}$$

Combining Eqs. (10) and (11) and rearranging yields

$$\begin{bmatrix} \cos(\tilde{k}\ell_f) \\ i\tilde{Z}\sin(\tilde{k}\ell_f) \\ i\dfrac{\sin(\tilde{k}\ell_f)}{\tilde{Z}} \\ \cos(\tilde{k}\ell_f) \end{bmatrix} = \begin{bmatrix} p_b & u_b & 0 & 0 \\ 0 & 0 & p_b & u_b \\ p'_b & u'_b & 0 & 0 \\ 0 & 0 & p'_b & u'_b \end{bmatrix}^{-1} \begin{bmatrix} p_a \\ u_a \\ p'_a \\ u'_a \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix}, \tag{12}$$

where $T_1 = T_4$ is noted. In Eq. (12), acoustic pressures (p_a, p_b, p'_a, p'_b) and velocities (u_a, u_b, u'_a, u'_b) at the sample surfaces a and b can be calculated from the pressure measurements at four microphones $(p_{m1}, p_{m2}, p_{m3}, p_{m4}, p'_{m1}, p'_{m2}, p'_{m3}, p'_{m4})$ in Fig. 3. The transfer matrix between the sample surface a and microphone 2 [for a straight pipe of constant cross-sectional area (Munjal, 1987)] is given by, for the first termination

$$\begin{bmatrix} p_a \\ u_a \end{bmatrix} = \begin{bmatrix} \cos(k_0\ell_2) & iZ_0\sin(k_0\ell_2) \\ i\dfrac{1}{Z_0}\sin(k_0\ell_2) & \cos(k_0\ell_2) \end{bmatrix}^{-1} \begin{bmatrix} p_{m2} \\ u_{m2} \end{bmatrix}. \tag{13}$$

In Eq. (13), the acoustic velocity u_{m2} at microphone 2 can be expressed in terms of pressure measurements, p_{m1} and p_{m2} at microphones 1 and 2. In light of

$$\begin{bmatrix} p_{m2} \\ u_{m2} \end{bmatrix} = \begin{bmatrix} \cos(k_0\ell_1) & -iZ_0\sin(k_0\ell_1) \\ -i\dfrac{1}{Z_0}\sin(k_0\ell_1) & \cos(k_0\ell_1) \end{bmatrix} \begin{bmatrix} p_{m1} \\ u_{m1} \end{bmatrix}, \tag{14}$$

u_{m2} is given by

$$u_{m2} = \frac{p_{m1} - p_{m2}\cos(k_0\ell_1)}{iZ_0\sin(k_0\ell_1)}. \tag{15}$$

Substitution of Eq. (15) into Eq. (13) yields

$$\begin{bmatrix} p_a \\ u_a \end{bmatrix} = \begin{bmatrix} \cos(k_0\ell_2) & iZ_0\sin(k_0\ell_2) \\ i\dfrac{1}{Z_0}\sin(k_0\ell_2) & \cos(k_0\ell_2) \end{bmatrix}^{-1} \times \begin{bmatrix} p_{m2} \\ \dfrac{p_{m1} - p_{m2}\cos(k_0\ell_1)}{iZ_0\sin(k_0\ell_1)} \end{bmatrix}, \tag{16}$$

which expresses the acoustic pressure and velocity at the sample surface a in terms of pressures alone measured at microphones 1 and 2. Similarly, the acoustic pressure and velocity at the sample surface a for the second termination is given by

$$\begin{bmatrix} p'_a \\ u'_a \end{bmatrix} = \begin{bmatrix} \cos(k_0\ell_2) & iZ_0\sin(k_0\ell_2) \\ i\dfrac{1}{Z_0}\sin(k\ell_2) & \cos(k_0\ell_2) \end{bmatrix}^{-1} \times \begin{bmatrix} p'_{m2} \\ \dfrac{p'_{m1} - p'_{m2}\cos(k_0\ell_1)}{iZ_0\sin(k_0\ell_1)} \end{bmatrix}. \tag{17}$$

The acoustic pressure and velocity at sample surface b can be similarly expressed in terms of pressures measured at microphones 3 and 4 for the first termination as

$$\begin{bmatrix} p_b \\ u_b \end{bmatrix} = \begin{bmatrix} \cos(k_0\ell_3) & iZ_0\sin(k_0\ell_3) \\ i\dfrac{1}{Z_0}\sin(k_0\ell_3) & \cos(k_0\ell_3) \end{bmatrix} \times \begin{bmatrix} p_{m3} \\ \dfrac{p_{m3}\cos(k_0\ell_4) - p_{m4}}{iZ_0\sin(k_0\ell_4)} \end{bmatrix}, \tag{18}$$

and for the second termination

$$\begin{bmatrix} p'_b \\ u'_b \end{bmatrix} = \begin{bmatrix} \cos(k_0\ell_3) & iZ_0\sin(k_0\ell_3) \\ i\dfrac{1}{Z_0}\sin(k_0\ell_3) & \cos(k_0\ell_3) \end{bmatrix} \times \begin{bmatrix} p'_{m3} \\ \dfrac{p'_{m3}\cos(k_0\ell_4) - p'_{m4}}{iZ_0\sin(k_0\ell_4)} \end{bmatrix}. \tag{19}$$

The complex characteristic impedance and wave number of the sample can then be obtained from Eq. (12) and Eqs. (16)–(19). The complex characteristic impedance $\tilde{Z}$ is given, in terms of T_2 and T_3 of Eq. (12), by

$$\tilde{Z} = \sqrt{\frac{T_2}{T_3}} \tag{20}$$

and the complex wave number, in terms of T_1 or T_4 of Eq. (12), by

$$\tilde{k} = \frac{1}{\ell_f}\cos^{-1}(T_1) = \frac{1}{\ell_f}\cos^{-1}(T_4). \tag{21}$$

Measured T_1 and T_4 of Eq. (12) are expected to be identical. However, that may not be exactly true in experiments due,

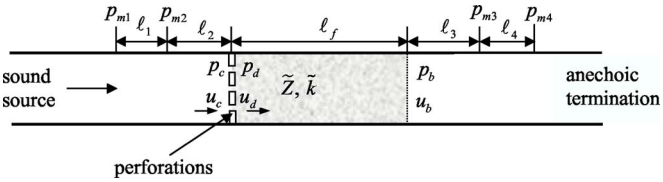


FIG. 4. Schematic of the experimental setup for obtaining the acoustic impedance of perforations in the absence of mean flow.

for example, to inhomogeneity in the filling and imperfect sample surfaces. Thus, the complex wave number is obtained in the present study using an average of T_1 and T_4 as

$$\tilde{k} = \frac{1}{\ell_f} \cos^{-1}\left(\frac{T_1 + T_4}{2}\right). \tag{22}$$

B. Measurement of acoustic impedance of perforations

Once the acoustic properties of the absorbing material are obtained, the acoustic impedance of perforations in contact with the absorbent on one side can be also determined experimentally. Unlike the acoustic properties of the absorbing material, only a single experiment with one termination is adequate for the measurement of perforation impedance. An anechoic termination is applied to reduce the interferences at microphones by the reflected wave from the termination. In view of Fig. 4, replacing u_1 of Eq. (4) by u_c gives the acoustic impedance of a perforated plate as

$$\tilde{\zeta}_p = \frac{p_c - p_d}{Z_0 u_c}, \tag{23}$$

where p_c and p_d are the acoustic pressures at the perforation surfaces contacting air and absorbing material, respectively, and u_c is the acoustic velocity in the duct at the perforation surface. The quantities p_c and u_c are determined in terms of pressures measured at microphones 1 and 2 from Eq. (16) as

$$\begin{bmatrix} p_c \\ u_c \end{bmatrix} = \begin{bmatrix} \cos(k_0\ell_2) & iZ_0\sin(k_0\ell_2) \\ i\frac{1}{Z_0}\sin(k_0\ell_2) & \cos(k_0\ell_2) \end{bmatrix}^{-1} \times \begin{bmatrix} p_{m2} \\ \frac{p_{m1} - p_{m2}\cos(k_0\ell_1)}{iZ_0\sin(k_0\ell_1)} \end{bmatrix}. \tag{24}$$

The acoustic pressure and velocity at the perforation surface in contact with the fibrous material are expressed in terms of a series of transfer matrices and the pressures measured at microphones 3 and 4 as

$$\begin{bmatrix} p_d \\ u_d \end{bmatrix} = [T_{ab}][T_{b3}] \begin{bmatrix} p_{m3} \\ \frac{p_{m3}\cos(k_0\ell_4) - p_{m4}}{iZ_0\sin(k_0\ell_4)} \end{bmatrix}, \tag{25}$$

where

$$[T_{ab}] = \begin{bmatrix} \cos(\tilde{k}\ell_f) & i\tilde{Z}\sin(\tilde{k}\ell_f) \\ i\frac{\sin(\tilde{k}\ell_f)}{\tilde{Z}} & \cos(\tilde{k}\ell_f) \end{bmatrix} \tag{26}$$

is the transfer matrix between surfaces a and b, and

$$[T_{b3}] = \begin{bmatrix} \cos(k_0\ell_3) & iZ_0\sin(k_0\ell_3) \\ i\frac{1}{Z_0}\sin(k_0\ell_3) & \cos(k_0\ell_3) \end{bmatrix} \tag{27}$$

is the transfer matrix between surface b and microphone 3. Finally, the perforation impedance can be determined from Eqs. (23)–(27) and sound pressure measurements at four microphones. In Eq. (23), experimentally determined acoustic velocities u_c and u_d at the perforation surface may not be exactly identical in experiments. Thus, an average of the two velocities is used for the calculation of perforation impedance as

$$\tilde{\zeta}_p = \frac{p_c - p_d}{Z_0\frac{(u_c + u_d)}{2}}. \tag{28}$$

The same setup and the data reduction technique are also used for the measurements of acoustic impedance of perforations without the fibrous material.

IV. ACOUSTIC PROPERTIES OF FIBROUS MATERIAL

A modified impedance tube setup described in the preceding section is used to experimentally obtain the complex characteristic impedance and wave number of fibrous material. The present study focuses on two filling densities ($\rho_f = 100$ and 200 kg/m^3 representing approximate bounds in practical applications), and experimentally determines their acoustic properties. These properties are then directly applied to the measurements of acoustic impedance of perforation in contact with the fibrous material. The precise measurements of the distances between the sample surfaces and microphones or among microphones are critical in the experiments. The accuracy within 0.1 mm for the materials with well-defined surfaces is recommended by ASTM 1050-98 (1998). However, such a precise physical measurement may not be possible due, for example, to the difficulty in finding the centers of the microphones with their diameter of 6.35 mm (1/4 in.). Thus, the distances ℓ_1 through ℓ_4 and the sample length ℓ_f in Figs. 3 and 4 are acoustically estimated using quarter wave resonances (Katz, 2000). The estimated distances by this method are as follows: $\ell_1 = \ell_4 = 3.55$ cm, $\ell_2 = 12.81$ cm, $\ell_3 = 12.71$ cm, and $\ell_f = 10.16$ cm. When ℓ_f is too short, the impact of inhomogeneous filling tends to magnify. On the other hand, too long of a sample is detrimental to transferring sufficient sound energy to the downstream microphones due to wave dissipation through the material, particularly at high frequencies. Therefore, ℓ_f was chosen to balance these competing factors. An accurate

estimation of speed of sound is also important for the measurements, which is calculated here by

$$c_0 = 20.047\sqrt{273.15 + T_C} \quad [\mathrm{m/s}], \tag{29}$$

as suggested by ASTM 1050-98 (1998). All experiments were performed at the room temperature, $T_C = 19 \pm 1\ ^\circ\mathrm{C}$.

A. Fiber samples

Advantex® continuous-strand glass fiber developed by Owens Corning (Novi, MI) is used in the experiments. The texturization process separates filament roving strands of fiber glass into individual filaments by turbulent air flow (Silentex™ process). The average diameter of the individual filaments in the roving strand is 24 μm. In addition to the acoustic properties, high thermal and corrosion resistance of the Advantex® glass fiber may also be desirable particularly for the exhaust systems. The physical and chemical properties of various absorptive materials and their relative durability in automotive silencer operations have been described by Beranek (1988) and Huff (2001).

Absorbent samples, 10 cm long and 4.9 cm in diameter and with two different filling densities $\rho_f = 100$ and 200 kg/m^3, are placed in the test tube (Fig. 2). Densities significantly lower than $\rho_f = 100$ kg/m^3 may cause the movement of fibers inside the silencers and higher than $\rho_f = 200$ kg/m^3 may be undesirable in terms of cost, weight, and in some cases acoustic attenuation.

Random filling is applied to achieve overall isotropic and homogeneous conditions. However, partially or locally nonisotropic or inhomogeneous regions may exist due, for example, to the hand filling of the absorbent into the test tube. Thus, experiments with five different samples for each filling density are performed to examine the effect of variations in filling conditions. An average of five individual experiments is then used as the final result. The present study uses a well-texturized condition, which has mostly individual filaments. The front surfaces of the absorbent sample also need to be well defined in the experiments. The surfaces of the sample perpendicular to the direction of wave propagation are identified by two screens. The screens have a high porosity not to influence the surface properties of the absorbent and sufficiently high tension to hold the absorbent flat at the surface.

B. Characteristic impedance and wave number

The normalized empirical complex characteristic impedance $\tilde{Z}$ and the wave-number $\tilde{k}$ of Delany and Bazley (1970) for fibrous materials in mks units

$$\frac{\tilde{Z}}{Z_0} = [1 + 0.0511(f/R_f)^{-0.75}] + i[-0.0768(f/R_f)^{-0.73}] \tag{30}$$

TABLE I. Fit coefficients of the measured characteristic impedance and wave number.

Filling density (kg/m^3)	a_1	b_1	a_2	b_2	a_3	b_3	a_4	b_4
100	33.20	0.7523	28.32	0.6512	39.20	0.6841	38.39	0.6285
200	25.69	0.5523	71.97	0.7072	56.03	0.6304	62.05	0.5980

$$\frac{\tilde{k}}{k_0} = [1 + 0.0858(f/R_f)^{-0.70}] + i[-0.1749(f/R_f)^{-0.59}] \tag{31}$$

have been widely used due to its simplicity and accuracy. Here f is the frequency in hertz and R_f the flow resistivity which is primarily dependent on the fiber diameter, layer orientation, and filling density. Although the Delany and Bazley formulation is well representative of the acoustic behavior of fibrous material in general, the present study experimentally determines the complex characteristic impedance and wave number of specific fibrous material since the process of data reduction for perforation impedance requires accurate (rather than simply somewhat representative) acoustic properties of the material in contact with the perforations. The present study has retained the forms of Eqs. (30) and (31), since they provide suitable fits. The curve-fitted complex characteristic impedance and wave number based on the averaged values of five experiments are expressed by

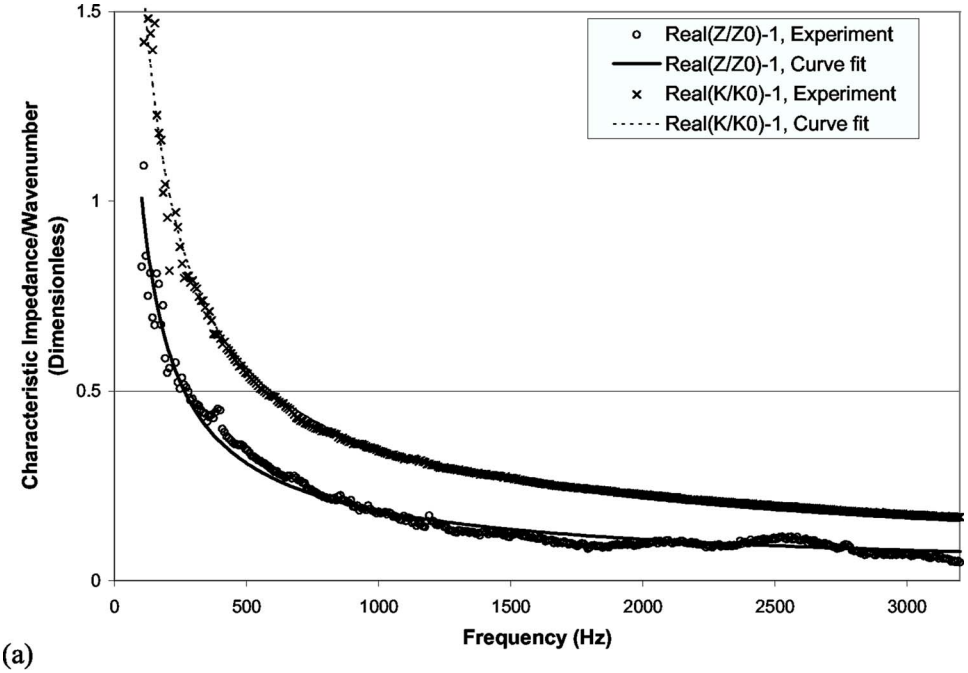
$$\frac{\tilde{Z}}{Z_0} = (1 + a_1 f^{-b_1}) - i a_2 f^{-b_2}, \tag{32}$$

$$\frac{\tilde{k}}{k_0} = (1 + a_3 f^{-b_3}) - i a_4 f^{-b_4}, \tag{33}$$

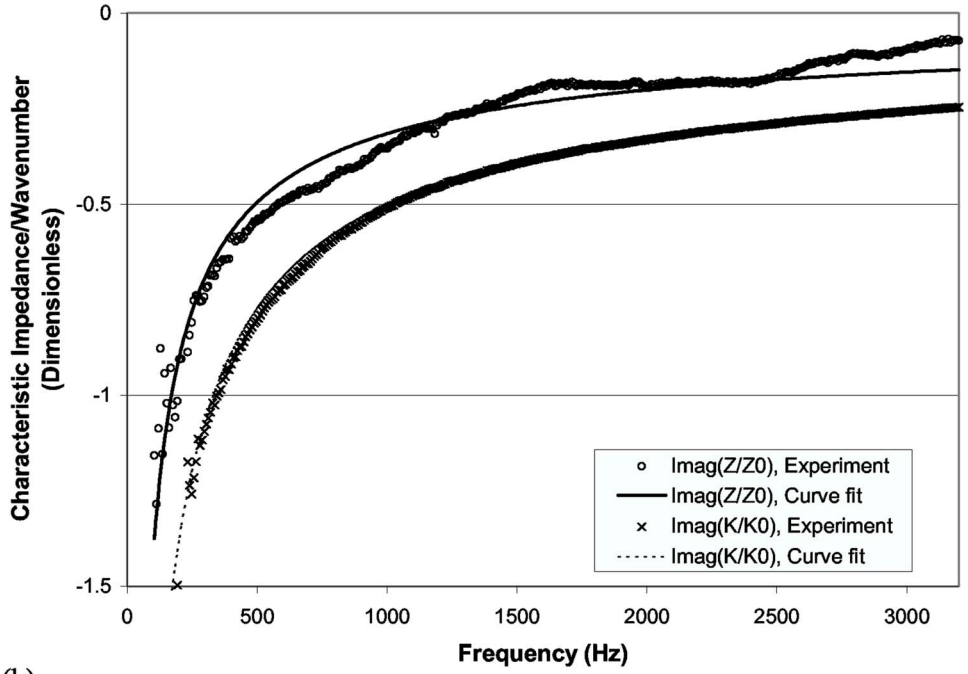
with the coefficients presented in Table I for both filling densities of $\rho_f = 100$ and 200 kg/m^3. The function "fit" of MATLAB® is applied for all curve fits in the frequency range of 100–3200 Hz. At lower frequencies, two distinct termination conditions may not be achieved, which are necessary for accurate results, due to the imperfect anechoic boundary. The accuracy of the measurement at low frequency is also limited by the microphone spacing (ASTM 1050-98, 1998). Figure 5 shows, for example, the characteristic impedance and wave number for $\rho_f = 100$ kg/m^3. The curve fits represent most of the data points well, particularly for the wave number.

V. ACOUSTIC IMPEDANCE OF PERFORATIONS

The acoustic impedance of perforations in contact with air-fibrous material is different than the one facing air-air due to the influence of absorbing material. The fiber filling density and texturization condition in contact with the perforation are particularly important. Ingard (1954) claims that the fiber properties only within a distance of one hole diameter are significant. Thus, the air gap between the perforation and the fibrous material may also alter the perforation impedance. The present study considers only the perforations directly in contact with the fibrous material without the air gap.



(a)



(b)

FIG. 5. Characteristic impedance and wave number of the fibrous material with $\rho_f = 100\,\mathrm{kg/m^3}$; (a) real part and (b) imaginary part.

Acoustic impedance of perforations facing air only and air-fibrous material is measured in the present study using the impedance tube presented in Fig. 4 and methodology described in Sec. II. The averaged results from five experiments with different fibrous material samples are utilized in the data reduction.

A. Perforation plate samples

Table II presents 11 samples of circular plates considered in this study with different porosity ϕ, wall thickness t_w, and hole diameter d_h to investigate the effect of such parameters on the perforation impedance. Figures 6 and 7 provide, respectively, the detailed dimensions and the photos of the circular perforated plates used in the present study. Perfora-

TABLE II. Perforation samples.

Sample No.	Hole diameter, d_h (cm)	Wall thickness, t_w (cm)	Porosity, ϕ (%)
A1	0.249	0.08	2.1
A2			8.4
A3			13.6
A4			25.2
B1		0.16	2.1
B2			8.4
B3			13.6
B4			25.2
C2	0.498	0.08	8.4
C3			13.6
C4			25.2

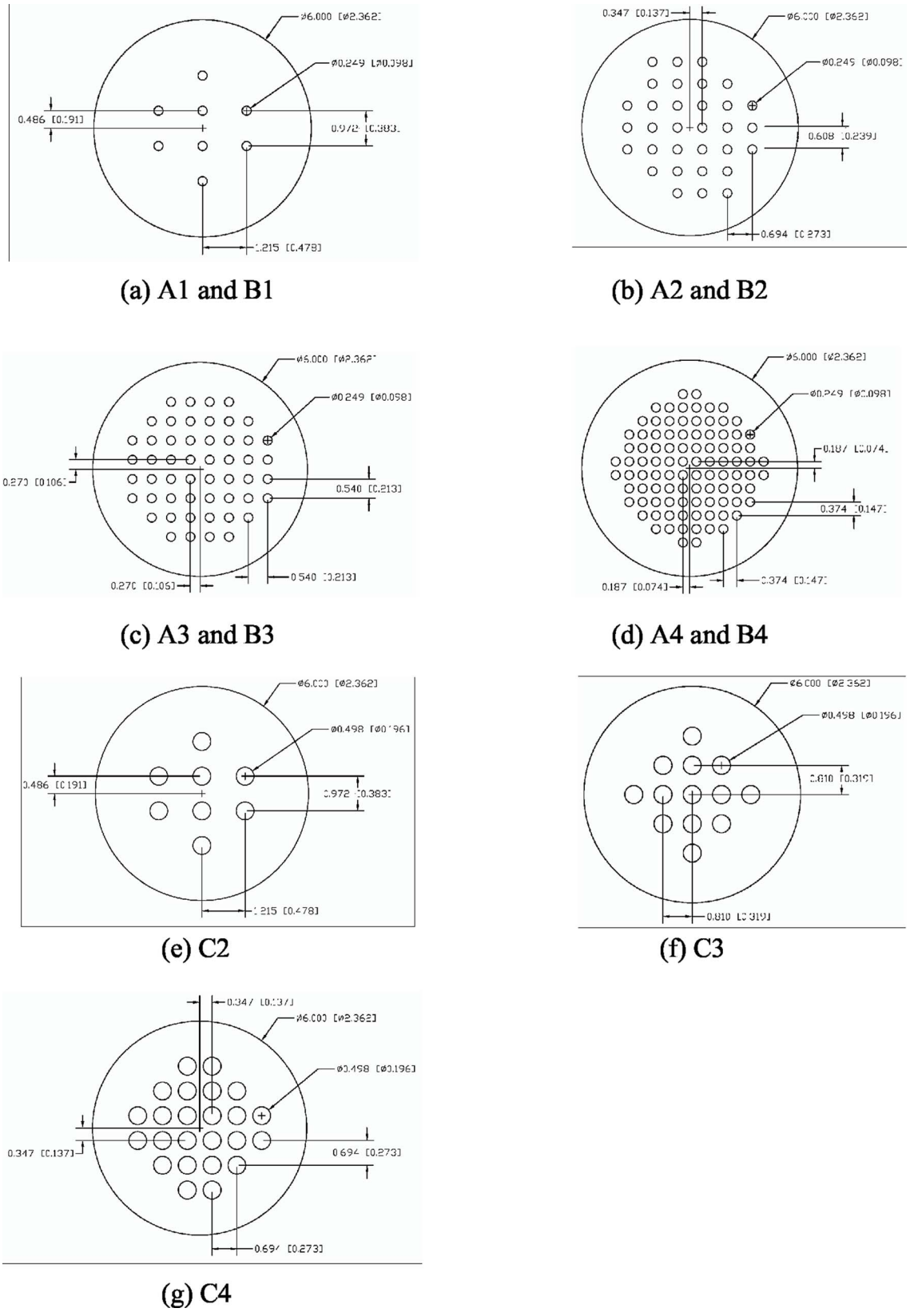


FIG. 6. Dimensions of perforated plate samples (unit: centimeters [inch]).

tions facing air-air with higher than 10% porosity have not been, in general, considered in the literature since the effect of varying perforations at high porosities on the transmission loss of perforated silencers is insignificant. However, with the fibrous material, the transmission loss may be affected by the perforation impedance even at higher porosities. Thus, four porosities (ϕ=2.1, 8.4, 13.6, and 25.2 %) are used in the present study, defined by the area ratio of all holes to the plate. The samples are categorized into three groups (Table II); (1) group A is the base with d_h=0.249 cm and t_w

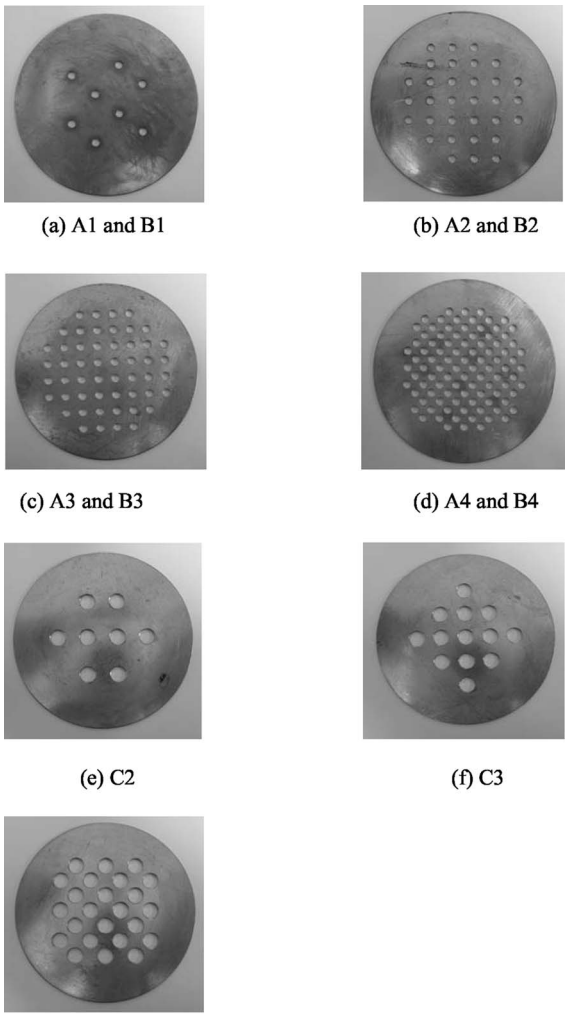


FIG. 7. Pictures of perforated plate samples.

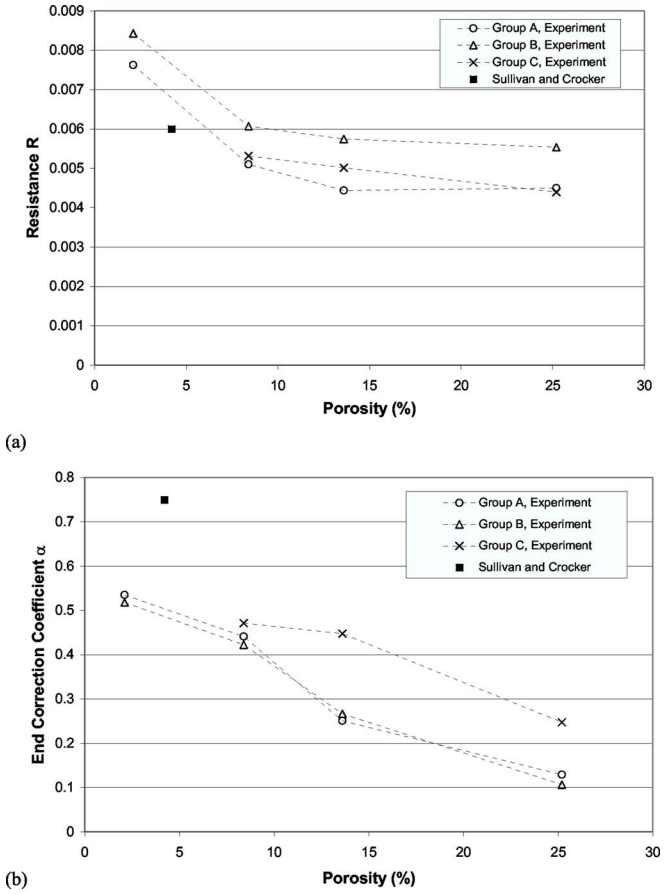


FIG. 8. Comparison of measurements of the acoustic impedance of a hole in contact with air-air; (a) resistance and (b) end correction coefficient.

=0.08 cm; (2) group B has the same hole diameter as group A and doubles the wall thickness to t_w=0.16 cm; and (3) group C has the same wall thickness as group A, while doubling the hole diameter to d_h=0.498 cm. The numbers (1–4) after the group identification by (A–C) in Table II designate different duct porosities. The holes on the plates are distributed to achieve similar distances between the holes in the vertical and horizontal directions. For impedance measurements, the circular sample plates are installed in the impedance tube of 4.9 cm inner diameter, as shown in Fig. 4. The distance from the tube wall to the nearest hole is chosen to be longer than the distance between holes in order to minimize the effect of the interaction between holes and the wall.

B. Acoustic impedance of perforations in contact with air alone

The dimensionless resistance R and the end correction coefficient α in Eq. (6) of the perforation samples with air-air contact are experimentally determined using the impedance tube setup depicted in Fig. 4. While this setup is originally designed to measure the impedance of perforation facing fibrous material, it can also be utilized for the perforation in contact with air only. For each perforation plate, five experiments are performed, and their averaged values are then used to obtain R and α. The end correction coefficients are calculated from the curve-fitted reactance in the frequency range of 100–3200 Hz for most samples. The only exception is A1 and B1 with ϕ=2.1% which are curve fitted in the range of 100–2600 Hz due to the unstable signals at frequencies higher than 2600 Hz. The mean values of resistance in the same frequency range are also obtained.

Figure 8 presents the measured R and α (also given in Table III) for the perforation samples of Table II, facing air-air. Figure 8(a) shows that the resistance R decreases as the porosity increases, particularly from ϕ=2.1 to 8.4 %. Group B (t_w=0.16 cm) has higher resistance than group A (t_w =0.08 cm), due to the doubled wall thickness. The resistance 0.006 obtained by Sullivan and Crocker (1978) for ϕ =4.2% falls in between the range of ϕ=2.1 and 8.4 % in Fig. 8(a). Figure 8(b) illustrates the effect of porosity, wall thickness, and hole diameter on α, thus reactance of the perforation impedance. As the porosity increases, α decreases because of the shorter distance between the holes, hence a stronger interaction among the holes. Groups A and B have similar values of α indicating that the end correction coefficient is essentially independent of the wall thickness. Group C has higher values of α than those of group A presumably

TABLE III. Measured R and α of the perforations.

Sample No.	R			α		
	Air alone	$\rho_f=100$ (kg/m^3)	$\rho_f=200$ (kg/m^3)	Air alone	$\rho_f=100$ (kg/m^3)	$\rho_f=200$ (kg/m^3)
A1	0.007624	0.03054	0.04935	0.5350	0.6989	0.7769
A2	0.005101	0.02996	0.04575	0.4409	0.6471	0.7412
A3	0.004437	0.02298	0.03610	0.2506	0.4576	0.4999
A4	0.004500	0.02716	0.04447	0.1286	0.4717	0.4590
B1	0.008429	0.03313	0.05770	0.5179	0.6514	0.7663
B2	0.006074	0.02793	0.04900	0.4224	0.6026	0.6758
B3	0.005744	0.02444	0.04142	0.2666	0.4661	0.5048
B4	0.005539	0.02657	0.05076	0.1066	0.3951	0.3980
C2	0.005318	0.04728	0.09015	0.4707	0.6206	0.7167
C3	0.005013	0.04598	0.09308	0.4473	0.6269	0.7142
C4	0.004395	0.03973	0.07604	0.2471	0.4504	0.4926

due to the larger hole diameter resulting in a longer distance among the holes, thereby reducing the interaction. The end corrections obtained in the present study are lower than 0.75 of Sullivan and Crocker (1978) for all porosities, possibly due to different distance between holes.

C. Acoustic impedance of perforations in contact with fibrous material

The experimental results for R and α of the perforations facing an absorbent with $\rho_f=100$ and 200 kg/m^3 as well as air alone are shown in Figs. 9–11, for groups A, B, and C, respectively, and Table III. Figures 9–11 show that fibrous material substantially increases both resistance and end correction coefficients for all three groups. The resistance R of $\rho_f=200$ kg/m^3 are higher than those of lower filling density $\rho_f=100$ kg/m^3 for all samples. The effect of different filling density on α diminishes for the samples A4 and B4. The average increases of R over all 11 samples (recall Table II) due to the fibrous material are reflected by factors of nearly 5 and 10 relative to the baseline with no fiber for $\rho_f=100$ and 200 kg/m^3, respectively, whereas the average α is approximately only doubled for both fiber densities. Thus, the impact of fibrous material on the resistance is more significant than on the end correction coefficient. The imaginary part of the perforation impedance depends also on the frequency, the hole diameter, and the wall thickness [recall Eq. (3)].

For groups A (Fig. 9) and B (Fig. 10), as the porosity increases, α decreases for both filling densities except for one sample (A4 with $\phi=25.2\%$ and $\rho_f=100$ kg/m^3). For the same groups, R decreases with porosity up to $\phi=13.6\%$, and then increases to that of $\phi=25.2\%$. Group C (Fig. 11) with $d_h=0.498$ cm exhibits some differences from groups A and B with $d_h=0.249$ cm; R of $\phi=25.2\%$ is lower than that of $\phi=13.6\%$, and the change of α from $\phi=13.6$ to 25.2 % is more dramatic than from $\phi=8.4$ to 13.6 %.

Table IV presents relative standard deviations (σ_R/R and σ_α/α) for R and α calculated from five repeat experiments with air alone and with fibrous material. σ_R/R is usually higher than σ_α/α, and both deviations, in general, increase with the porosity. σ_R/R and σ_α/α for $\rho_f=200$ kg/m^3 are, in most cases, greater than those for $\rho_f=100$ kg/m^3, possibly due to relatively nonhomogeneous filling conditions, including texturization of the fiber, for the higher density. The texturization of the absorbent may substantially affect the acoustic impedance of perforations in contact with this material (Lee, 2005). The present study has provided the impedance of perforation in contact with well texturized fibrous material.

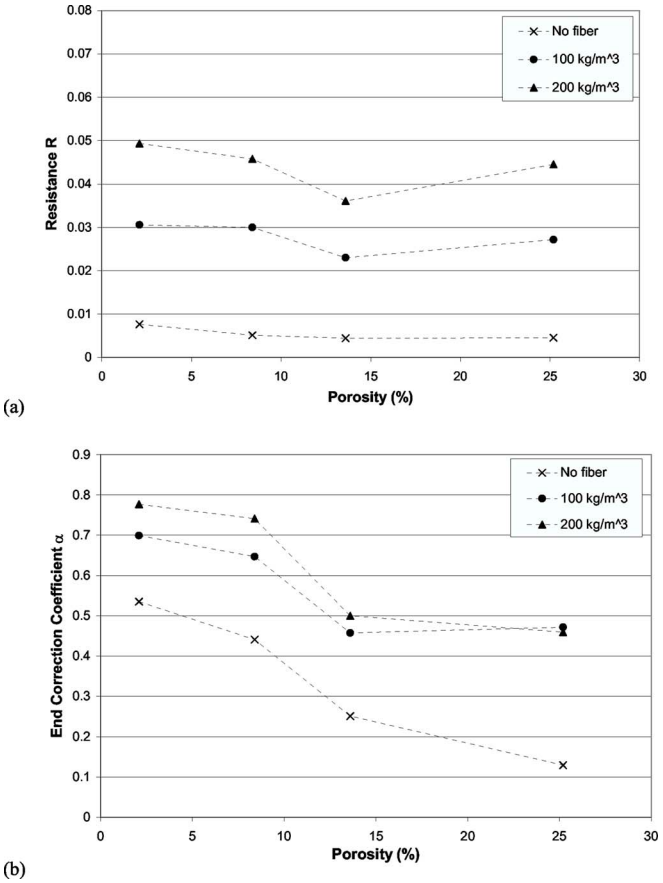
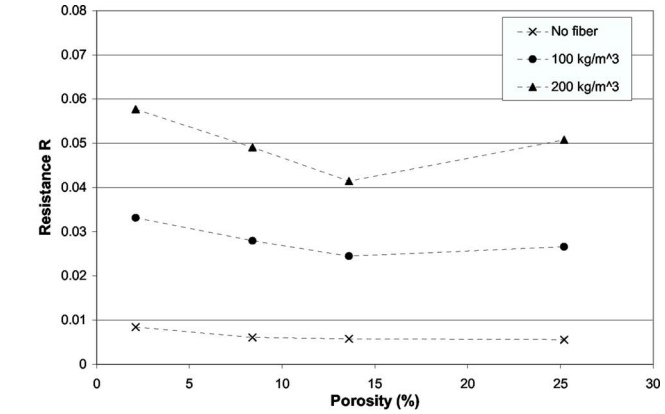
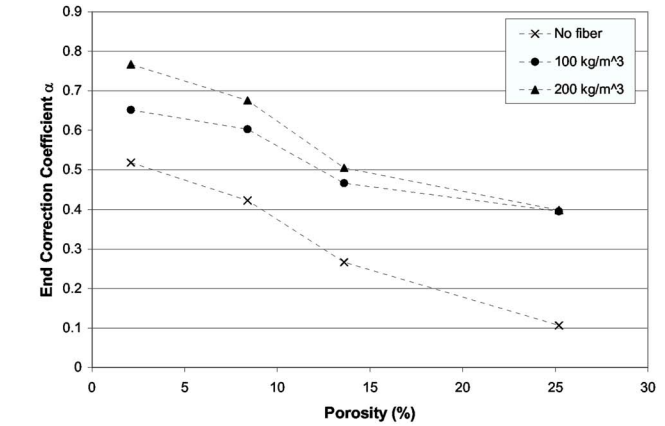


FIG. 9. Acoustic impedance of a hole, group A; (a) resistance and (b) end correction coefficient.

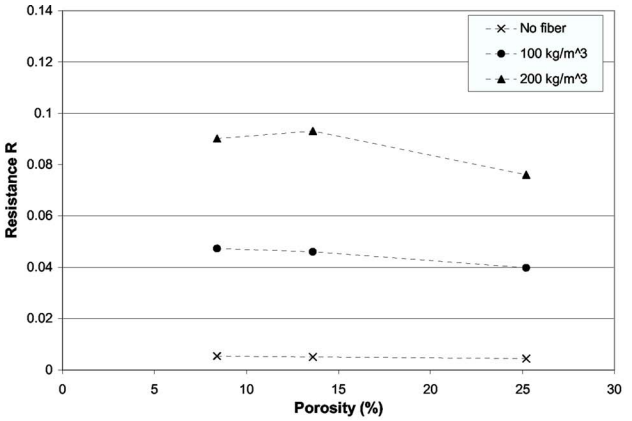


(a)

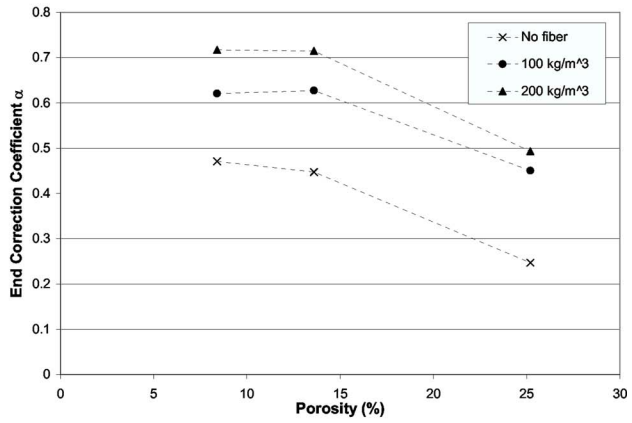


(b)

FIG. 10. Acoustic impedance of a hole, group B; (a) resistance and (b) end correction coefficient.



(a)



(b)

FIG. 11. Acoustic impedance of a hole, group C; (a) resistance and (b) end correction coefficient.

VI. CONCLUSIONS

The understanding of the acoustic behavior of the fibrous material and perforations is critical for the design of dissipative silencers. Therefore, the present study has developed experimental setups to measure: (a) the complex characteristic impedance and the complex wave number of the fibrous material, and (b) the acoustic impedance of perforations, particularly in the presence of absorbing material.

The characteristic impedance and wave number of the fibrous material with two filling densities (ρ_f=100 and 200 kg/m^3) are obtained using a modified impedance tube setup with four microphones and two different termination conditions. The averaged values from five samples are curve fitted as a function of frequency. These curve fits represent most of the data points well, particularly for the wave number at both filling densities. The characteristic impedance and wave number inferred from the experiments are then used for the measurement of acoustic impedance of perforations in contact with the fibrous material.

A new experimental setup has been developed in the present study for the measurement of acoustic impedance of perforations in contact with the fibrous material as well as with air alone in the absence of mean flow. The resistance R and the end correction coefficient α of 11 perforation samples with different porosities (ϕ=2.1, 8.4, 13.6, and 25.2 %), hole diameters (d_h=0.249 and 0.498 cm), and wall thickness (t_w=0.08 and 0.16 cm) are measured along with the fibrous material of ρ_f=100 and 200 kg/m^3. For the perforations without fibrous material, the resistance R and the end correction coefficient α decrease as the porosity increases. The fibrous material significantly increases both R and α compared to the perforation in contact with air only. For both filling densities, both R and α decrease as the porosity increases, except for some of the samples with highest porosity (ϕ=25.2%).

TABLE IV. Standard deviations relative to R and α (%).

Sample No.	$\frac{\sigma_R}{R}$ Air alone	$\frac{\sigma_R}{R}$ ρ_f=100 (kg/m^3)	$\frac{\sigma_R}{R}$ ρ_f=200 (kg/m^3)	$\frac{\sigma_\alpha}{\alpha}$ Air alone	$\frac{\sigma_\alpha}{\alpha}$ ρ_f=100 (kg/m^3)	$\frac{\sigma_\alpha}{\alpha}$ ρ_f=200 (kg/m^3)
A1	24.5	16.1	25.4	7.4	3.4	7.7
A2	10.3	15.7	16.0	4.9	3.7	6.9
A3	11.3	25.4	31.8	8.9	7.2	13.8
A4	17.5	44.2	38.2	30.8	10.9	25.5
B1	7.3	15.3	18.1	3.4	4.2	4.6
B2	6.9	17.1	23.0	4.2	4.1	7.5
B3	9.4	21.8	31.8	8.4	9.2	16.1
B4	15.3	25.8	36.1	35.2	15.9	36.9
C2	18.7	19.4	22.1	3.0	4.2	5.9
C3	16.6	18.3	26.9	3.5	5.1	6.3
C4	20.7	29.0	35.9	10.0	7.8	13.9

The results obtained in the present study can be directly used for the predictions of acoustic behavior of silencers that use similar fibrous material and perforation geometry. When the material and geometry deviate substantially than those of the current work, caution should be exercised in the use of perforation impedance.

Nomenclature

A_1 = cross-sectional area of the duct before the perforated plate
A_h = cross-sectional area of a hole
$a_{1,2,3,4}$ $b_{1,2,3,4}$ = coefficients of wave number and characteristic impedance of fibrous material
$\tilde{c}$ = complex speed of sound in the absorbing material
c_0 = speed of sound in air
d_h = perforate hole diameter
f = frequency
i = imaginary unit ($=\sqrt{-1}$)
$\tilde{k}$ = complex wave number in the absorbing material
k_0 = wave number in air
L_s = specific inertance
ℓ_1 = length between microphones 1 and 2
ℓ_2 = length between microphone 2 and sample upstream surface
ℓ_3 = length between microphone 3 and sample downstream surface
ℓ_4 = length between microphones 3 and 4
ℓ_{eff} = effective length of a hole
ℓ_f = length of a fiber sample
n_h = number of holes
p_1 = acoustic pressure before the perforate plate
p_2 = acoustic pressure after the perforate plate
p_a = acoustic pressure at the upstream of the fiber sample with the first termination condition
p'_a = acoustic pressure at the upstream of the fiber sample with the second termination condition
p_b = acoustic pressure at the downstream of the fiber sample with the first termination condition
p'_b = acoustic pressure at the downstream of the fiber sample with the second termination condition
p_c = acoustic pressure at the perforations contacting air
p_d = acoustic pressure at the perforations contacting fiber
p_{m1} = acoustic pressure at microphone 1 with the first termination condition
p_{m2} = acoustic pressure at microphone 2 with the first termination condition
p_{m3} = acoustic pressure at microphone 3 with the first termination condition
p_{m4} = acoustic pressure at microphone 4 with the first termination condition
p'_{m1} = acoustic pressure at microphone 1 with the second termination condition
p'_{m2} = acoustic pressure at microphone 2 with the second termination condition
p'_{m3} = acoustic pressure at microphone 3 with the second termination condition
p'_{m4} = acoustic pressure at microphone 4 with the second termination condition
R = dimensionless resistance of perforation impedance
R_f = flow resistivity of the absorbing material
R_s = specific resistance
t_w = wall thickness of perforated duct
$[T_{ab}]$ = transfer matrix of a fiber sample
$[T_{b3}]$ = transfer matrix between the downstream of a fiber sample and microphone 3
T_C = room temperature
$T_{1,2,3,4}$ = matrix elements
u_1 = particle velocity before the perforated plate
u_2 = particle velocity after the perforated plate
u_h = particle velocity at a hole
u_{m1} = particle velocity at microphone 1 with the first termination condition
u_{m2} = particle velocity at microphone 2 with the first termination condition
u_a = particle velocity at upstream of the fiber sample with the first termination condition
u'_a = particle velocity at upstream of the fiber sample with the second termination condition
u_b = particle velocity at downstream of the fiber sample with the first termination condition
u'_b = particle velocity at downstream of the fiber sample with the second termination condition
u_c = particle velocity at perforations contacting air
u_d = particle velocity at perforations contacting fiber
Z_0 = $\rho_0 c_0$, characteristic impedance of air
$\tilde{Z}$ = $\widetilde{\rho c}$, complex characteristic impedance of the absorbing material
Z_h = acoustic impedance of a hole
α = end correction coefficient of perforation impedance
σ_α = one standard deviation of α
σ_R = one standard deviation of R
ϕ = porosity
ρ_0 = density of air
ρ_f = filling or bulk density of fibrous material
$\tilde{\rho}$ = effective density of the absorbing material

ω = angular velocity
ζ_p = nondimensionalized acoustic impedance of perforated plate
$\tilde{\zeta}_p$ = nondimensionalized acoustic impedance of perforated plate in contact with fibrous material
ζ_h = nondimensionalized acoustic impedance of a hole
$\tilde{\zeta}_h$ = nondimensionalized acoustic impedance of a hole in contact with fibrous material

ASTM E1050-98 (**1998**). "Standard test method for impedance and absorption of acoustical materials using a tube, two microphones and a digital frequency analysis system" (American Society for Testing and Materials, Pennsylvania), 951–961.

Beranek, L. L. (**1988**). *Noise and Vibration Control* (Institute of Noise Control Engineering, Washington, DC).

Bolt, R. H. (**1947**). "On the design of perforated facings for acoustical materials," J. Acoust. Soc. Am. **19**, 917–921.

Callaway, D. B., and Ramer, L. G. (**1952**). "The use of perforated facings in designing low frequency resonant absorbers," J. Acoust. Soc. Am. **24**, 309–312.

Cummings, A. (**1986**). "The effects of grazing turbulent pipe-flow on the impedance of an orifice," Acustica **61**, 233–242.

Davern, W. A. (**1977**). "Perforated facings backed with porous materials as sound absorbers-an experimental study," Appl. Acoust. **10**, 85–112.

Dean, P. D. (**1974**). "An in situ method of wall acoustic impedance measurement in flow ducts," J. Sound Vib. **34**, 97–130.

Delany, M. E., and Bazley, E. N. (**1970**). "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116.

Dickey, N. S., Selamet, A., and Ciray, M. S. (**2001**). "An experimental study of the impedance of perforated plates with grazing flow," J. Acoust. Soc. Am. **110**, 2360–2370.

Dickey, N. S., Selamet, A., and Novak, J. M. (**2000**). "The effect of high-amplitude sound on the attenuation of perforated tube silencers," J. Acoust. Soc. Am. **108**, 1068–1081.

Huff, N. T. (**2001**). "Materials for absorptive silencer systems," *SAE Noise and Vibration Conference and Exposition*, April 30–May 3, SAE Paper No. 2001-01-1458 Traverse City, MI (Society of Automotive Engineers, Pennsylvania).

Ingard, U. (**1953**). "On the theory and design of acoustic resonators," J. Acoust. Soc. Am. **25**, 1037–1061.

Ingard, U. (**1954**). "Perforated facing and sound absorption," J. Acoust. Soc. Am. **26**, 151–154.

Ingard, U., and Bolt, R. H. (**1951**). "Absorption characteristics of acoustic material with perforated facings," J. Acoust. Soc. Am. **23**, 533–540.

Katz, B. F. G. (**2000**). "Method to resolve microphone and sample location errors in the two-microphone duct measurement method," J. Acoust. Soc. Am. **108**, 2231–2237.

Kirby, R. (**2001**). "Simplified techniques for predicting the transmission loss of a circular dissipative silencer," J. Sound Vib. **243**, 403–426.

Kirby, R., and Cummings, A. (**1998**). "The impedance of perforated plates subjected to grazing gas flow and backed by porous media," J. Sound Vib. **217**, 619–636.

Lee, I.-J. (**2005**). "Acoustic characteristics of perforated dissipative and hybrid silencers," Ph.D. dissertation (The Ohio State University, Columbus, OH).

Melling, T. H. (**1973**). "The acoustic impedance of perforates at medium and high sound pressure levels," J. Sound Vib. **29**, 1–65.

Munjal, M. L. (**1987**). *Acoustics of Ducts and Mufflers* (Wiley, New York).

Rao, K. N., and Munjal, M. L. (**1986**). "Experimental evaluation of impedance of perforates with grazing flow," J. Sound Vib. **108**, 283–295.

Selamet, A., Dickey, N. S., and Novak, J. M. (**1994**). "The Herschel-Quincke tube: A theoretical, computational, and experimental investigation," J. Acoust. Soc. Am. **96**, 3177–3185.

Selamet, A., Lee, I.-J., Ji, Z. L., and Huff, N. T. (**2001**). "Acoustic attenuation performance of perforated concentric absorbing silencers," *SAE Noise and Vibration Conference and Exposition*, April 30–May 3, Traverse City, MI, SAE Paper No. 2001-01-1435 (Society of Automotive Engineers, Pennsylvania).

Sullivan, J. W., and Crocker, M. J. (**1978**). "Analysis of concentric-tube resonators having unpartitioned cavities," J. Acoust. Soc. Am. **64**, 207–215.

Evaluation of a boss model and subtraction technique for predicting wideband scattering phenomena in room acoustics

Georgios Natsiopoulos[a] and Mendel Kleiner
Department of Civil and Environmental Engineering, Division of Applied Acoustics, Chalmers University of Technology, S-41296 Gothenburg, Sweden

(Received 6 October 2005; revised 20 February 2006; accepted 21 February 2006)

Scattering from a hard hemisphere on an infinite plane is studied experimentally and theoretically in a laboratory environment using a subtraction technique. With this technique, the scattered field is determined by subtracting the transient responses from the plane with and without the boss present. Scattering for different combinations of transducer and hemisphere positions are then classified into a number of cases, denoted in-plane/off-plane and back-/forward/mixed scattering. Theoretical and experimental results are compared and evaluated from an auralization point of view using a one-third octave band smoothing of the contribution from the hemisphere. With the smoothing used, average intensity level discrepancies typically less than 3 dB are obtained for a 6 octaves wide *ka*-interval, where *a* equals the hemisphere radius. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2188332]

PACS number(s): 43.55.Br, 43.20.Px, 43.20.Fn [NX] Pages: 2798–2803

I. INTRODUCTION

Acoustic scattering from isolated objects on planar surfaces, such as semicylindrical columns or hemispherical armatures on walls, is a common phenomenon in room acoustics. Therefore, fast and sufficiently accurate prediction models of this for auralization purposes are of great importance in the design, improvement, and analysis of auditoria.

Analytical solutions are rare[1,2] but generally preferred, especially since no spatial discretization is needed. In the middle of the last century, Twersky[3,4] developed analytical expressions for scattering from rigid semicylinders or hemispheres on a rigid infinite plane, where a boss model was used — a combination of image methods, separation of variables, and addition theorems for coupling phenomena. Later, Goodman and Stern[5] studied acoustic plane wave scattering from an elastic hemispherical shell on a rigid infinite plane, both analytically and numerically. They found that the external field was similar to that produced by a rigid hemisphere.

Comparisons with measurements were not made in the studies above, although they would elucidate several questions of practical importance. For example, possible turbulence effects at sharp edges will not be visible in the analytical solution, since they are derived by adding results for smooth spheres or cylinders in free space. In addition, theoretical results for hemispheres show strong focusing effects that may not be equally pronounced in reality. Moreover, the incident wave front is never perfectly plane or spherical, and this will also affect the agreement with situations encountered in practice.

An experiment with a hemispherical hardened plastic shell placed on a smooth concrete floor was carried out to investigate whether the boss model could be a candidate for applications in room acoustics. Microphone and loudspeaker were of a cylindrical shape and were pointed toward the hemisphere base center from distances $3a$ and $9a$, respectively, where $a=11.5$ cm is the hemisphere radius. Loudspeaker and microphone angles were varied in steps of 30° to include examples of in- and off-plane scattering, as well as back-, forward, and mixed scattering. Each measurement was made with and without the hemisphere present, making an estimate of the scattered part obtainable by subtraction. Details of all measurements have been documented.[6] Corresponding theoretical results were readily produced for comparison using the method of separation of variables[1] and the principle of superposition.[7]

II. THEORY

A. Scattering from a hard sphere

The problem has rotational symmetry about the axis that runs through the point source location $\boldsymbol{r}_0$ and the center of the sphere $\boldsymbol{r}_a$. Let $\boldsymbol{r}$ be the observation point. See Fig. 1. Note that

$$\cos\alpha = \frac{\boldsymbol{r}-\boldsymbol{r}_a}{|\boldsymbol{r}-\boldsymbol{r}_a|}\cdot\frac{\boldsymbol{r}_0-\boldsymbol{r}_a}{|\boldsymbol{r}_0-\boldsymbol{r}_a|}. \tag{1}$$

In the frequency domain, the exact analytical solution for a sphere present is written as a sum of the incident field $\hat{p}_{\mathrm{in}}$ and a scattered part $\hat{p}_{\mathrm{sc}}$, which satisfies the homogeneous Helmholtz equation

$$\hat{p}(\boldsymbol{r},\omega)=\hat{p}_{\mathrm{in}}(\boldsymbol{r},\omega)+\hat{p}_{\mathrm{sc}}(\boldsymbol{r},\omega), \tag{2}$$

$$\nabla^2\hat{p}_{\mathrm{in}}(\boldsymbol{r},\omega)+k^2\hat{p}_{\mathrm{in}}(\boldsymbol{r},\omega)=4\pi\delta(\boldsymbol{r}-\boldsymbol{r}_0)\hat{U}(\omega), \tag{3}$$

$$\nabla^2\hat{p}_{\mathrm{sc}}(\boldsymbol{r},\omega)+k^2\hat{p}_{\mathrm{sc}}(\boldsymbol{r},\omega)=0. \tag{4}$$

Here, δ is the Dirac delta function. The point source sends out a signal $U(t)$ and the Fourier transform pair used is

$$p(\boldsymbol{r},t)=\frac{1}{2\pi}\int_{\omega=-\infty}^{\infty}\hat{p}(\boldsymbol{r},\omega)e^{-i\omega t}d\omega, \tag{5}$$

[a]Electronic mail: georgios.natsiopoulos@chalmers.se

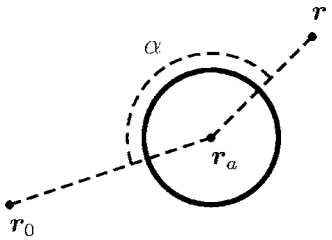


FIG. 1. Geometry for scattering from a sphere centered at $\boldsymbol{r}_a$. The source is located at $\boldsymbol{r}_0$ and the receiver at $\boldsymbol{r}$. By definition, $0^\circ \leqslant \alpha \leqslant 180^\circ$.

$$\hat{p}(\boldsymbol{r},\omega) = \int_{t=-\infty}^{\infty} p(\boldsymbol{r},t) e^{i\omega t} dt. \tag{6}$$

Neumann conditions demand that the normal derivative of the total field is zero at the boundary of the sphere:

$$[\boldsymbol{n}(\boldsymbol{r}) \cdot \nabla \hat{p}(\boldsymbol{r},\omega)]_{|\boldsymbol{r}-\boldsymbol{r}_a|=a} = 0. \tag{7}$$

Here, $\boldsymbol{n}$ is the normal vector of unit length. An explicit analytical solution exists[8] by separation of variables:

$$\hat{p}_{\mathrm{sc}}(\boldsymbol{r},\omega) = -ik\hat{U}(\omega) \sum_{m=0}^{\infty} Q_m(ka) h_m(k|\boldsymbol{r}-\boldsymbol{r}_a|) \times h_m(k|\boldsymbol{r}_0-\boldsymbol{r}_a|) P_m(\cos\alpha), \tag{8}$$

$$Q_m(ka) = (2m+1) \frac{m j_m(ka) - ka j_{m+1}(ka)}{m h_m(ka) - ka h_{m+1}(ka)}, \tag{9}$$

where P_m is the Legendre polynomial of degree m, j_m is the spherical Bessel function of order m, and h_m is the spherical Hankel function of the first kind and order m.

B. A hard hemisphere on a hard infinite plane

Let a hemisphere be situated on an infinite plane with unit normal $\boldsymbol{n}$ pointing into the room. Above the plane, the problem is mathematically equivalent to scattering from a sphere by two sources — the original source and a virtual mirror source.[3] See Fig. 2. The position of the minor source $\boldsymbol{r}_0^*$ is

$$\boldsymbol{r}_0^* = \boldsymbol{r}_0 - 2|(\boldsymbol{r}_0 - \boldsymbol{r}_a) \cdot \boldsymbol{n}|\boldsymbol{n}. \tag{10}$$

The exact analytical solution for a hemisphere present is

$$\hat{p}(\boldsymbol{r},\omega) = \hat{U}(\omega)\left[\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}_0|}}{|\boldsymbol{r}-\boldsymbol{r}_0|} + \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}_0^*|}}{|\boldsymbol{r}-\boldsymbol{r}_0^*|}\right] + \hat{p}_{\mathrm{sc}}(\boldsymbol{r},\omega) + \hat{p}_{\mathrm{sc}}^*(\boldsymbol{r},\omega), \tag{11}$$

where the sum of the scattered fields are

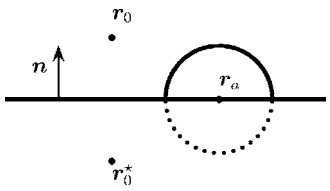


FIG. 2. The hemisphere with its base center at $\boldsymbol{r}_a$, the reflection plane with its unit normal vector $\boldsymbol{n}$, and the virtual mirror source $\boldsymbol{r}_0^*$.

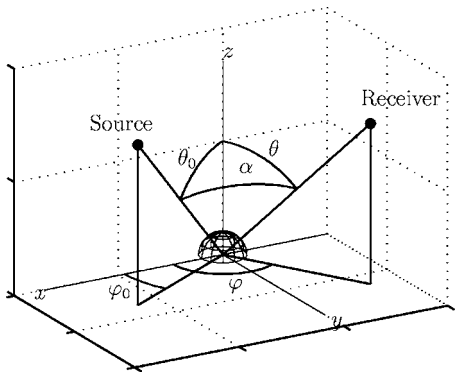


FIG. 3. Illustration that shows the coordinates used in specifying various measurement setups.

$$\hat{p}_{\mathrm{sc}}(\boldsymbol{r},\omega) + \hat{p}_{\mathrm{sc}}^*(\boldsymbol{r},\omega) = -ik\hat{U}(\omega) \sum_{m=0}^{\infty} Q_m(ka) \times h_m(k|\boldsymbol{r}-\boldsymbol{r}_a|) h_m(k|\boldsymbol{r}_0-\boldsymbol{r}_a|)[P_m(\cos\alpha) + P_m(\cos\alpha^*)], \tag{12}$$

since $|\boldsymbol{r}_0^*-\boldsymbol{r}_a| = |\boldsymbol{r}_0-\boldsymbol{r}_a|$. Here,

$$\cos\alpha^* = \frac{\boldsymbol{r}-\boldsymbol{r}_a}{|\boldsymbol{r}-\boldsymbol{r}_a|} \cdot \frac{\boldsymbol{r}_0^*-\boldsymbol{r}_a}{|\boldsymbol{r}_0^*-\boldsymbol{r}_a|}, \tag{13}$$

where α^* corresponds to the angle between the image source and the observation point, measured from the hemisphere base center. Compare with Eq. (1).

The principle of superposition can be used to take into account N point sources placed at $\boldsymbol{r}_n, n=1,2,\ldots,N$ instead of a single one at $\boldsymbol{r}_0$:

$$\hat{p}(\boldsymbol{r},\omega) = \sum_{n=1}^{N} \left[\hat{U}_n(\omega)\left\{\frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}_n|}}{|\boldsymbol{r}-\boldsymbol{r}_n|} + \frac{e^{ik|\boldsymbol{r}-\boldsymbol{r}_n^*|}}{|\boldsymbol{r}-\boldsymbol{r}_n^*|}\right\} + \hat{p}_{\mathrm{sc},n}(\boldsymbol{r},\omega) + \hat{p}_{\mathrm{sc},n}^*(\boldsymbol{r},\omega)\right]. \tag{14}$$

III. MEASUREMENT RESULTS

Figure 3 introduces the coordinates used in the measurement setup, while Fig. 4 shows the equivalent situation from a mathematical point of view. See Appendix A for details regarding coordinate relations. Scattering from the hemisphere on the infinite plane was calculated with the help of Eq. (14) by modeling the loudspeaker as a cluster of monopoles, further discussed in Sec. IV A. The measured scattered part was determined by subtracting the result with the hemisphere present to the result with the hemisphere absent.

Loudness is roughly independent of the energy distribution inside critical bands,[9] whose width above 500 Hz approximates to one-third octave. In the comparison between measurement and model, a one-third octave smoothing was therefore applied, see Eq. (B1). Figure 5 shows typical results for comparisons in both the time and frequency domains. In Table I, results for all comparisons between the measured and modeled scattered parts are compiled in terms of a single value, ΔL_{MA}, see Eq. (B3). A compilation of

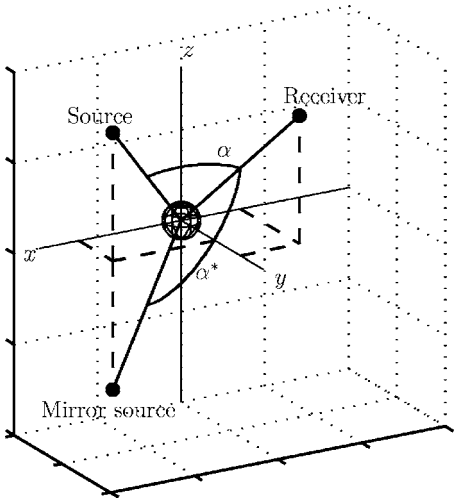


FIG. 4. A situation that is mathematically equivalent to the one in Fig. 3 for $z \geq 0$.

graphs similar to those in Fig. 5 for every measurement is available.[6] In order to analyze the measurements group wise, instead of individually, a number of scattering cases were defined:

- Backscattering: $\alpha \leq 90^\circ$ and $\alpha^* \leq 90^\circ$,
- Forward scattering: $\alpha \geq 90^\circ$ and $\alpha^* \geq 90^\circ$,
- Mixed scattering: $\alpha \leq 90^\circ$ and $\alpha^* \geq 90^\circ$, or $\alpha \geq 90^\circ$ and $\alpha^* \leq 90^\circ$,
- In-plane: $\boldsymbol{r}, \boldsymbol{r}_a, \boldsymbol{r}_0$, and $\boldsymbol{r}_0^*$ all in the same plane,
- Off-plane: $\boldsymbol{r}, \boldsymbol{r}_a, \boldsymbol{r}_0$, and $\boldsymbol{r}_0^*$ *not* in the same plane.

Moreover, focusing occurs for $\alpha^* = 180^\circ$. Table II shows ΔL_{MA} for different groups of measurements. The repeatability of measurements is studied in Table III in terms of standard deviations of ΔL_{MA} for measurements that ideally should yield identical results due to geometrical symmetry.

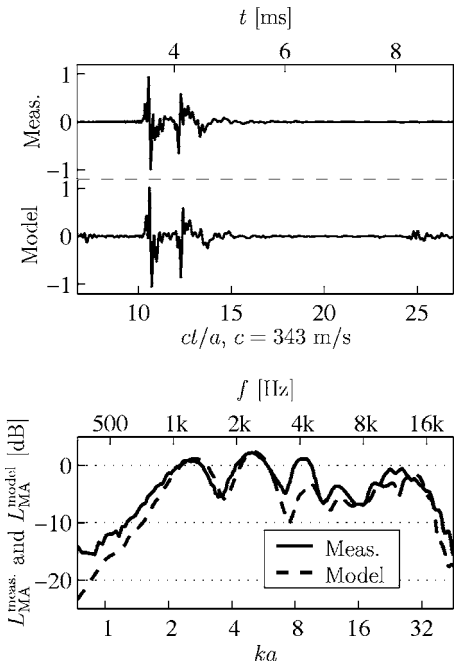


FIG. 5. Typical example of scattered field: $\theta_0 = 0^\circ$, $\theta = 30^\circ$, and $\varphi = 90^\circ$. Frequency response (arbitrary normalization): $L_{\mathrm{MA}}^{\mathrm{meas.}}$ (continuous) and $L_{\mathrm{MA}}^{\mathrm{model}}$(dashed). Transient response (normalized to $\max|p_{\mathrm{model}}|$): Measurement (upper) and model (lower).

TABLE I. $\Delta L_{\mathrm{MA}}[\hat{u}, \hat{v}]$ for all comparisons, with $\hat{u} = \hat{p}_{\mathrm{sc,meas.}} + \hat{p}^*_{\mathrm{sc,meas.}}$ and $\hat{v} = \hat{p}_{\mathrm{sc,\ model}} + \hat{p}^*_{\mathrm{sc,\ model}}$, see Eq. (12). In all measurements, $r_0 = 9a, r = 3a, a = 11.5$ cm, and $\varphi_0 = 0^\circ$.

θ	φ	$\theta_0 = 0^\circ$	$\theta_0 = 30^\circ$	$\theta_0 = 60^\circ$
0		2.3	2.9	2.4
30	0	2.9	1.7	3.2
30	30	2.0	1.8	3.1
30	60	2.2	2.1	2.7
30	90	2.2	1.9	2.4
30	120	2.0	2.9	2.4
30	150	2.9	1.8	2.8
30	180	2.5	3.0	5.2
60	0	2.8	3.3	4.4
60	30	2.1	2.9	3.7
60	60	2.3	2.8	4.8
60	90	2.3	2.4	4.4
60	120	2.1	2.9	2.3
60	150	2.2	2.6	5.4
60	180	3.8	2.1	2.4
90	0	1.6	1.7	
90	30	1.6	1.9	2.0
90	60	1.9	2.0	1.9
90	90	1.9	1.6	1.9
90	120	1.8	1.9	3.5
90	150	19	1.7	2.4
90	180	2.3	2.3	2.3

IV. DISCUSSION

A. Directivity measurements

Equation (14) can be used to compare measurements with theory if the loudspeaker is modeled as a cluster of monopoles. In the scattering measurements, a cylindrical loudspeaker was used as a sound source. The transient response was measured a full circle around the loudspeaker at 5° steps in an anechoic chamber. Measurements showed that the loudspeaker directivity was approximately axially symmetric. For simplicity, the monopoles were therefore placed along the symmetry axis. Using a linear least-squares fit method,[10] the amplitudes and phases of the monopoles were adjusted to reproduce the directivity measurements. Figure 6 indicates that the cluster of monopoles could reproduce the

TABLE II. $\Delta L_{\mathrm{MA}}[\hat{u}, \hat{v}]$ for various groups and *ka*-intervals. Here $\hat{u} = \hat{p}_{\mathrm{sc,meas.}} + \hat{p}^*_{\mathrm{sc,meas.}}$ and $\hat{v} = \hat{p}^*_{\mathrm{sc,model}} + \hat{p}_{\mathrm{sc,model}}$.

ka octave band(s)	1–32	1	2	4	8	16	32
All	2.7	4.2.	1.7	1.4	4.0	2.4	2.7
Back	2.9	3.8	1.6	1.5	4.7	2.8	2.7
Forward	3.1	4.3	2.2	1.7	4.0	2.7	3.3
Mixed	2.6	4.2	1.6	1.4	3.9	2.3	2.6
In plane	2.3	4.2	1.5	1.4	3.6	1.9	2.3
Off plane	3.1	4.2	1.8	1.4	4.4	3.0	3.1
$\theta_0 = 0^\circ$	2.2	4.1	1.4	1.2	3.6	1.8	2.2
$\theta_0 = 30^\circ$	2.3	4.5	1.6	1.3	3.1	1.9	2.4
$\theta_0 = 60^\circ$	3.5	3.9	2.0	1.8	5.2	3.5	3.5

TABLE III. Standard deviations of ΔL_{MA}. $\theta_0=0^\circ$ in all cases.

ka octave band(s)	1–32	1	2	4	8	16	32
$\theta=30^\circ$	0.4	0.9	0.2	0.2	0.4	0.7	0.5
$\theta=60^\circ$	0.6	1.0	0.2	0.6	1.2	0.9	0.5
$\theta=90^\circ$	0.2	0.6	0.4	0.3	0.3	0.4	0.3

directivity well at the measurement positions. Since the loudspeaker was directed toward the hemisphere base center in all measurements, it was sufficient to use directivity measurements for the angles $0^\circ, 5^\circ, \ldots, 30^\circ$ relative to the loudspeaker front. This reduced the number of monopoles that needed to be used, and so reduced calculation time.

Although the amplitudes and phases have been adjusted to reproduce the directivity at the measurement positions, it is not certain that the cluster of monopoles may reproduce the loudspeaker directivity at other positions. This results in an error, whose contribution to the discrepancies in Table I is unknown. The method has been studied extensively,[7,11,12] but a definitive scheme for choosing the number and placement of the monopoles has not been proposed.[13]

B. The sphere — theoretical sensitivity study

A sensitivity study for small variations of transducer positions was made in order to assess what accuracy to expect when theory is to be compared with measurements. Figure 7 shows some typical results for the scattered part in the case of the sphere. Because of the principle of superposition, this applies for hemispheres on planes too, see Eq. (12). One notices that the scattered field becomes stronger and varies more as α increases. Because of symmetry, strong focusing occurs around $\alpha=180^\circ$, which may make scattering phenomena audible even at long distances for spheres.

Table IV shows that small variations in α ($\pm 5^\circ$ in steps of 1°) may result in large variations for the scattered field, especially for large values of α and high frequencies. One notices that the variations are largest around $\alpha_0=150^\circ$. This corresponds well to α_s, the angle that separates the optically illuminated region to the optical shadow:

$$\alpha_s = \arccos\frac{a}{|\boldsymbol{r}_0-\boldsymbol{r}_a|} + \arccos\frac{a}{|\boldsymbol{r}-\boldsymbol{r}_a|} = 154^\circ. \qquad (15)$$

To conclude, slightly inaccurate transducer positions in measurements may result in arge errors when comparing with theory, especially when either α and/or α^* is close to α_s.

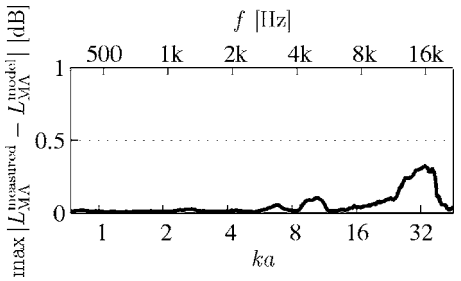


FIG. 6. Reproduction of the loudspeaker directivity with a cluster of monopoles. The graph shows $\max|L_{\mathrm{MA}}^{\mathrm{measured}}-L_{\mathrm{MA}}^{\mathrm{model}}|$ over all angles $0^\circ, 5^\circ, \ldots, 30^\circ$ re frontal.

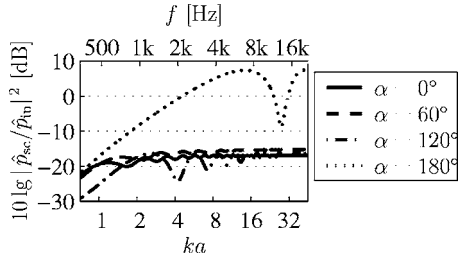


FIG. 7. $10\lg|\hat{p}_{\mathrm{sc}}/\hat{p}_{\mathrm{in}}|^2$ for the angles $\alpha=0^\circ, 60^\circ, 120^\circ$, and 180°, respectively. In all four cases, $|\boldsymbol{r}-\boldsymbol{r}_a|=3a, |\boldsymbol{r}_0-\boldsymbol{r}_a|=9a$, with $a=0.115$ m. Results for $1/\sqrt{2}\leq ka\leq 32\sqrt{2}$ (600 values, logarithmic spacing) where $k=2\pi f/c$ and $c=343$ m/s.

C. Scattering measurements

On average, the measurements and calculations show good agreement; with ΔL_{MA} typically less than 3 dB for most scattering cases. Therefore, although normal hearing subjects typically show a level-difference threshold around 1.5 dB for similar cases,[14] masking from other reflections suggests that this threshold is higher for applications in room acoustics. Since the scattered part $\hat{p}_{\mathrm{sc}}+\hat{p}_{\mathrm{sc}}^*$ is always accompanied by the incident field and the reflection from the infinite plane $\hat{p}_{\mathrm{in}}+\hat{p}_{\mathrm{in}}^*$, significant masking is expected. Except for focusing, i.e., around $\alpha^*=180^\circ$, the scattered sound pressure level is of the order of 20 dB weaker than the incident field, see Fig. 7.

Results are worse for forward scattering, which is expected since the results are sensitive for large angles α and

TABLE IV. Small angular variations may lead to large variations in $10\lg|\hat{p}_{\mathrm{sc}}/\hat{p}_{\mathrm{in}}|^2$. First, octave band averages (centered around $ka=1,2,4,8,16$, and 32) of $|\hat{p}_{\mathrm{sc}}/\hat{p}_{\mathrm{in}}|^2$ were calculated for $\alpha=0^\circ, 1^\circ, 2^\circ, \ldots, 180^\circ$. The table entries show (in decibels) how much these averages vary in the intervals $\alpha\in[\alpha_0-5^\circ, \alpha_0+5^\circ]$. In all calculations $|\boldsymbol{r}-\boldsymbol{r}_a|=3a, |\boldsymbol{r}_0-\boldsymbol{r}_a|=9a$, with $a=0.115$ m. Results for $1/\sqrt{2}\leq ka\leq 32\sqrt{2}$ (600 values, logarithmic spacing) where $k=2\pi f/c$ and $c=343$ m/s.

α_0	1	2	4	8	16	32
0°	0.0	0.0	0.0	0.0	0.0	0.0
10°	0.1	0.1	0.1	0.1	0.1	0.1
20°	0.2	0.2	0.2	0.2	0.3	0.3
30°	0.2	0.3	0.3	0.3	0.3	0.3
40°	0.1	0.2	0.3	0.3	0.3	0.4
50°	0.1	0.2	0.2	0.3	0.3	0.3
60°	0.3	0.1	0.2	0.2	0.3	0.3
70°	0.6	0.1	0.0	0.0	0.2	0.2
80°	0.9	0.1	0.3	0.1	0.0	0.1
90°	1.3	0.1	0.7	0.2	0.2	0.1
100°	1.5	0.1	0.6	0.8	0.4	0.3
110°	0.7	0.2	0.8	0.9	0.9	0.7
120°	1.1	0.3	0.9	1.1	1.5	1.3
130°	2.1	1.7	1.7	1.1	2.4	2.5
140°	2.0	2.6	4.8	2.9	2.4	2.6
150°	1.5	2.4	5.6	6.7	9.3	11.9
160°	1.0	1.6	4.1	7.9	8.0	9.4
170°	0.5	0.8	2.1	6.4	5.1	3.4
180°	0.1	0.1	0.3	0.8	2.3	5.8

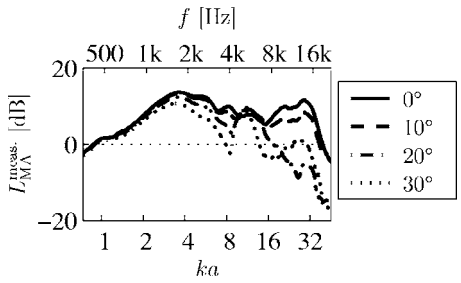


FIG. 8. The measured loudspeaker directivity for the angles $0^\circ, 10^\circ, 20^\circ, 30^\circ$ re frontal. Arbitrary normalization.

α^*. More surprising, from this point of view, is that backscattering results are worse than the total average, too. This could be related to the fact that the microphone generally is closer to the loudspeaker for backscattering in comparison with other cases, and that the loudspeaker enclosure itself may reflect sound. The loudspeaker is however modeled as a number of infinitesimally small monopole sources, that cannot by themselves scatter any secondary incident waves (see Fig. 8).

The number of samples are fewer for low octaves, since the measurements were made with constant sampling frequency. Larger discrepancies are therefore to be expected at the lower octave bands. Another factor that may affect the magnitude of the discrepancies is the averaging process.[15] In comparison with $\Delta L_{\mathrm{MA}}, \Delta L_{\mathrm{MG}}$ enhances peaks and ignores dips to a higher degree (see Appendix B). Table V shows that ΔL_{MG} produces better overall numerical agreement.

As a final comment, significantly better agreements could be achieved by assuming that the measurements for $\theta_0 = 60^\circ$ actually had been carried out for $\theta_0 = 55^\circ$. The improved statistics for those measurements are shown in Table VI and illustrates the sensitivity to small angular variations well.

V. CONCLUSION

Scattering phenomena play an important part in determining the acoustic quality of auditoria, although its relation to, e.g., diffusion is not yet understood.[16] Accurate scattering models are available through boundary element models for example, but they are not fast enough for use in prediction of concert halls acoustics,[17] whose size often exceeds 20000 m^3. As a result, the necessary accuracy needed for auralization purposes is largely unknown.

To address the mentioned challenges, scattering of sound from a hemisphere on a plane was measured and the results compared with the boss model — a combination of the method of images and separation of variables. The hemispherical shape was chosen partly because an analytical reference solution for Neumann boundary conditions with point source excitation exists. Together with the principle of superposition, the point source excitation model was sufficient, since the loudspeaker directivity was modeled as a cluster of point sources. Moreover, since many objects designed to scatter sound in auditoria are rigid and immobile, solutions for Neumann boundary conditions are applicable in many situations of practical interest. Another reason was that scattering from hemispheres may be significant even at long distances because it focuses sound from the virtual image source, and this may find useful applications in the design of auditoria.

Motivated by psychoacoustic results,[9] which states that loudness is roughly independent of the energy distribution inside critical bands, a moving one-third octave band level smoothing was applied for more relevant comparison between measured and theoretical results. In this case, mean intensity level errors typically less than 3 dB were obtained. These results are promising, but whether this accuracy is sufficient or not in each situation of practical interest can ultimately be answered only by psychoacoustic experiments. An important result was, however, that small variations in transducer positions could lead to large variations in the results; an inherent physical phenomenon that cannot be circumvented by alternative modeling methods.

TABLE V. ΔL_{MA} vs ΔL_{MG}.

ka octave band(s)	1–32	1	2	4	8	16	32
ΔL_{MA}	2.7	4.2	1.7	1.4	4.0	2.4	2.7
ΔL_{MG}	2.6	3.0	1.2	1.5	3.7	2.4	2.6

TABLE VI. ΔL_{MA} assuming that θ_0 was 55° instead of 60° in the measurements. Values for $\theta_0 = 60^\circ$ are included for comparison.

ka octave band(s)	1–32	1	2	4	8	16	32
$\theta_0 = 55^\circ$	3.1	4.1	2.1	1.5	4.6	2.5	3.3
$\theta_0 = 60^\circ$	3.5	3.9	2.0	1.8	5.2	3.5	3.5

ACKNOWLEDGMENTS

This work would not have been possible without the research funding provided by Vetenskapsrådet. The authors also wish to thank Christopher Cichetti and Jovanie Zayas at Worcester Polytechnic Institute who performed all the measurements during their stay at the department, and their major advisor, Professor Richard H. Campbell.

APPENDIX A: MEASUREMENT COORDINATE SYSTEM

A spherical coordinate system (x, y, z) is used to identify the microphone and loudspeaker center positions $\boldsymbol{r} = (r, \theta, \varphi)$ and $\boldsymbol{r}_0 = (r_0, \theta_0, \varphi_0)$, (see Figs. 3 and 4).

$$\begin{cases} x = r \sin\theta \cos\varphi, \\ y = r \sin\theta \sin\varphi, \\ z = r \cos\theta. \end{cases} \tag{A1}$$

The hemisphere base center lies in the origin, and the z axis (and $\boldsymbol{n}$) is pointing upward from the concrete floor. This implies that $\varphi_0^* = \varphi_0$ and $\theta_0^* = \pi - \theta_0$. The relations between $(\theta, \varphi), (\theta_0, \varphi_0), (\theta_0^*, \varphi_0^*), \alpha$, and α^* become

$$\cos\alpha = \sin\theta \sin\theta_0 \cos(\varphi - \varphi_0) + \cos\theta \cos\theta_0, \tag{A2}$$

$$\cos \alpha^* = \sin \theta \sin \theta_0 \cos(\varphi - \varphi_0) - \cos \theta \cos \theta_0. \quad \text{(A3)}$$

In all measurements, $r_0 = 9a, r = 3a, a = 11.5$ cm, and $\varphi_0 = 0°$.

APPENDIX B: VARIOUS LEVEL MEASURES

Let $\hat{u}$ be a function of ω and let $\hat{u}_{\text{ref}}$ be some reference value. We define the moving one-third octave band *arithmetic* mean level $L_{\text{MA}}[\hat{u}](\omega)$, a function of ω, as

$$L_{\text{MA}}[\hat{u}](\omega) = 10 \lg \left[\frac{1}{\omega \sqrt[3]{2}} \int_{\eta=\omega/\sqrt[6]{2}}^{\omega\sqrt[6]{2}} \left| \frac{\hat{u}(\eta)}{\hat{u}_{\text{ref}}} \right|^2 d\eta \right], \quad \text{(B1)}$$

An alternative is the moving one-third octave band *geometric* mean level $L_{\text{MG}}[\hat{u}](\omega)$:

$$L_{\text{MG}}[\hat{u}](\omega) = \frac{1}{\omega \sqrt[3]{2}} \int_{\eta=\omega/\sqrt[6]{2}}^{\omega\sqrt[6]{2}} 10 \lg \left| \frac{\hat{u}(\eta)}{\hat{u}_{\text{ref}}} \right|^2 d\eta. \quad \text{(B2)}$$

$\Delta L_{\text{MX}}[\hat{u}, \hat{v}]$ (X stands for either A or G) is a single number used to estimate the overall similarity between $\hat{u}(\omega)$ and $\hat{v}(\omega)$:

$$\Delta L_{\text{MX}}[\hat{u}, \hat{v}] = \frac{1}{\bar{\omega} - \underline{\omega}} \int_{\eta=\underline{\omega}}^{\bar{\omega}} |L_{\text{MX}}[\hat{u}](\eta) - L_{\text{MX}}[\hat{v}](\eta)| d\eta. \quad \text{(B3)}$$

[1] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw–Hill, New York, 1953), Chap. 5.1.

[2] J. J. Bowman, T. B. A. Senior, and P. L. E. Uslenghi, *Electromagnetic and Acoustic Scattering by Simple Shapes* (Taylor and Francis, Abingdon, 1988).

[3] V. Twersky, "On the nonspecular reflection of plane waves of sound," J. Acoust. Soc. Am. **22**, 539–546 (1950).

[4] V. Twersky, "On the nonspecular reflection from planes with absorbent bosses," J. Acoust. Soc. Am. **23**, 336–338 (1951).

[5] R. R. Goodman and R. Stern, "Reflection and transmission of sound by elastic spherical shells," J. Acoust. Soc. Am. **34**, 338–344 (1962).

[6] G. Natsiopoulos, "Acoustic scattering for a hard hemisphere on a hard plane," Report 2005:5, Chalmers University of Technology Sweden, Division of Applied Acoustics (2005).

[7] G. Pavić, "An engineering technique for the computation of sound radiation by vibrating bodies using substitute sources," Acta. Acust. Acust. **91**, 1–16 (2005).

[8] F. W. Byron, Jr. and R. W. Fuller, *Mathematics of Classical and Quantum Physics* (Dover, New York, 1992), Chap. 7.5.

[9] E. Zwicker and H. Fastl, *Psychoacoustics* (Springer, New York, 1999), 158–164.

[10] M. T. Heath, *Scientific Computing: An Introductory Survey New York* (McGraw–Hill, New York, 1997), Chap. 3.3.

[11] P. R. Stepanishen, "Acoustic axisymmetric radiation and scattering from bodies of revolution using the internal source density and fourier methods," J. Acoust. Soc. Am. **102**, 726–732 (1997).

[12] Y. Gounot, R. E. Musafir, and J. G. Slama, "A comparative study of two variants of the equivalent sources method in scattering problems," Acta. Acust. Acust. **91**, 860–872 (2005).

[13] M. B. S. Magalhães and R. A. Tenenbaum, "Sound sources reconstruction techniques: A review of their evolution and new trends," Acta. Acust. Acust. **90**, 199–220 (2004).

[14] E. Zwicker and H. Fastl, *Psychoacoustics* (Springer, New York, 1999), 333–334.

[15] H. Cox, "Linear versus logarithmic averaging," J. Acoust. Soc. Am. **39**, 688–690 (1966).

[16] L. Beranek, *Concert Halls and Opera Houses* (Springer, New York, 2004), 521–524.

[17] U. P. Svensson, "Modeling room acoustics," *Proceedings of the Baltic-Nordic Acoustics Meeting*, Mariehamn, Åland (Acoustical Society of Finland 2004).

The influence of circumferential edge constraint on the acoustical properties of open-cell polyurethane foam samples

Huoy-Shyi Tsay[a] and Fung-Huei Yeh
Department of Mechanical and Electro-Mechanical Engineering, Tamkang University, 151 Ying-Chuan Road, Tamsui Taipei Hsien, Taiwan 25137, R.O.C.

(Received 16 August 2005; revised 15 February 2006; accepted 16 February 2006)

In this paper, a three-dimensional Laplace domain finite element model with four-node tetrahedral porous elements is developed and applied for the investigations. Without the incident acoustic field, the exact acoustical properties of a laterally infinite planar foam, which is backed with a rigid plane, are accurately predicted by assuming the circumferential edge constraints as rollers. It is found that the rigidity of the circumferential edge constraint has a profound effect on the predictions. If the circumferential edge constraints are all fixed, the amplitude of the first sound absorption peak is increased, and the peak is moved to a lower frequency as compared with the exact solutions. Measurements obtained by the two-microphone impedance tube method are bounded by the predictions of samples with roller as well as with fixed constraints. They are much closer to that predicted with fixed edge constraints. Upon increasing the diameter-to-thickness ratio of the sample, the influence of the fixed edge constraint is reduced. The measured acoustic results can be the representatives of the laterally infinite planar foam from which the sample was cut only if the roller edge constraint of the sample is managed or the diameter-to-thickness ratio of the sample is increased. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2185494]

PACS number(s): 43.55.Ev, 43.20.Gp, 43.58.Ta, 43.58.Bh [NX] Pages: 2804–2814

I. INTRODUCTION

Acoustical properties (dynamic stiffness, acoustic impedance, reflection coefficient, and absorption coefficient) are indices for evaluating the quality of the sound absorption materials. Generally, the acoustical properties can be measured in free field,[1,2] in a reverberant room,[3] and in a tube.[4–7] Techniques of measurements in a tube mainly found are one-[4] and two-microphone[5] impedance tube techniques. The two-microphone impedance tube method is considerably automatically calibrated if the sensor switching procedure is performed. Other than the measurement techniques, analytical models generally achieved for predicting the acoustical properties of the sound absorption material are the equivalent fluid model and the sophisticated model based on Biot's theory of poroelasticity. The equivalent fluid model assumes the sound absorption material as an isotropic mixed fluid and the propagation of sound is determined by two complex quantities, the characteristic impedance and the wave number. These two quantities are empirically expressed by power law expressions of a single parameter, air resistance divided by frequency.[8,9] In the sophisticated model, the Biot theory[10–12] assumes the sound absorption material is constructed such that a solid material forms a skeleton that contains statistically distributed small pores that are filled with a Newtonian, viscous compressible fluid. The bulk material is assumed homogeneous on a macroscopic scale and the pores are assumed interconnected. The solid skeleton is taken to be linear elastic and undergoes small deformation, and the fluid inertia and the friction are characterized by the density, viscosity, and pore dimensions.

The exact acoustical property solutions to the sound absorption material are restricted to a laterally infinite, finite-thickness planar sheet.[13] Commonly, the acoustical properties of the sound absorption material with complex geometry are numerically pursued by applying the porous model coupled by an incident acoustic field with which the sound absorption material is in contact. The transmission and reflection waves that characterized acoustical properties are then solved. Based on the equivalent medium properties, the resistivity and the porosity, Craggs[14,15] examined a one-dimensional, axially symmetric acoustic absorption system for obtaining the absorption coefficient. Easwaran and Munjal[16] predicted the reflection characteristics of rectangular and conical wedges in an impedance tube. Subsequently, Chen *et al.*[17] investigated the acoustic absorption characteristics of porous materials with triangle, semicircle, and convex rectangular surface shapes. There is only one longitudinal wave to be examined, if the equivalent fluid model is applied. If the prediction and the design of acoustical properties of the sound absorption material are mainly considered, methods that combine the acoustic field with the Biot theory are more effective. Behaviors of two longitudinal waves and one shear wave in the porous material can be examined with the use of Biot theory.[11] Using the Biot complete poroelastic theory, Kang and Bolton[18–20] studied the optimal design of axisymmetric foam wedges and examined the effect of the circumferential edge constraint on the absorption coefficient as well as the vibration modes of a porous material sample placed in the standing wave tube. In any case, the incident acoustic field is needed to perform simulations.

Numerous researchers[6,7,19,21–23] have found the acoustical properties measured from an impedance tube strongly

[a]Electronic mail: huoyshyi@mail.tku.edu.tw

 0001-4966/2006/119(5)/2804/11/$22.50

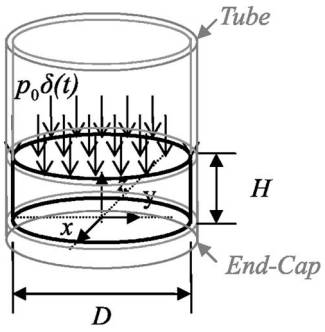


FIG. 1. Cylindrical foam sample subjected to an impulsive sound pressure loading.

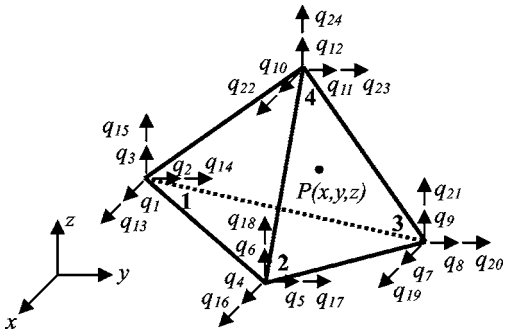


FIG. 2. Four-node tetrahedral porous element.

dependent on the uniformity as well as the circumferential edge of the sample, and the backing between the sample and the end-cap of the tube. The measured results are different from those of the laterally infinite, finite-thickness planar sheet from which the samples were cut. It has been numerically and experimentally revealed[19,21] that with the edge constraint effects applied, the predictions are much closer to the measurements than those of the unconstrained predictions. Heretofore, the numerical methods usually applied are the two-dimensional finite element methods. However, with the influence of the circumferential edge constraint, it is unlikely that a two-dimensional result is representative of that of a sample of three-dimensional shape (a cylindrical porous column with low diameter-to-thickness ratio).

In this paper, a new approach using an impulse excitation to examine the influence of the circumferential edge constraint on the three-dimensional acoustic frequency response functions of a rigidly backed polyurethane foam sample placed in the impedance tube is presented. Using this method, the efforts on calculating the incident acoustic field can be omitted. During the analysis, Biot's poroelastic dynamic equation set is first transferred to the Laplace domain. Thereafter the four-node tetrahedral porous elements are used for deriving the three-dimensional finite element equations. After applying a uniformly distributed impulse pressure excitation to the surface of the foam sample, the displacement response at the surface is then calculated by the derived Laplace domain finite element equations. The acoustic response of the foam sample is characterized by a dynamic transfer function $[T(s)]$ that is defined as the ratio of the transformed impulsive pressure excitation $[I(s)]$ to the transformed displacement response $[O(s)]$. Upon obtaining the dynamic transfer function, frequency response functions of the sample representing the dynamic stiffness, surface impedance, reflection, and absorption coefficients in the frequency domain can be obtained by simply replacing the Laplace transform parameter s by $j\omega$ where ω is the radial frequency.[13,24,25] Subsequently, with the use of the roller circumferential edge constraint, the acoustical properties of a planar, laterally infinite, and rigidly backed open-cell polyurethane foam with permeable upper surface are validated first. Then, the influences of the fixed circumferential edge constraint and the diameter-to-thickness ratio on the acoustical properties of polyurethane foam samples are examined. Comparisons of results predicted with those measured by the two-microphone impedance tube method are made as well.

II. THEORY

A. Governing equations of foams

In this work, Biot's theory of poroelasticity is employed to simulate the dynamic behaviors of polyurethane foam samples placed in the impedance tube. The equations governing the deformation of poroelastic material are written as

$$N\nabla^2\mathbf{u} + \nabla[(A+N)e + Q\varepsilon] = \rho_{11}\ddot{\mathbf{u}} + \rho_{12}\ddot{\mathbf{U}} + b(\dot{\mathbf{u}} - \dot{\mathbf{U}}), \quad (1)$$

$$\nabla(Qe + R\varepsilon) = \rho_{12}\ddot{\mathbf{u}} + \rho_{22}\ddot{\mathbf{U}} - b(\dot{\mathbf{u}} - \dot{\mathbf{U}}), \quad (2)$$

where N, A, Q, R are elastic constants, b is the dissipation coefficient. $\mathbf{u}$ is the average solid skeleton displacement vector, and $\mathbf{U}$ is the average fluid displacement vector. Similarly, the solid dilatation e and the fluid dilatation ε are defined in the usual manner for small deformation. ∇^2 is the Laplacian and ∇ is the gradient. The quantities ρ_{11}, ρ_{22}, and ρ_{12} are apparent mass densities. ρ_{11} represents the total effective mass of the solid moving in the fluid, ρ_{22} represents the total effective mass of that part of the fluid, and ρ_{12} represents a mass coupling parameter between the fluid and the solid.[11] When there is no relative motion between the fluid and solid skeleton, it is found that $\rho = \rho_{11} + \rho_{22} + 2\rho_{12}$. Here, ρ denotes the total mass density of the fluid-saturated elastic foam. ρ can be expressed in terms of the mass density of solid ρ_s, the mass density of fluid ρ_f, and the porosity ϕ as $\rho = (1-\phi)\rho_s + \phi\rho_f$.

The dynamic behavior of the cylindrical polyurethane foam sample in the impedance tube is investigated by assuming its lower surface fixed on the infinite rigid end-cap and subjected to a uniform distributed impulsive sound pressure loading on its permeable upper surface. The cylindrical sample of diameter D and thickness H is presented in Fig. 1. In the figure, p_o is the amplitude of the pressure loading and $\delta(t)$ is the impulse at time t. The foam considered is assumed to be deformed from the initially un-deformed position and all stresses are zero if the foam is stress free at the initial state.

Taking the Laplace transform with respect to the time variable of Eqs. (1) and (2) and then using the initial condi-

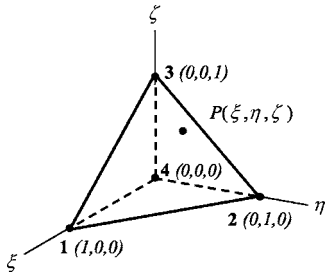


FIG. 3. Four-node master element in ξ, η, and ζ space.

tions mentioned and the Cartesian coordinates shown, the dynamic equations for the three-dimensional foam problems yields

$$L_1(\bar{u}_x,\bar{u}_y,\bar{u}_z,\bar{U}_x,\bar{U}_y,\bar{U}_z) = N\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\bar{u}_x + (A+N)\frac{\partial}{\partial x}\left(\frac{\partial \bar{u}_x}{\partial x}+\frac{\partial \bar{u}_y}{\partial y}+\frac{\partial \bar{u}_z}{\partial z}\right) + Q\frac{\partial}{\partial x}\left(\frac{\partial \bar{U}_x}{\partial x}+\frac{\partial \bar{U}_y}{\partial y}+\frac{\partial \bar{U}_z}{\partial z}\right) - s^2(\rho_{11}\bar{u}_x+\rho_{12}\bar{U}_x) - bs(\bar{u}_x-\bar{U}_x)=0, \quad (3)$$

$$L_2(\bar{u}_x,\bar{u}_y,\bar{u}_z,\bar{U}_x,\bar{U}_y,\bar{U}_z) = N\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\bar{u}_y + (A+N)\frac{\partial}{\partial y}\left(\frac{\partial \bar{u}_x}{\partial x}+\frac{\partial \bar{u}_y}{\partial y}+\frac{\partial \bar{u}_z}{\partial z}\right) + Q\frac{\partial}{\partial y}\left(\frac{\partial \bar{U}_x}{\partial x}+\frac{\partial \bar{U}_y}{\partial y}+\frac{\partial \bar{U}_z}{\partial z}\right) - s^2(\rho_{11}\bar{u}_y+\rho_{12}\bar{U}_y) - bs(\bar{u}_y-\bar{U}_y)=0, \quad (4)$$

$$L_3(\bar{u}_x,\bar{u}_y,\bar{u}_z,\bar{U}_x,\bar{U}_y,\bar{U}_z) = N\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\bar{u}_z + (A+N)\frac{\partial}{\partial z}\left(\frac{\partial \bar{u}_x}{\partial x}+\frac{\partial \bar{u}_y}{\partial y}+\frac{\partial \bar{u}_z}{\partial z}\right) + Q\frac{\partial}{\partial z}\left(\frac{\partial \bar{U}_x}{\partial x}+\frac{\partial \bar{U}_y}{\partial y}+\frac{\partial \bar{U}_z}{\partial z}\right) - s^2(\rho_{11}\bar{u}_z+\rho_{12}\bar{U}_z) - bs(\bar{u}_z-\bar{U}_z)=0, \quad (5)$$

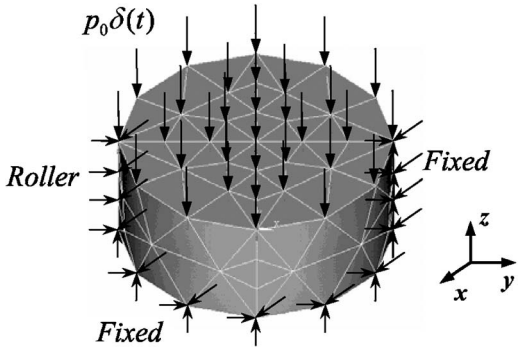


FIG. 4. Boundary conditions of a three-dimensional polyurethane foam.

$$L_4(\bar{u}_x,\bar{u}_y,\bar{u}_z,\bar{U}_x,\bar{U}_y,\bar{U}_z) = \frac{\partial}{\partial x}\left[Q\left(\frac{\partial \bar{u}_x}{\partial x}+\frac{\partial \bar{u}_y}{\partial y}+\frac{\partial \bar{u}_z}{\partial z}\right)+R\left(\frac{\partial \bar{U}_x}{\partial x}+\frac{\partial \bar{U}_y}{\partial y}+\frac{\partial \bar{U}_z}{\partial z}\right)\right] - s^2(\rho_{12}\bar{u}_x+\rho_{22}\bar{U}_x) + bs(\bar{u}_x-\bar{U}_x)=0, \quad (6)$$

$$L_5(\bar{u}_x,\bar{u}_y,\bar{u}_z,\bar{U}_x,\bar{U}_y,\bar{U}_z) = \frac{\partial}{\partial y}\left[Q\left(\frac{\partial \bar{u}_x}{\partial x}+\frac{\partial \bar{u}_y}{\partial y}+\frac{\partial \bar{u}_z}{\partial z}\right)+R\left(\frac{\partial \bar{U}_x}{\partial x}+\frac{\partial \bar{U}_y}{\partial y}+\frac{\partial \bar{U}_z}{\partial z}\right)\right] - s^2(\rho_{12}\bar{u}_y+\rho_{22}\bar{U}_y) + bs(\bar{u}_y-\bar{U}_y)=0, \quad (7)$$

$$L_6(\bar{u}_x,\bar{u}_y,\bar{u}_z,\bar{U}_x,\bar{U}_y,\bar{U}_z) = \frac{\partial}{\partial z}\left[Q\left(\frac{\partial \bar{u}_x}{\partial x}+\frac{\partial \bar{u}_y}{\partial y}+\frac{\partial \bar{u}_z}{\partial z}\right)+R\left(\frac{\partial \bar{U}_x}{\partial x}+\frac{\partial \bar{U}_y}{\partial y}+\frac{\partial \bar{U}_z}{\partial z}\right)\right] - s^2(\rho_{12}\bar{u}_z+\rho_{22}\bar{U}_z) + bs(\bar{u}_z-\bar{U}_z)=0, \quad (8)$$

where $\bar{u}_x$, $\bar{u}_y$, $\bar{u}_z$, $\bar{U}_x$, $\bar{U}_y$, and $\bar{U}_z$ are the Laplace transforms of displacement components of the solid skeleton (u_x,u_y,u_z) and the fluid (U_x,U_y,U_z), respectively. s above is the Laplace transform parameter.

B. Finite element model

The Galerkin finite element method is applied to obtain the acoustical properties of polyurethane foam samples in the study. Figure 2 shows a four-node tetrahedral porous element. $P(x,y,z)$ in Fig. 2 represents an arbitrary point within an element. If three Cartesian components of displacement of the solid skeleton and of fluid at each node are defined, the four-node tetrahedral porous element will have 24 degrees of freedom, given by the column vector of displacements $[\mathbf{q}]=[q_1,\ldots,q_{24}]^T$.

In terms of a natural coordinate system (ξ, η, ζ), the tetrahedral porous element defined with reference to the Cartesian coordinate system (x, y, z) can then be transformed into a master element as shown in Fig. 3. $P(\xi, \eta, \zeta)$ becomes the coordinate system transformation of $P(x, y, z)$. According to the natural coordinates defined, four Lagrange-type shape functions for the tetrahedral element can be written as

$$N_1 = \xi,$$

$$N_2 = \eta,$$

$$N_3 = \zeta, \tag{9}$$

$$N_4 = 1 - \xi - \eta - \zeta.$$

With the use of the column vector of displacements $[\mathbf{q}]$ and shape functions defined, solid skeleton and fluid displacement components can be written in a compact form as

$$[\mathbf{u}] = [\mathbf{N}][\mathbf{q}], \tag{10}$$

where $[\mathbf{u}]=[\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z]^T$ and $[\mathbf{N}]$ is the shape function matrix. The shape function matrix for the tetrahedral element can be written as

$$[\mathbf{N}] = \begin{bmatrix} [\mathbf{N}]_1 \\ [\mathbf{N}]_2 \\ [\mathbf{N}]_3 \\ [\mathbf{N}]_4 \\ [\mathbf{N}]_5 \\ [\mathbf{N}]_6 \end{bmatrix} = \begin{bmatrix} N_1 & & 0 & N_2 & & 0 & N_3 & & 0 & N_4 & & 0 \\ 0 & N_1 & 0 & 0 & N_2 & 0 & 0 & N_3 & 0 & 0 & N_4 & 0 \\ 0 & 0 & \cdots 0 & 0 & 0 & \cdots 0 & 0 & 0 & \cdots 0 & 0 & 0 & \cdots 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & N_1 & 0 & 0 & N_2 & 0 & 0 & N_3 & 0 & 0 & N_4 \end{bmatrix}.$$

Then, following the Galerkin approach, the equation governing the behavior of an element becomes

$$\begin{aligned} \sum_{e=1}^{n} \int_{V_e} & [\mathbf{N}]_1^T L_1(\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z) \\ & + [\mathbf{N}]_2^T L_2(\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z) \\ & + [\mathbf{N}]_3^T L_3(\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z) \\ & + [\mathbf{N}]_4^T L_4(\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z) \\ & + [\mathbf{N}]_5^T L_5(\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z) \\ & + [\mathbf{N}]_6^T L_6(\bar{u}_x, \bar{u}_y, \bar{u}_z, \bar{U}_x, \bar{U}_y, \bar{U}_z) dV = 0. \end{aligned} \tag{11}$$

If we integrate Eq. (11) by parts using the Green-Gauss theorem or the divergence theorem and then summarize the results, we obtain

$$\begin{aligned} \sum_{e=1}^{n} \int_{V_e} & [[N]_1^T, [N]_2^T, [N]_3^T, [N]_4^T, [N]_5^T, [N]_6^T] \\ & \times \begin{Bmatrix} s^2(\rho_{11}\bar{u}_x + \rho_{12}\bar{U}_x) + bs(\bar{u}_x - \bar{U}_x) \\ s^2(\rho_{11}\bar{u}_y + \rho_{12}\bar{U}_y) + bs(\bar{u}_y - \bar{U}_y) \\ s^2(\rho_{11}\bar{u}_z + \rho_{12}\bar{U}_z) + bs(\bar{u}_z - \bar{U}_z) \\ s^2(\rho_{12}\bar{u}_x + \rho_{22}\bar{U}_x) - bs(\bar{u}_x - \bar{U}_x) \\ s^2(\rho_{12}\bar{u}_y + \rho_{22}\bar{U}_y) - bs(\bar{u}_y - \bar{U}_y) \\ s^2(\rho_{12}\bar{u}_z + \rho_{22}\bar{U}_z) - bs(\bar{u}_z - \bar{U}_z) \end{Bmatrix} dV \\ & - \sum_{e=1}^{n} \int_{l_e} [[\mathbf{N}]_1^T, [\mathbf{N}]_2^T, [\mathbf{N}]_3^T, [\mathbf{N}]_4^T, [\mathbf{N}]_5^T, [\mathbf{N}]_6^T] \\ & \times \begin{Bmatrix} \bar{\sigma}_{xx} l_x + \bar{\tau}_{xy} l_y + \bar{\tau}_{xz} l_z \\ \bar{\tau}_{xy} l_x + \bar{\sigma}_{yy} l_y + \bar{\tau}_{yz} l_z \\ \bar{\tau}_{xz} l_x + \bar{\tau}_{yz} l_x + \bar{\sigma}_{zz} l_z \\ \bar{\sigma} l_x \\ \bar{\sigma} l_y \\ \bar{\sigma} l_z \end{Bmatrix} dl + \sum_{e=1}^{n} \int_{V_e} [\mathbf{k}]^T [\hat{\mathbf{q}}] dV = 0 \end{aligned} \tag{12}$$

where $\bar{\sigma}_{xx}$, $\bar{\sigma}_{yy}$, $\bar{\sigma}_{zz}$, $\bar{\tau}_{xy}$, $\bar{\tau}_{xz}$, $\bar{\tau}_{yz}$, and $\bar{\sigma}$ are the Laplace transforms of σ_{xx}, σ_{yy}, σ_{zz}, τ_{xy}, τ_{xz}, τ_{yz}, and σ, respectively. σ_{xx}, σ_{yy}, σ_{zz}, τ_{xy}, τ_{xz}, and τ_{yz} are stresses acting on the solid skeleton of the foam and σ is the stress of the air in the foam;

$$[\hat{\mathbf{q}}] = \left\{ \frac{\partial \bar{u}_x}{\partial x}, \frac{\partial \bar{u}_x}{\partial y}, \frac{\partial \bar{u}_x}{\partial z}, \frac{\partial \bar{u}_y}{\partial x}, \frac{\partial \bar{u}_y}{\partial y}, \frac{\partial \bar{u}_y}{\partial z}, \frac{\partial \bar{u}_z}{\partial x}, \frac{\partial \bar{u}_z}{\partial y}, \frac{\partial \bar{u}_z}{\partial z}, \right.$$
$$\left. \frac{\partial \bar{U}_x}{\partial x}, \frac{\partial \bar{U}_x}{\partial y}, \frac{\partial \bar{U}_x}{\partial z}, \frac{\partial \bar{U}_y}{\partial x}, \frac{\partial \bar{U}_y}{\partial y}, \frac{\partial \bar{U}_y}{\partial z}, \frac{\partial \bar{U}_z}{\partial x}, \frac{\partial \bar{U}_z}{\partial y}, \frac{\partial \bar{U}_z}{\partial z} \right\}^T,$$

$$[\mathbf{k}] = [k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10}, 0, 0, 0, k_{11}, 0, 0, 0, k_{12}],$$

in which

$$k_1 = (A + 2N) \frac{\partial [N]_1^T}{\partial x} + A \left(\frac{\partial [N]_2^T}{\partial y} + \frac{\partial [N]_3^T}{\partial z} \right) + Q \left(\frac{\partial [N]_4^T}{\partial x} + \frac{\partial [N]_5^T}{\partial y} + \frac{\partial [N]_6^T}{\partial z} \right),$$

$$k_2 = k_4 = \left(\frac{\partial [N]_1^T}{\partial y} + \frac{\partial [N]_2^T}{\partial x} \right),$$

$$k_3 = k_7 = \left(\frac{\partial [N]_1^T}{\partial z} + \frac{\partial [N]_3^T}{\partial x} \right),$$

$$k_5 = (A + 2N) \frac{\partial [N]_2^T}{\partial y} + A \left(\frac{\partial [N]_1^T}{\partial x} + \frac{\partial [N]_3^T}{\partial z} \right) + Q \left(\frac{\partial [N]_4^T}{\partial x} + \frac{\partial [N]_5^T}{\partial y} + \frac{\partial [N]_6^T}{\partial z} \right),$$

$$k_6 = k_8 = \left(\frac{\partial [N]_2^T}{\partial z} + \frac{\partial [N]_3^T}{\partial y} \right),$$

$$k_9 = (A + 2N) \frac{\partial [N]_3^T}{\partial z} + A \left(\frac{\partial [N]_1^T}{\partial x} + \frac{\partial [N]_2^T}{\partial y} \right) + Q \left(\frac{\partial [N]_4^T}{\partial x} + \frac{\partial [N]_5^T}{\partial y} + \frac{\partial [N]_6^T}{\partial z} \right),$$

and

$$k_{10} = k_{11} = k_{12} = Q \left(\frac{\partial [N]_1^T}{\partial x} + \frac{\partial [N]_2^T}{\partial y} + \frac{\partial [N]_3^T}{\partial z} \right) + R \left(\frac{\partial [N]_4^T}{\partial x} + \frac{\partial [N]_5^T}{\partial y} + \frac{\partial [N]_6^T}{\partial z} \right).$$

Substituting shape functions [Eq. (9)] and displacement components [Eq. (10)] into Eq. (12) and rearranging terms gives

$$\sum_{e=1}^{n} \{ [\mathbf{K}^{(e)}][\mathbf{q}] - [\mathbf{F}^{(e)}] \} = 0, \tag{13}$$

where $[\mathbf{K}^{(e)}]$ is the stiffness coefficient matrix for the element and $[\mathbf{F}^{(e)}]$ may be viewed as the loading vector which contains boundary conditions.

The three-dimensional foam sample can be analyzed by discretizing it with finite elements. The equations of individual elements would then be assembled to form the equation of the foam. After manipulation, the Laplace domain finite element equation of the foam is simplified as

$$[\mathbf{K}^*][\mathbf{q}^*] = [\mathbf{F}^*], \tag{14}$$

where $[\mathbf{K}^*]$ is the global stiffness matrix, $[\mathbf{q}^*]$ is the global displacement vector, and $[\mathbf{F}^*]$ is the global loading vector. Once the Laplace transformed dynamic stiffness $\bar{K}$ of the foam sample is calculated, the Laplace transformed surface acoustic impedance $\bar{Z}$, the reflection coefficient R_0, and the sound absorption coefficient A_0 for the normal incidence case can be defined as

$$\bar{Z}(s) = \frac{-p_0}{(1-\phi) s \bar{\mathbf{u}}|_{z=\mathrm{H}} + \phi s \bar{\mathbf{U}}|_{z=\mathrm{H}}} = \frac{\bar{K}(s)}{s}, \tag{15}$$

$$R_0 = \frac{\bar{Z} - Z_{\mathrm{air}}}{\bar{Z} + Z_{\mathrm{air}}} = \frac{\bar{K}(s)/s - Z_{\mathrm{air}}}{\bar{K}(s)/s + Z_{\mathrm{air}}}, \tag{16}$$

$$A_0 = 1 - |R_0|^2, \tag{17}$$

where Z_{air} is the characteristic impedance of air. Frequency response functions of K, Z, R_0, and A_0 can then be rapidly obtained by replacing the Laplace transform parameter s that appears in Eqs. (15)–(17) by $j\omega$, where ω is the radial frequency.

C. Boundary conditions in Laplace domain

As shown in Fig. 4, the impulsive sound pressure loading applied on the upper surface of the foam sample is presented as $p = p_o \delta(t)$. If the top surface of the foam is permeable, the pressure loading is shared by the solid skeleton and the fluid of the three-dimensional sample depending on the porosity. Thus, the boundary conditions of the top surface after the Laplace transform can be described as $\bar{\sigma}_f = \phi p_o$ and $\bar{\sigma}_s = (1-\phi) p_o$. The lower surface is presumed to be fixed on an infinite rigid plane. Hence, the Laplace transformed solid skeleton and fluid displacements ($\bar{u}_x$, $\bar{u}_y$, $\bar{u}_z$, $\bar{U}_x$, $\bar{U}_y$, and $\bar{U}_z$) must be zero on this face.

The constraints on the circumferential edge of the foam sample can be set as rollers for simulating a laterally infinite, finite-thickness planar foam. The roller condition implies that the node can freely move up and down but not to the side as shown on the left of the foam sample in Fig. 4; hence, the Laplace transformed solid skeleton and fluid displacements in x and y coordinates, $\bar{u}_x$, $\bar{u}_y$, $\bar{U}_x$, and $\bar{U}_y$, are zero on the circumferential edge. For the case that the circumferential edge is fully constrained, the node constraint can be simulated as fixed as shown on the right of the foam sample. Same as the boundary conditions of the lower surface, fixed condition implies that the solid skeleton and fluid displacements are zero on the surface.

III. MATERIAL PROPERTIES OF POLYURETHANE FOAM

In order to predict the acoustical properties in terms of frequency response functions, the apparent mass densities ρ_{11}, ρ_{12}, ρ_{22}, the dissipation coefficient b, and the elastic constants N, A, Q, R that appeared in Eqs. (1) and (2) can all be related to the foam constituents, the mass density of solid ρ_s, the mass density of fluid ρ_f, porosity ϕ, tortuosity α_∞, pore cross-section shape correction factor c, and DC flow resistivity σ_0. The typical material properties of the polyurethane foam saturated with air are summarized by Tsay and Yeh[13] and presented as follows.

The coupling apparent mass ρ_{12} can be represented by porosity, tortuosity, and the mass density of fluid as

$$\rho_{12} = -\phi \rho_f (\alpha_\infty - 1). \tag{18}$$

The total effective masses of the solid ρ_{11} and fluid ρ_{22} are defined accordingly as

$$\rho_{11} = (1-\phi)\rho_s - \rho_{12}, \tag{19}$$

$$\rho_{22} = \phi \rho_f - \rho_{12}. \tag{20}$$

The porosity of polyurethane foam is around 0.75–0.99 and the density is in the range of 20–300 kg/m^3 usually.

TABLE I. Measured and estimated (*) material properties of polyurethane foams.[13]

	Porosity ϕ	Shear modulus N (10^4 N/m^2)	Poisson coefficient ν	DC flow resistivity σ_o (N.s/m^4)	Tortuosity α_∞	Density of frame ρ_s (kg/m^3)	Shape factor c	Thickness of foam H (m)
Foam 1	0.95	175(1+0.2i)	0.35*	5500	1.0*	45.7	1.0*	0.016
Foam 2	0.95	93(1+0.2i)	0.35*	5500	1.0*	45.7	1.0*	0.030
Foam 3	0.95	55(1+0.2i)	0.35*	5500	1.0*	45.7	1.0*	0.051

The Laplace transformed dissipation coefficient $\bar{b}$ defined in the dynamic poroelastic theory is formulated as

$$\bar{b} = \frac{-\bar{\sigma}_0}{4}\phi^2 \left\{ \frac{[m_{r_0} J_1(m_{r_0})]/[J_0(m_{r_0})]}{1-[2J_1(m_{r_0})]/[m_{r_0} J_0(m_{r_0})]} \right\}, \quad (21)$$

where $\bar{\sigma}_0$ is the DC flow resistivity and is expressed as $\bar{\sigma}_0 = (8c\mu\alpha_\infty)/(\phi r_0^2)$. c is a cross-section shape correction factor, μ is the viscosity of fluid (air), and r_0 is the radius of the pore. For pores of square, equilateral triangle, and rectangular slit, c is 1.07, 1.14, and 0.78, respectively. J_0 and J_1 are Bessel functions of the first kind of order zero and one. m_{r_0} in the Bessel functions is found as $m_{r_0} = c\sqrt{-8s\rho_{22}/(\bar{\sigma}_0\phi^2)}$ in which s is the Laplace transform parameter.

For most acoustic foams, elastic constants A, Q, and R can be expressed by the shear modulus of the solid skeleton in vacuum N, the bulk modulus of the solid skeleton K_b, and the bulk modulus of the air K_f as

$$A = \frac{(1-\phi)^2}{\phi}K_f + K_b - \frac{2}{3}N, \quad (22)$$

$$Q = (1-\phi)K_f, \quad (23)$$

$$R = \phi K_f. \quad (24)$$

N, K_b, and K_f are complex coefficients. K_b can be evaluated by

$$K_b = \frac{2N(1+\nu)}{3(1-2\nu)}, \quad (25)$$

where ν is the Poisson's ratio of the foam. K_f in pores inside the foam can be derived using the equation of state of air and the equation of thermal conduction of air. The Laplace transformed bulk modulus of air $\bar{K}_f$ is formulated as

$$\bar{K}_f = \frac{p_{\text{amb}}\gamma}{\gamma-(\gamma-1)\{1-2/Bm_{r_0}[J_1(Bm_{r_0})]/[J_0(Bm_{r_0})]\}}, \quad (26)$$

where p_{amb} is the ambient mean pressure, B is the square root of the Prandtl number, and $\gamma = c_p/c_v$. c_v and c_p are specific heats per unit mass at constant volume and pressure, respectively.

IV. MEASUREMENT SYSTEM

The two-microphone acoustic measurement system used for the earlier study[13] was applied for measuring the acoustical properties of foam samples. The measurement system was established according to the ASTM E1050-98 standard.[5] The system contains a driving signal generator (ACO 3024), a driving signal amplifier (LDS PA-100E), a 6-in. speaker, two impedance tubes of 0.0395- and 0.1268-m diameter, two pairs of acoustic microphones (B&K 4187) and amplifiers (Norsonic 336), and a dynamic signal analyzer (SD-390). When the reliable measurement errors are kept under 0.5 dB, the applicable frequency ranges of the small and large impedance tubes are from 100 to 4000 Hz and from 100 to 1600 Hz, respectively.

In the acoustic measurement system, the phasor of sound pressure measured at position 1 through microphone 1 can be derived as $A_1(\omega)e^{i\phi_1(\omega)}M_1(\omega)e^{i\varphi_1(\omega)}$, and the phasor of sound pressure measured at position 2 through microphone 2 is then found as $A_2(\omega)e^{i\phi_2(\omega)}M_2(\omega)e^{i\varphi_2(\omega)}$. Here, $A_1(\omega)e^{i\phi_1(\omega)}$ and $A_2(\omega)e^{i\phi_2(\omega)}$ are the phasors of the true sound pressure appeared in the positions 1 and 2. Accordingly, $M_1(\omega)e^{i\varphi_1(\omega)}$ and $M_2(\omega)e^{i\varphi_2(\omega)}$ are the phasors of the mechanical property of microphones 1 and 2, respectively. The transfer function H_{21}^o that describes the sound pressure ratio of position 2 to position 1 can be defined as $H_{21}^o(\omega) = [A_2(\omega)e^{i\theta_2(\omega)}M_2(\omega)e^{i\varphi_2(\omega)}]/[A_1(\omega)e^{i\theta_1(\omega)}M_1(\omega)e^{i\varphi_1(\omega)}]$. After the two microphones are switched and the second measurement of the same sample is performed, the phasor of sound pressure measured at position 1 through microphone 2 is $A_1(\omega)e^{i\theta_1(\omega)}M_2(\omega)e^{i\varphi_2(\omega)}$, and the phasor of sound pressure measured at position 2 through microphone 1 is $A_2(\omega)e^{i\theta_2(\omega)}M_1(\omega)e^{i\varphi_1(\omega)}$. A_1, A_2, M_1, and M_2 are amplitudes. θ_1, θ_2, φ_1, and φ_2 are phase angles. The sound pressure ratio of position 2 to position 1 after the switching of two microphones is $H_{21}^s(\omega) = [A_2(\omega)e^{i\theta_2(\omega)}M_1(\omega)e^{i\varphi_1(\omega)}]/[A_1(\omega)e^{i\theta_1(\omega)}M_2(\omega)e^{i\varphi_2(\omega)}]$. If H_{21}^o and H_{21}^s are multiplied, we can eliminate the mechanical influences of microphones 1 and 2 and receive the true sound pressure transfer function of position 2 to position 1 as $H_{21} = \sqrt{H_{21}^o H_{21}^s}$. With the use of this automatically calibrated transfer function, the reflection coefficient of the sample can be formulated[7] as

$$R_0(\omega) = \frac{H_{21} - e^{-ik_w d_m}}{e^{-ik_w d_m} - H_{21}} e^{i2k_w l_1}, \quad (27)$$

and the sound absorption coefficient is calculated subsequently ($A_0 = 1 - |R_0|^2$). In Eq. (27), k_w is the wave number of air, d_m is the length between positions 1 and 2, and l_1 is the distance measured from the base of the foam sample to position 1.

V. VALIDATION OF THE THREE-DIMENSIONAL LAPLACE DOMAIN FINITE ELEMENT METHOD

To validate the three-dimensional Laplace domain finite element method (3D LDFEM), the dynamic stiffness of a laterally infinite, finite-thickness, and planar foam backed with a rigid plane with its permeable upper surface subjected to an impulsive pressure excitation, $p_0\delta(t)$, is examined. Results predicted by the 3D LDFEM in comparison with that predicted by the 2D LDFEM[24] as well as the 1D exact results are performed. The 1D exact solution of the Laplace transformed dynamic stiffness $\bar{K}$ of the foam derived in a previous work[13] is rearranged as

$$\bar{K}(s) = \frac{-p_0}{(1-\phi)\bar{\mathbf{u}}|_{z=H} + \phi\bar{\mathbf{U}}|_{z=H}} = \frac{(\bar{P}\bar{R}-\bar{Q}^2)(C_3-C_1)}{[(1-\phi)\Delta_{11}+\phi\Delta_{31}][(1-\phi)(\bar{Q}+\bar{R}C_3)-\phi(\bar{P}+\bar{Q}C_3)]\delta_1\tanh(\delta_1 H)+[(1-\phi)\Delta_{13}+\phi\Delta_{33}][\phi(\bar{P}+\bar{Q}C_1)-(1-\phi)(\bar{Q}+\bar{R}C_1)]\delta_2\tanh(\delta_2 H)}, \tag{28}$$

where

$$\bar{P} = \bar{A} + 2\bar{N},$$

$$C_1 = \frac{-\bar{\rho}_{11}(s)^2 - \bar{b}s + \bar{P}(\delta_1)^2}{\bar{\rho}_{12}(s)^2 - \bar{b}s - \bar{Q}(\delta_1)^2},$$

$$C_3 = \frac{-\bar{\rho}_{11}(s)^2 - \bar{b}s + \bar{P}(\delta_2)^2}{\bar{\rho}_{12}(s)^2 - \bar{b}s - \bar{Q}(\delta_2)^2},$$

$$\delta_1 = \pm\sqrt{\frac{s\psi_4}{2\bar{\beta}_1}}, \quad \delta_2 = \pm s\sqrt{\frac{2\psi_2}{\psi_4}},$$

$$\psi_1 = \bar{\alpha}_1 s + \bar{\beta}_2,$$

$$\psi_2 = \bar{\rho}_0 s + \bar{b}\bar{\rho}, \quad \psi_3 = (\psi_1)^2 - 4\bar{\beta}_1 s(\psi_2), \quad \psi_4 = \sqrt{\psi_3} + \psi_1,$$

$$\alpha_1 = \bar{P}\rho_{22} + \bar{R}\rho_{11} - 2\bar{Q}\rho_{12},$$

$$\beta_1 = \bar{P}\bar{R} - \bar{Q}^2, \quad \beta_2 = (\bar{P} + \bar{R} + 2\bar{Q})\bar{b},$$

$$\bar{\rho}_0 = \bar{\rho}_{11}\bar{\rho}_{22} - \bar{\rho}_{12}^2, \quad \bar{\rho} = \bar{\rho}_{11} + \bar{\rho}_{22} + 2\bar{\rho}_{12},$$

$$\Delta_{11} = \frac{\Delta_1}{\Delta} + \frac{\Delta_2 C_1}{\Delta}, \quad \Delta_{13} = \frac{\Delta_1}{\Delta} + \frac{\Delta_2 C_3}{\Delta},$$

$$\Delta_{31} = \frac{\Delta_3}{\Delta} + \frac{\Delta_4 C_1}{\Delta}, \quad \Delta_{33} = \frac{\Delta_3}{\Delta} + \frac{\Delta_4 C_3}{\Delta},$$

$$\Delta = (\bar{\rho}_0 s + \bar{b}\bar{\rho})s^2,$$

$$\Delta_1 = (\bar{\rho}_{22}s + \bar{b})\bar{P} - (\bar{\rho}_{12}s - \bar{b})\bar{Q},$$

$$\Delta_2 = (\bar{\rho}_{22}s + \bar{b})\bar{Q} - (\bar{\rho}_{12}s - \bar{b})\bar{R},$$

$$\Delta_3 = (\bar{\rho}_{11}s + \bar{b})\bar{Q} - (\bar{\rho}_{12}s - \bar{b})\bar{P},$$

$$\Delta_4 = (\bar{\rho}_{11}s + \bar{b})\bar{R} - (\bar{\rho}_{12}s - \bar{b})\bar{Q}.$$

.

During the numerical comparisons, element models of polyurethane foam of thickness 0.016 m with diameters of 0.0395 m (diameter of small tube) and 0.1268 m (diameter of large tube) subjected to roller circumferential edge constraints are used to simulate the complex dynamic stiffness of the infinite foam 1. The material properties of foam 1 are presented in Table I. The real and imaginary parts of the complex dynamic stiffness predicted by the 2D and 3D LDFEMs for samples of different diameters and the exact solutions predicted from Eq. (28) for the infinite foam 1 are illustrated in Fig. 5. Nodes and elements for the analyses are shown in Table II. Since the constraints on the circumferential edge are assumed as rollers, the change of sample diameter as expected has no noticeable effect on the predictions of the acoustical properties of the infinite foam 1. All predicted results are complete matched with the exact solutions. This validates that the 3D LDFEM is suitable for the predictions of the acoustical properties of polyurethane foams.

VI. INFLUENCE OF EDGE CONSTRAINT ON THE ACOUSTICAL PROPERTY

Numerous researchers reported that if the sample is cut slightly undersize or slightly oversize, then either an air gap is created between the sample and the tube wall, or the circumferential edge of the sample is compressed after being pushed into the tube. Hence, the boundary conditions at the circumferential edge normally cannot be simulated precisely. In this section, the measured and the numerical simulated acoustic results of foams 1–3 (foams with the same constituents but different thicknesses) are shown to present this phenomena and to reveal that a porous material with nonzero Poisson's coefficient, the accuracy of predictions can be greatly enhanced through the three-dimensional simulations. The material properties of foams 1–3 are tabulated in Table I.

The sound absorption coefficients of the samples of diameter 0.0395 m with different thicknesses are shown in Figs. 6–8. In these figures, the exact solutions of foams with infinite extent, experimental data of exact cut (cut by a cylindrical cutter which has the same diameter as the tube) and

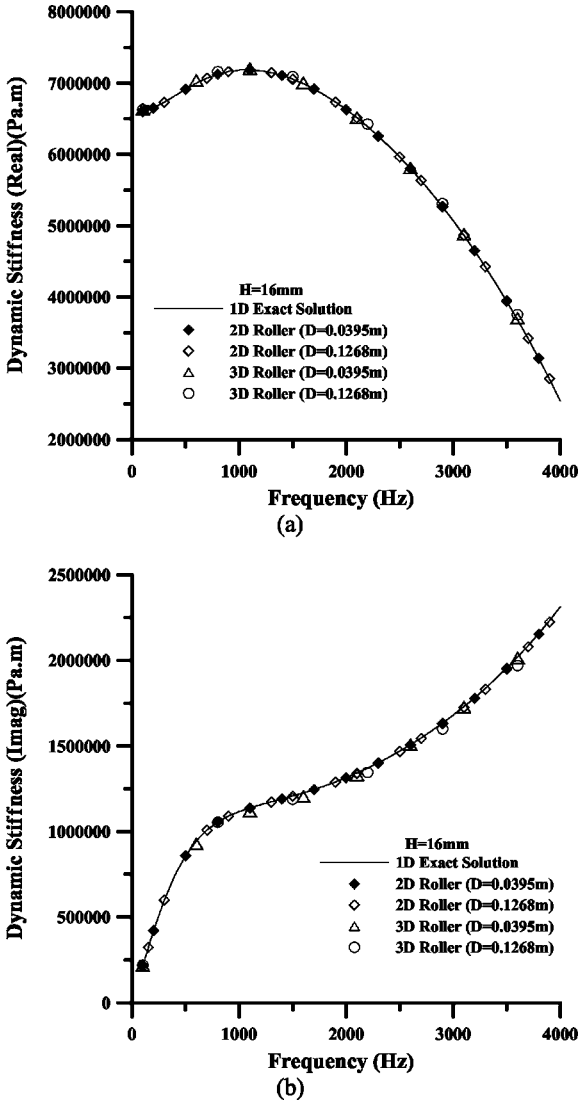


FIG. 5. Dynamic stiffness of foam 1 backed with a rigid wall. (a) Real part. (b) Imaginary part.

slightly undercut samples, and the predictions of samples with fixed as well as roller circumferential edge constraints simulated by the 2D and 3D LDFEMs are presented. It is clearly found that the experimental results are bounded by the predictions of 3D LDFEM with fixed and roller edge constraints. Before the first absorption peak occurred, the sound absorption coefficients of exact cut samples are better than that of undercut samples. The numerical predictions with fixed edge constraints are much closer to the measurements of exact cut samples than the predictions with roller edge constraints. Though the predictions of the 2D LDFEM can generally fit the results of the slightly undercut samples, underestimated predictions are found.

TABLE II. Nodes and elements of the finite element models of foam 1.

	$D=0.0395$ m		$D=0.1268$ m	
	Nodes	Elements	Nodes	Elements
2D LDFEM	377	336	490	432
3D LDFEM	1352	4992	1520	6564

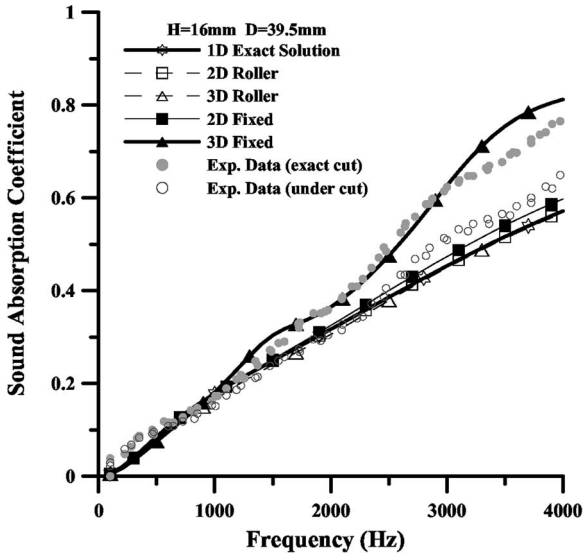


FIG. 6. Sound absorption coefficient of foam 1 backed with a rigid wall.

It is remarked that the unlikely simulations of the sound absorption coefficients reaching unity at low frequencies while reducing the rigidity of the edge constraint from fixed to roller, reported in the earlier study,[19] are not found in the present study. Owing to the influence of the fixed edge constraint, the results of Figs. 6–8 clearly show that the amplitude of the first sound absorption peak is increased and the frequency of this peak is shifted downwards. The sound absorption coefficient measured from a sample can be used to represent that of the infinite planar foam where it was cut from only if the ideal roller edge constraint is applied on the sample during the experiment. This conclusion also applied to the other acoustical properties. For demonstrations, the reflection coefficient, the acoustic impedance, and the dynamic stiffness of samples of foam 2 are presented in Figs. 9–11, respectively. It is further emphasized that if the results are presented in the form of the imaginary part of the normal incidence acoustic impedance, the variations in between the

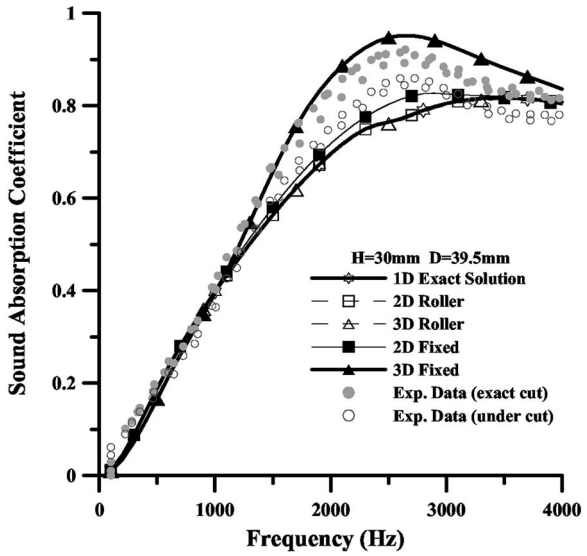


FIG. 7. Sound absorption coefficient of foam 2 backed with a rigid wall.

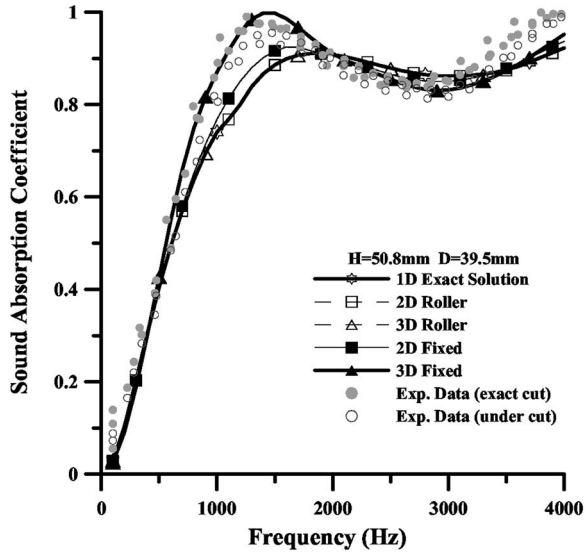


FIG. 8. Sound absorption coefficient of foam 3 backed with a rigid wall.

2D and 3D simulations become less. However, small discrepancies in the low-frequency regions can still be discovered.

VII. INFLUENCE OF SAMPLE DIAMETER ON THE ACOUSTICAL PROPERTY

It was validated that acoustic results of foams with roller edge constraints predicted by the 2D and 3D LDFEMs are completely matched with the 1D exact solutions. While the circumferential edge of a sample is fixed, the rigidity near the edge is increased accordingly. The averaged effect of the increasing of the local rigidity on the acoustical properties of the sample will become less if the diameter-to-thickness ratio (DTR) is raised. For illustration, the sound absorption results for samples of foam 2 of diameters of 0.0395 m (DTR =1.3) and 0.1268 m (DTR=4.2) predicted by the 3D LDFEM with fixed circumferential edge constraints applied, and the one-dimensional exact solutions are shown in Fig. 12. The results clearly show that the more the DTR is increased, the closer the predicted results are to the one-dimensional exact solutions. Generally, as the DTR approaches infinity, the predicted results of the model with fixed edge constraints on the circumferential edge will approach the exact solutions as well as the predicted results of the model with roller edge constraints.

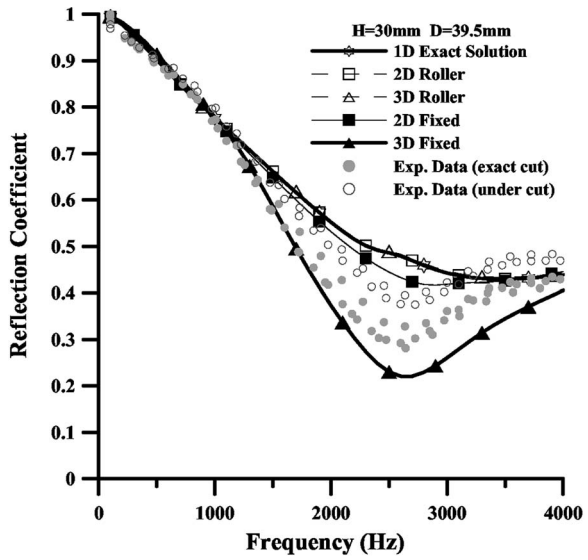


FIG. 9. Reflection coefficient of foam 2 backed with a rigid wall.

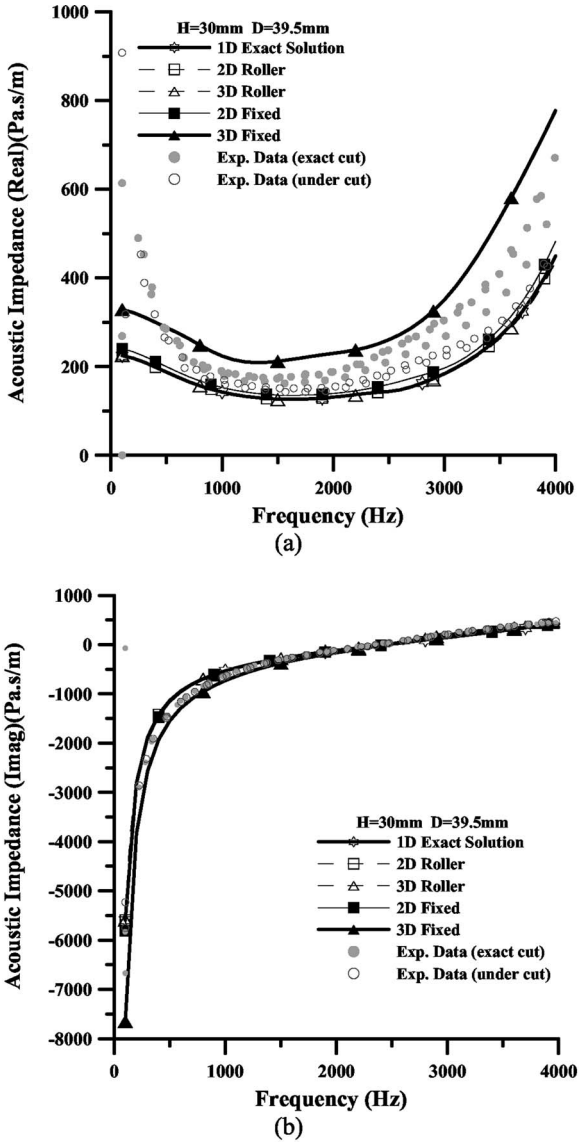
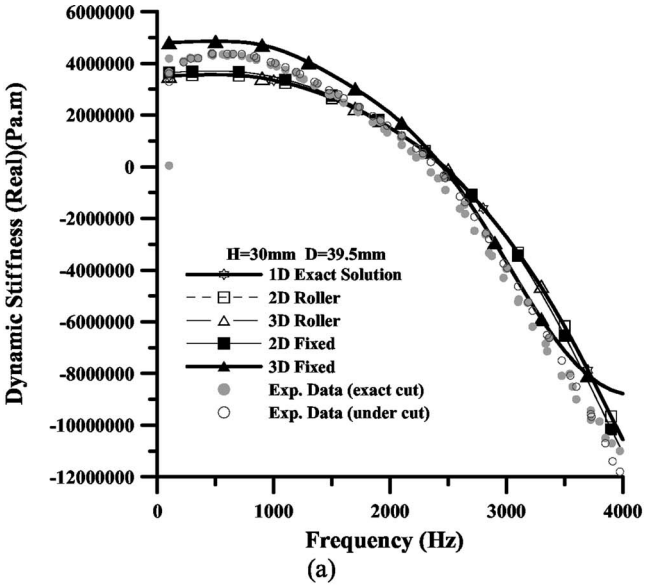


FIG. 10. Acoustic impedance of foam 2 backed with a rigid wall. (a) Real part. (b) Imaginary part.

VIII. CONCLUSIONS

The three-dimensional Laplace domain finite element method, based on Biot's dynamic poroelasticity theory, for simulating the acoustical properties of foams with four-node tetrahedral porous elements was established in this study. Combined with the acoustic measurement system, the influences of circumferential edge constraint and the sample diameter on the foam sample placed in the impedance tube were examined. The experimental and predicted results obtained in this study lead to the following conclusions.



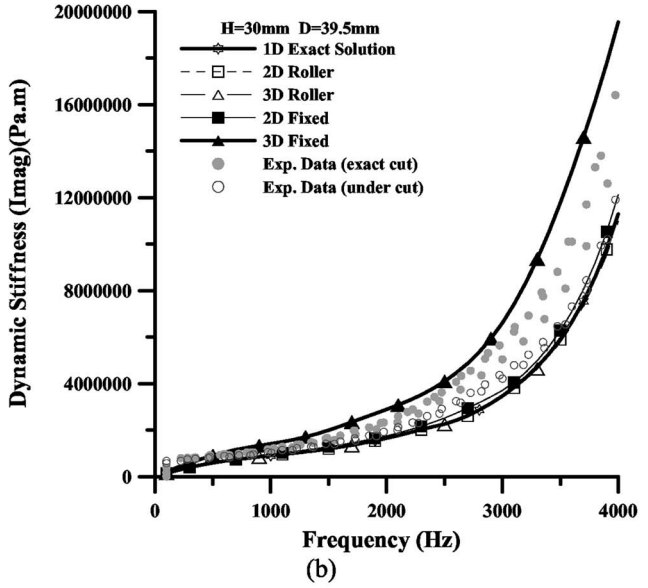


FIG. 11. Dynamic stiffness of foam 2 backed with a rigid wall. (a) Real part. (b) Imaginary part.

Without the use of the incident acoustic field, which was required in the earlier studies, the acoustic behaviors of the foam samples can be efficiently predicted by the three-dimensional Laplace domain finite method. Results predicted by the three-dimensional Laplace domain finite method for samples with roller circumferential edge constraints coincided with the one-dimensional exact solutions. However, in the circumstance of a foam with nonzero Poisson's coefficient and low diameter-to-thickness ratio, the rigidity of the circumferential edge constraint has profound effects on the acoustic predictions of the foam.

While the circumferential edge of the foam is fixed, the amplitude of the first sound absorption peak is increased and the peak is moved to a lower frequency as compared with that predicted with roller constraints. Measurements of samples are bounded by the predictions of samples with roller as well as fixed constraints and are much closer to that predicted with fixed edge constraints. Although the two-dimensional approach can predict the tendency of the acoustic results of samples, underestimations are often found. In summary, the measured acoustic results can be used to represent that of the laterally infinite planar foam from which the sample was cut only if the roller edge constraint is managed during the experiment.

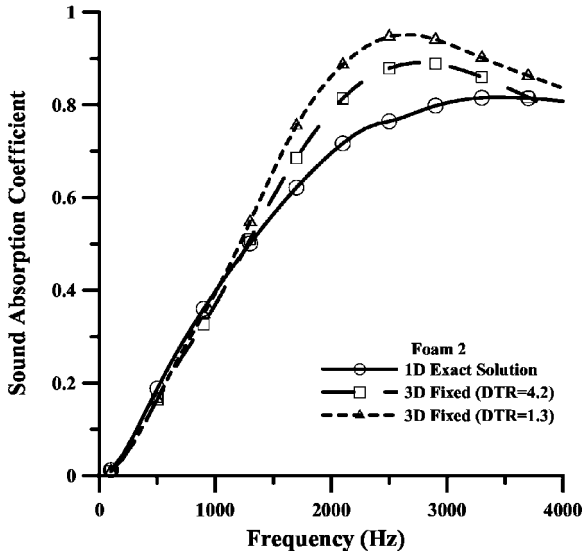


FIG. 12. Influence of sample diameter on the sound absorption coefficient.

The influence of the fixed circumferential edge constraint can be covered up by the increase of the diameter-to-thickness ratio. The more the diameter-to-thickness ratio is increased, the closer the predicted results are to that of the sample with roller edge constraints. When the diameter-to-thickness ratio approaches infinity, the predicted results of the sample with fixed edge constraints and that of the sample with roller edge constraints both approach the one-dimensional exact solutions.

[1]J. F. Allard, C. Depollier, and W. Lauriks, "Measurement and prediction of source impedance at oblique incidence of a plastic foam of high flow resistivity," J. Sound Vib. **132**(1), 51–60 (1989).
[2]W. Lauriks, A. Cops, J. F. Allard, C. Depollier, and P. Rebillard, "Modelization at oblique incidence of layered porous materials with impervious screens," J. Acoust. Soc. Am. **87**, 1200–1206 (1990).
[3]ASTM Standard C423–02a, "Standard Test Method for Sound Absorption and Sound Absorption Coefficients by the Reverberation Room Method" (American Society of Testing and Materials, Philadelphia, 2002).
[4]ASTM Standard C384–04, "Standard Test Method for Impedance and Absorption of Acoustical Materials by Impedance Tube Method" (American Society of Testing and Materials, Philadelphia, 2004).
[5]ASTM Standard E1050-98, "Standard Test Method for Impedance and Absorption of Acoustical Materials Using a Tube, Two Microphones, and a Digital Frequency Analysis System" (American Society of Testing and Materials, Philadelphia, 1998).
[6]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. I. Theory," J. Acoust. Soc. Am. **68**, 907–913 (1980).
[7]J. Y. Chung and D. A. Blaser, "Transfer function method of measuring in-duct acoustic properties. II. Experiment," J. Acoust. Soc. Am. **68**, 914–921 (1980).
[8]M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[9]N. Voronina, "An empirical model for rigid frame porous materials with high porosity," Appl. Acoust. **51**(2), 181–198 (1997).
[10]M. A. Biot, "Theory of elasticity and consolidation for a porous anisotropic solid," J. Appl. Phys. **26**(2), 182–185 (1955).
[11]M. A. Biot, "Theory of propagation of elastic waves in a fluid-saturated

porous solid I. low-frequency range, II: higher frequency range," J. Acoust. Soc. Am. **28**, 168–191 (1956).
[12]M. A. Biot and D. G. Willis, "The elastic coefficients of the theory of consolidation," J. Appl. Mech. **24**, 594–601 (1957).
[13]H.-S. Tsay and F.-H. Yeh, "Frequency response function for prediction of planar cellular plastic foam acoustic behavior," J. Cell. Plast. **41**(2), 101–131 (2005).
[14]A. Craggs, "A finite element model for rigid porous absorbing materials," J. Sound Vib. **61**(1), 101–111 (1978).
[15]A. Craggs, "Coupling of finite element acoustic absorption models," J. Sound Vib. **66**(4), 605–613 (1979).
[16]V. Easwaran and M. L. Munjal, "Finite element analysis of wedges used in anechoic chambers," J. Sound Vib. **160**(2), 333–350 (1993).
[17]W.-H. Chen, F.-C. Lee, and D.-M. Chiang, "On the acoustic absorption of porous materials with different surface shapes and perforated plates," J. Sound Vib. **237**(2), 337–355 (2000).
[18]Y. J. Kang and J. S. Bolton, "Finite element modeling of isotropic elastic porous materials coupled with acoustical finite elements," J. Acoust. Soc. Am. **98**, 635–643 (1995).
[19]B. H. Song, J. S. Bolton, and Y. J. Kang, "Effect of circumferential edge constraint on the acoustical properties of glass fiber materials," J. Acoust. Soc. Am. **110**, 2902–2916 (2001).
[20]B. H. Song and J. S. Bolton, "Investigation of the vibrational modes of edge-constrained fibrous samples placed in a standing wave tube," J. Acoust. Soc. Am. **113**, 1833–1849 (2003).
[21]J. S. Bolton, N. M. Shiau, and Y. J. Kang, "Sound transmission through multi-panel structures lined with elastic porous materials," J. Sound Vib. **191**(3), 317–347 (1996).
[22]A. Cummings, "Impedance tube measurements on porous media: The effect of air-gaps around the sample," J. Sound Vib. **151**(1), 63–75 (1991).
[23]T. E. Vigran, L. Kelders, W. Leclaire, and T. F. Johansen, "Prediction and measurements of the influence of boundary conditions in a standing wave tube," Acust. Acta Acust. 83, 419–423 (1997).
[24]F.-H. Yeh and H.-S. Tsay, "Dynamic behavior of a poroelastic slab subjected to uniformly distributed impulsive loading," Comput. Struct. **67**, 267–277 (1998).
[25]H.-S. Tsay and F.-H. Yeh, "Finite element frequency-domain acoustic analysis of open-cell plastic foams," Finite Elem. Anal. Design **42**(4), 314–339 (2006).

Piezoelectric vibration control by synchronized switching on adaptive voltage sources: Towards wideband semi-active damping

A. Badel
Institut National des Sciences Appliquées de Lyon, Laboratoire de Génie Electrique et Ferroélectricité, Villeurbanne 69621, France

G. Sebald
Intelligent Systems Laboratory, Institute of Fluid Science, Tohoku University, Sendai 980-8577, Japan

D. Guyomar, M. Lallart, E. Lefeuvre, and C. Richard
Institut National des Sciences Appliquées de Lyon, Laboratoire de Génie Electrique et Ferroélectricité, Villeurbanne 69621, France

J. Qiu
Intelligent Systems Laboratory, Institute of Fluid Science, Tohoku University, Sendai 980-8577, Japan

(Received 19 May 2005; revised 1 February 2006; accepted 9 February 2006)

Synchronized switch damping (SSD) principle and derived techniques have been developed to address the problem of structural damping. Compared with standard passive piezoelectric damping, these new semi-passive techniques offer the advantage of self-adaptation with environmental variations. Unlike active damping systems, their implementation does not require any sophisticated signal processing nor any bulky power amplifier. This paper presents an enhancement of the SSD technique on voltage source (SSDV) which is the most effective of the SSD techniques. The former SSDV technique uses a constant continuous voltage sources whereas the proposed enhancement uses an adaptive continuous voltage source which permits fitting the mechanical braking force resulting from the SSDV process to the vibration level. A theoretical analysis of the SSDV techniques is proposed. Experimental results for structural damping under single frequency and for vibration control of a smart board under white noise excitation are presented and confirm the interest of the enhanced SSDV compared to other SSD techniques. Depending on the excitation type, a 4- to 10-dB damping gain can be achieved. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2184149]

PACS number(s): 43.55.Wk, 43.55.Rg, 43.50.Ki, 43.38.Fx [KAC] Pages: 2815–2825

I. INTRODUCTION

Several semi-passive vibration control techniques have been recently developed. The state switching method proposed by Clark (2000) is a variable stiffness technique in which piezoelements are periodically held in the open-circuit state, then switched and held in the short-circuit state, synchronously with the structure motion. Cunefare (2002) successfully adapted previous works of Larson *et al.* (1998) on the state-switching concept to vibration damping in a rather different way and these were proposed as the state-switching absorber (SSA).

This paper deals with particular nonlinear techniques which are known as SSD for synchronized switch damping (Richard, 1998, 2000). These techniques consist in a nonlinear processing of the voltage on a piezoelement bonded on the structure. It is implemented with a simple switch synchronously driven during short periods with the structure motion. The switch connects the piezoelement to a circuit which can be either a simple short circuit (SSDS), a small inductor (SSDI), or voltage sources (SSDV). Due to this process, a voltage magnification is obtained and a phase shift appears between the piezoelement strain and the resulting voltage, thus creating energy dissipation. This method offers several advantages: It is insensitive to environmental changes due to its self-adaptive broadband behavior and it requires only a very low external power for the switch control and the voltage source in the case of the SSDV.

It has been shown that dissipated energy depends in particular on the voltage amplitude across the electrodes of piezoelectric elements. The role of the additional constant voltage sources in the SSDV technique is thus to increase the voltage amplitude, which increases the damping effect. In the original SSDV technique, the sign of the continuous voltage source is changed so that it increased the piezovoltage during the inversion process (Lefeurve, 2005). However, its absolute value remains constant, which can lead to stability problems. Actually, when the vibration level is very low, the original SSDV approach may excite the structure instead of limiting its vibrations. The enhancement proposed in this paper consists in fitting the continuous voltage amplitude to the vibration level. SSD techniques are presented and theoretically developed in Sec. II. Section III presents experimental results under single frequency excitation. Vibration damping experiment under white noise excitation of a plate structure

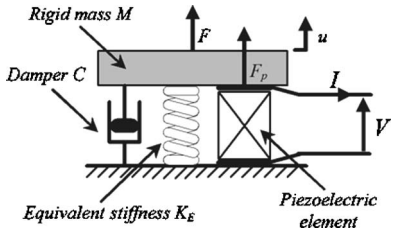


FIG. 1. Schematic representation of the electromechanical model.

exhibiting numerous modes on a 1-kHz frequency band is presented in Sec. IV and illustrates the broadband capability of the proposed enhanced SSDV technique.

II. SSD TECHNIQUES BASIS

This section briefly presents the different SSD techniques. For clarity reasons, a sinusoidal excitation is considered. In this simple case, if the structure is driven around one of its resonance frequency, it can be modeled using a second-order model.

A. Modeling

A mechanical model based on a spring-mass system having only one degree of freedom gives a good description of vibrating structure behavior near a resonance. The differential equation (1) is established assuming that the global structure including piezoelectric elements is linearly elastic. In this equation, M represents the equivalent rigid mass and C is the mechanical losses coefficient. K_E is the equivalent stiffness of the electromechanical structure when the piezoelectric elements are short-circuited, u is the rigid mass displacement and ΣF_i represents the sum of other forces applied to the equivalent rigid mass, comprising forces applied by piezoelectric elements. Piezoelectric elements bonded on the considered structure ensure the electromechanical coupling described by the equations (2) and (3), where F_p is the electrically dependent part of the force applied by piezoelectric elements on the structure, C_0 is the blocked capacitance of piezoelectric elements and α is the force factor. I is the outgoing current from piezoelectric elements. M, C_0, α, and K_E can be deduced from piezoelectric elements and structure characteristics and geometry, or starting from a few measurements, as shown in Sec. III A. A schematic representation of the model is given by Fig. 1:

$$M\ddot{u} + C\dot{u} + K_E u = \sum F_i, \tag{1}$$

$$F_p = -\alpha V, \tag{2}$$

$$I = \alpha\dot{u} - C_0\dot{V}. \tag{3}$$

Finally, ΣF_i applied to the rigid equivalent mass comprises F_p and an external excitation force F. Thus, the differential equation governing the mass motion can be written as Eq. (4). The energy equation (5) is obtained by multiplying both sides of Eq. (4) by the velocity and integrating over the time variable. The provided energy is divided into kinetic energy, potential elastic energy, mechanical losses, and transferred energy. The transferred energy corresponds to the part of the mechanical energy which is converted into electrical energy. Maximizing this energy amounts to minimizing the mechanical energy in the structure (kinetic + elastic):

$$M\ddot{u} + C\dot{u} + K_E u = F - \alpha V, \tag{4}$$

$$\int F\dot{u}\,dt = \frac{1}{2}M\dot{u}^2 + \frac{1}{2}K_E u^2 + \int C\dot{u}^2\,dt + \int \alpha V\dot{u}\,dt. \tag{5}$$

B. Open-circuit condition

It is considered that nothing is connected to the piezoelectric elements; the transferred energy is then clearly equal to zero. The current flowing out the equivalent piezo is null and then (3) written in the frequency domain leads to Eq. (6). Inserting Eq. (6) in Eq. (4) written in the frequency domain leads to the transfer function (7) linking the displacement to the driving force:

$$\tilde{V} = \frac{\alpha}{C_0}\tilde{u}, \tag{6}$$

$$\frac{\tilde{u}}{\tilde{F}} = \frac{1}{\left(K_E + \alpha^2/C_0 - M\omega^2 + jC\omega\right)}. \tag{7}$$

It is assumed that at the resonance frequency the force F and the speed $\dot{u}$ are in phase (this is a good approximation for structures with low viscous losses). The resonance angular frequency is then given by Eq. (8) and the amplitude of the displacement by Eq. (9) where F_M is this amplitude of the driving force:

$$\omega_0 = \sqrt{\frac{K_E + \alpha^2/C_0}{M}}, \tag{8}$$

$$u_M = \frac{F_M}{C\omega_0}. \tag{9}$$

C. SSDI technique

The SSDI technique consists of adding a switching device in parallel with the piezoelectric elements. This device is composed of a switch and an inductance L_I connected in series, as shown in Fig. 2. The switch is almost always open, except when a voltage extremum occurs, i.e., when a strain extremum occurs in the piezoelement. At this instant, the switch is closed. The capacitance C_0 of the piezoelectric elements and the inductance L_I then constitute oscillator. The switch S_W is kept closed until the voltage V on the piezoelectric elements has been reversed. It corresponds to a time t_i equal to a half period of the electric oscillator, as shown in Eq. (10):

$$t_i = \pi\sqrt{L_I C_0}. \tag{10}$$

When the switch is open and if no load is connected, the outgoing piezocurrent is null and then the voltage and the

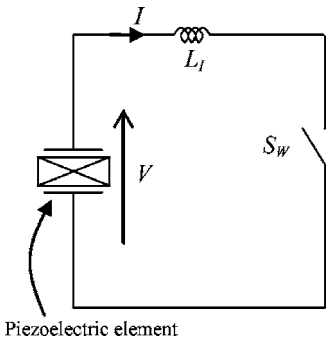


FIG. 2. SSDI electrical circuit.

displacement vary proportionally. The voltage and displacement typical waveforms are shown in Fig. 3.

The voltage inversion is not perfect, because a part of the energy stored on the piezoelectric element's capacitance is lost in the switching network (electronic switch +inductance). These losses are modeled by an electrical quality factor Q_I. The relation between Q_I and the voltage of the piezoelectric element before and after the inversion process is given by Eq. (11):

$$V_m = \gamma V_M = V_M e^{-\pi/2Q_I}. \tag{11}$$

The transferred energy can be expressed as the following integral function of voltage V and displacement u:

$$\int \alpha V \dot{u}\, dt = \int \alpha V\, du. \tag{12}$$

Considering a periodic motion of the structure, transferred energy can then be represented as the cycle shown in Fig. 4. Two open circuit phases and two voltage inversions phases can be distinguished on this cycle. In open circuit, voltage variations are proportional to displacement with a factor α/C_0, whereas there is no displacement variation during the voltage inversions, since the inversion time is much smaller than the mechanical motion period. The surface circumscribed by this cycle is the image of transferred energy with a factor $1/\alpha$. Between two voltage inversions, the voltage and the displacement vary proportionally, which leads to Eq. (13) where u_M is the displacement amplitude. The transferred energy E_t during a period is given by Eq. (14):

$$V_M = V_m + \frac{2\alpha}{C_0} u_M, \tag{13}$$

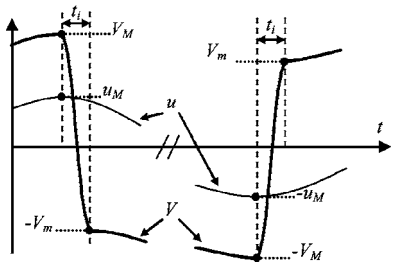


FIG. 3. Displacement and voltage typical waveforms.

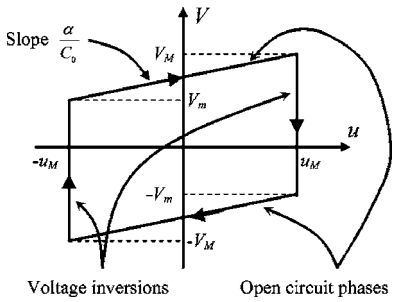


FIG. 4. Energy cycle.

$$E_t = \int_{\text{period}} \alpha V\, du = 2\alpha u_M (V_m + V_M). \tag{14}$$

E_t can be expressed as a unique function of the displacement amplitude using Eqs. (11), (13), and (14):

$$E_t = \frac{4\alpha^2}{C_0} \frac{1+\gamma}{1-\gamma} u_M^2. \tag{15}$$

It is assumed that the displacement remains purely sinusoidal. If the driving force angular frequency is ω, the displacement can be expressed by Eq. (16). The piezovoltage can be expressed as the sum of two functions, one which is an image of the displacement and one of which is a crenel function. This is summarized in Eq. (17), where h is the crenel function defined in Eq. (18). In the case of a periodic motion, h is also periodic and can be expressed as Fourier series, as shown in Eq. (19).

$$u(t) = u_M \sin(\omega t + \varphi), \tag{16}$$

$$V = \frac{\alpha}{C_0}(u + h), \tag{17}$$

$$h = u_M \frac{1+\gamma}{1-\gamma} \operatorname{signe}(\dot{u}), \tag{18}$$

$$h = u_M \frac{1+\gamma}{1-\gamma} \sum_{n=0}^{\infty} \left(\frac{4}{\pi(2n+1)} \cos((2n+1)\omega t + \varphi) \right). \tag{19}$$

Inserting Eq. (17) in Eq. (4) leads to Eq. (20). Since the displacement is assumed to remain sinusoidal, only the first harmonic of the crenel can be considered, which leads to the transfer function (21) linking the displacement to the driving force:

$$F = M\ddot{u} + \left(K_E + \frac{\alpha^2}{C_0}\right)u + C\dot{u} + \frac{\alpha^2}{C_0}h, \tag{20}$$

$$\frac{\tilde{u}}{\tilde{F}} = \frac{1}{\left(K_E + \alpha^2/C_0 - M\omega^2 + j\left(C\omega + (4\alpha^2/\pi C_0)(1+\gamma)/(1-\gamma)\right)\right)}. \tag{21}$$

It is assumed that at the resonance frequency the force F and the speed $\dot{u}$ are in phase. The resonance angular frequency is then given by Eq. (8) and the amplitude of the displacement by Eq. (22). Equation (23) gives the theoretical value of the SSDI damping

$$u_M = \frac{F_M}{C\omega_0 + (4\alpha^2/\pi C_0)(1+\gamma)/(1-\gamma)}, \tag{22}$$

$$A_{\mathrm{SSDI}} = 20\log\left(\frac{C\omega_0}{C\omega_0 + (4\alpha^2/\pi C_0)(1+\gamma)/(1-\gamma)}\right) \tag{23}$$

D. Original SSDV technique (Lefeuvre, 2006)

The electrical circuit of the original SSDV semi-passive damping technique is represented in Fig. 5. The difference between this circuit and the previously presented SSDI circuit is that the piezoelectric element is switched on a positive or a negative voltage source across the inductive shunt circuit. It has been shown that dissipated energy depends in particular on the voltage amplitude across the electrodes of piezoelectric elements. The role of these additional constant voltage sources is to increase the piezovoltage amplitude, thus increasing the damping effect.

The control strategy of the switches consists in closing S_{W1} when a maximum of the strain in the piezoelectric elements occurs, corresponding to a maximum of the voltage V. Closing S_{W1} starts a pseudo-periodic voltage oscillation, until S_{W1} is opened. The switch is kept closed during a time t_i equal to a half the pseudo-period of the electrical oscillator composed by the inductance and the capacitance of the piezoelectric elements. Thus, the voltage sign of V become negative when S_{W1} is opened.

The switch S_{W2} is controlled in a symmetrical way to S_{W1}: it is closed when a minimum of displacement strain occurs, corresponding to a minimum of the voltage V. This switch is then opened after a closing time t_i. Thus, the sign of V becomes positive. Theoretical corresponding waveforms of the strain and the voltage V are plotted in Fig. 6.

The transferred energy E_t during a period can be expressed as in the previous case by Eq. (14), since the energy cycle described in Fig. 4 also corresponds to the SSDV displacement and voltage waveforms. The voltage variation between two inversions still leads to Eq. (13) but the voltage inversion across the voltage source V_{cc} leads to Eq. (24). E_t can then be expressed as a function of u_M and V_{cc}, as shown by Eq. (25):

$$V_m - V_{cc} = \gamma(V_M + V_{cc}), \tag{24}$$

$$E_t = \left(4\frac{\alpha^2}{C_0}u_M^2 + 4\alpha u_M V_{cc}\right)\frac{1+\gamma}{1-\gamma}. \tag{25}$$

It is clearly shown that the SSDV technique increases the uV cycle along the voltage axis, which results in increasing the transferred energy.

As in the previous case, it is assumed that the displacement remains purely sinusoidal, thus Eqs. (16) and (17) are still valid. The crenel function h is now given by Eqs. (26) and (27) using Fourier series:

$$h = \frac{1+\gamma}{1-\gamma}\left(u_M + \frac{C_0}{\alpha}V_{cc}\right)\operatorname{signe}(\dot{u}), \tag{26}$$

$$h = \frac{1+\gamma}{1-\gamma}\left(u_M + \frac{C_0}{\alpha}V_{cc}\right) \times \sum_{n=0}^{\infty}\left(\frac{4}{\pi(2n+1)}\cos\left((2n+1)\omega t + \varphi\right)\right). \tag{27}$$

As in the previous case, only the first harmonic of the crenel can be considered, which leads to Eq. (28). For a given angular frequency ω, the displacement magnitude u_M as well as

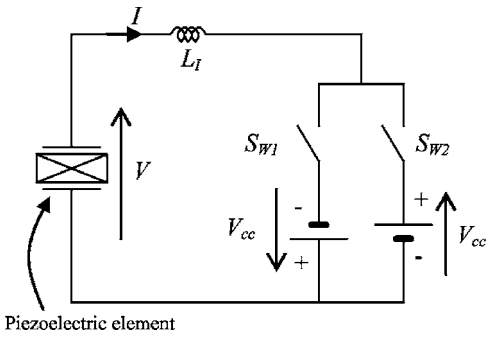


FIG. 5. Original SSDV electrical circuit.

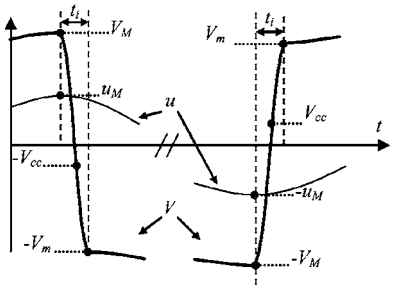


FIG. 6. Displacement and voltage typical waveforms.

its phase difference φ compared to the driving force can be obtained solving the system (29):

$$u_M \sin(\omega t+\varphi)\left(K_E+\frac{\alpha^2}{C_0}-\omega^2 M\right)+\cos(\omega t+\varphi) \times\left(u_M\left(C\omega+\frac{4}{\pi}\frac{\alpha^2}{C_0}\frac{1+\gamma}{1-\gamma}\right)+\frac{4}{\pi}\frac{1+\gamma}{1-\gamma}\alpha V_{cc}\right) = F_M \sin(\omega t), \tag{28}$$

$$u_M\left(K_E+\frac{\alpha^2}{C_0}-\omega^2 M\right)\sin\varphi + \left(u_M\left(C\omega+\frac{4}{\pi}\frac{\alpha^2}{C_0}\frac{1+\gamma}{1-\gamma}\right)+\frac{4}{\pi}\frac{1+\gamma}{1-\gamma}\alpha V_{cc}\right)\cos\varphi=0,$$

$$u_M\left(K_E+\frac{\alpha^2}{C_0}-\omega^2 M\right)\cos\varphi+\left(u_M\left(C\omega+\frac{4}{\pi}\frac{\alpha^2}{C_0}\frac{1+\gamma}{1-\gamma}\right)+\frac{4}{\pi}\frac{1+\gamma}{1-\gamma}\alpha V_{cc}\right)\sin\varphi=F_M. \tag{29}$$

At the resonance $\varphi=\pi/2$ and then the angular resonance frequency is still given by Eq. (8). In this case, the displacement amplitude can be expressed as a function of the force magnitude and of the continuous voltage, as shown by Eq. (30). The theoretical value of the SSDV damping is given by Eq. (31):

$$u_M=\frac{F_M(4/\pi)[(1+\gamma)/(1-\gamma)]\alpha V_{cc}}{C\omega_0+(4\alpha^2/\pi C_0)(1+\gamma)(1-\gamma)}, \tag{30}$$

$$A_{\mathrm{SSDV}}=20\log\left(\frac{C\omega_0}{C\omega_0+(4\alpha^2/\pi C_0)(1+\gamma)/(1-\gamma)} \times\left(1-\frac{4}{\pi}\frac{1+\gamma}{1-\gamma}\frac{\alpha V_{cc}}{F_M}\right)\right). \tag{31}$$

This equation shows that, under a given excitation force, a value- of the voltage V_{cc} that totally cancels the vibration can be found. The expression of this particular value of the voltage is given in Eq. (32). In this case, the force applied by the piezoelectric element on the structure, which is proportional to V_{cc}, totally compensates the effect of external excitation force. Applying a voltage higher than $V_{cc\,\max}$ leads to stability problems because in this case the voltage comes to have a driving instead of a braking effect. Moreover, when V_{cc} approaches this critical value, the previous theoretical development is no longer valid due to the displacement higher harmonics that cannot be neglected any longer compared to the fundamentals:

$$V_{cc\,\max}=\frac{\pi}{4\alpha}\frac{1-\gamma}{1+\gamma}F_M. \tag{32}$$

E. Enhanced SSDV technique

The enhancement proposed in this paper consists in adapting the voltage in order to fit the piezoelectric force due to the SSDV process to the vibration amplitude. This can be basically done by forcing the continuous voltage V_{cc} to be of opposite sign and proportional to the strain. The strain can be monitored using the voltage on an additional piezoelement bonded in the same area as the patches used for the vibration control. Considering the simple model previously described, the voltage on an additional piezoelement is given by Eq. (33), and V_{cc} by Eq. (34), where β is proportional coefficient set by the user. The electrical circuit of the enhanced SSDV is a represented in Fig. 7. As in the SSDI technique, the control strategy of the switch consists in closing S_w when an extremum of strain in the piezoelectric elements occurs, corresponding to an extremum of the voltage V. S_W is open after the voltage inversion. In the case of a sinusoidal steady-state excitation, the voltage V_{cc} during the voltage inversion is given by Eq. (35). Theoretical waveforms of the strain and the voltage V are similar to the ones plotted in Fig. 6:

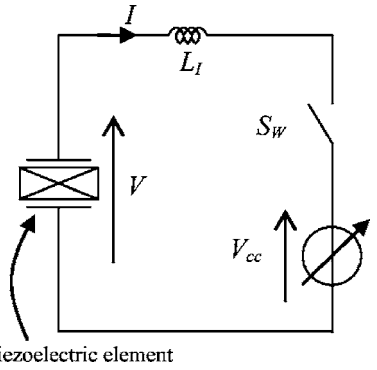


FIG. 7. Enhanced SSDV electrical circuit.

$$V_s=\frac{\alpha}{C_0}u, \tag{33}$$

$$V_{cc}=-\beta\frac{\alpha}{C_0}u, \tag{34}$$

$$V_{cc}=-\beta\frac{\alpha}{C_0}u_M \quad \text{on strain maximum},$$
$$V_{cc}=\beta\frac{\alpha}{C_0}u_M \quad \text{on strain minimum}. \tag{35}$$

The transferred energy E_t during a period can be expressed as in previous cases by Eq. (14), since the energy cycle described in Fig. 4 is still valid. The voltage variation between two inversions still leads to Eq. (13) and the voltage inversion across the voltage source V_{cc} leads to Eq. (36). E_t can then be expressed as a function of u_M, as shown by Eq. (37).

$$V_m-\beta\frac{\alpha}{C_0}u_M=\gamma\left(V_M+\beta\frac{\alpha}{C_0}u_M\right), \tag{36}$$

$$E_t=\frac{4\alpha^2}{C_0}(1+\beta)\frac{1+\gamma}{1-\gamma}u_m^2. \tag{37}$$

As the original SSDV technique, the enhanced SSDV increases the uV cycle along the voltage axis, which results in increasing the transferred energy.

As in previous cases, it is assumed that the displacement remains purely sinusoidal, thus Eqs. (16) and (17) are still

valid. The crenel function h is now given by Eqs. (38) and (39) using Fourier series. It is shown that, as in the SSDI case, the crenel function is proportional to the displacement amplitude, which is not the case with the original SSDV.

$$h = u_M(1+\beta)\frac{1+\gamma}{1-\gamma}\,\mathrm{signe}(\dot{u}), \tag{38}$$

$$h = u_M(1+\beta)\frac{1+\gamma}{1-\gamma}\sum_{n=1}^{\infty}\left(\frac{4}{\pi(2n+1)}\cos((2n+1)\omega t+\varphi)\right). \tag{39}$$

As previously, only the first harmonic of the crenel can be considered. Since the crenel function is very similar to the crenel function related to the SSDI technique, the transfer function linking the displacement to the driving force can easily be deduced and is given by Eq. (40). At the resonance frequency, still given by Eq. (8), the amplitude of the displacement is given by Eq. (41) and the related damping by Eq. (42):

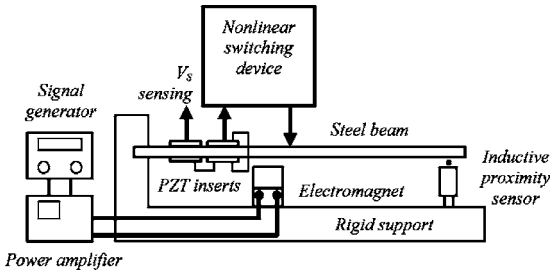


FIG. 8. Experimental setup.

$$\frac{\tilde{u}}{\tilde{F}} = \frac{1}{\left(K_E + (\alpha^2/C_0) - M\omega^2 + j\left(C\omega + (4\alpha^2/\pi C_0)(1+\beta)(1+\gamma)/(1-\gamma)\right)\right)}, \tag{40}$$

$$u_M = \frac{F_M}{C\omega_0 + (4\alpha^2/\pi C_0)(1+\beta)(1+\gamma)/(1-\gamma)}, \tag{41}$$

$$A_{\mathrm{SSDV_{enh}}} = 20\log\left(\frac{C\omega_0}{C\omega_0 + (4\alpha^2/\pi C_0)(1+\beta)(1+\gamma)/(1-\gamma)}\right). \tag{42}$$

This equation shows that, for a given parameter β, the damping is not sensitive to the amplitude of the applied force. This is the critical point of the SSDV enhancement. Using the original SSDV technique with a given continuous voltage amplitude leads to stability problems when reducing the driving force amplitude and to a lack of effectiveness when increasing it. These drawbacks are efficiently overpassed by the proposed enhancement. Actually, the enhanced SSDV technique is a closed loop process where the piezoelectric force is adjusted to the vibration amplitude, whereas the original SSDV is closer to an open loop process.

It must be noted that for high values of β, the previous theoretical development is no longer valid because the displacement higher harmonics cannot be neglected any longer compared to the fundamental.

III. EXPERIMENTS UNDER SINGLE FREQUENCY EXCITATION

A. Experimental setup

The experimental test structure is a cantilever beam clamped at one end in a rigid structure, as shown in Fig. 8. The beam vibrations are driven with an electromagnet that generates a periodic force. The piezoelectric ceramic inserts are bonded on the beam close to the clamped end where the bending radius is at minimum. The poling direction of the piezoelectric inserts is perpendicular to the plate and it is the lateral coupling k_{31} which mainly drives the piezoelectric response. The measurement of the cantilever free end displacement is obtained with an inductive proximity sensor. The characteristics of the beam and the piezoelectric inserts are summarized in Tables I and II II, respectively.

TABLE I. Steel beam characteristics.

Length L	180 mm
Width	90 mm
Thickness	2.05 mm

B. Model identification

The considered piezoelectric structure is more complex than the simple resonator described in Sec. II A. The various parameters cannot be simply obtained from the properties of the piezoelectric inserts. However, the structure presents all the features of a piezoelectric electromechanical device as defined in Sec. II A. Its properties can therefore be identified using the experimental measurements defined in Table III and Eqs. (43). The identified model corresponds then to the

TABLE II. Piezoelectric P189 characteristics.

Length L_P	60 mm
Width	90 mm
Thickness	300 μm
Location x_P	14 mm

TABLE III. Experimental measurements.

f_0	Short circuit resonance frequency
f_1	Open circuit resonance frequency
ξ	Open circuit damping coefficient
λ	Piezo open circuit voltage to the beam free end displacement ratio
C_0	Clamped capacitance of the piezoelectric elements

actual setup around the considered resonance frequency which corresponds, in this case, to the first flexural mode. Numerical data corresponding to the experimental setup are summarized in Table IV:

$$\alpha = \lambda C_0, \quad K_E = \alpha\lambda \frac{f_0^2}{f_1^2 - f_0^2}, \quad M = \frac{K_E}{4\pi^2 f_0^2},$$
$$C = 4\pi\xi M f_1. \tag{43}$$

C. Results

The beam is driven around its first resonance frequency. The driving force amplitude is 0.086N. Without control and at the resonance frequency the displacement amplitude of the beam free end is 2.2 mm. Both theoretical and experimental normalized displacement amplitudes in dB are plotted in Fig. 9, versus the driving force frequency for the different techniques. The inversion quality factor Q_I is 2.1. It was artificially decreased adding a 1-kΩ resistance in series with the inductor for clarity reasons to highlight the damping enhancement due to the SSDV voltage magnification. For the original SSDV technique V_{cc} is set to 3 V and for the enhanced SSDV β is set to 0.95. Theoretical displacements are calculated using Eqs. (7), (21), (29), and (40).

TABLE IV. Measurements and model parameters (beam).

f_0	56.06 Hz
f_1	56.30 Hz
ξ	0.0036
λ	12500 V/m
C_0	238 nF
α	0.0030 N/V
k^2	0.0086
K_E	4330 Nm^{-1}
M	34.9 g
C	0.0883 Nm^{-1} s^{-1}

V_{cc} and β were chosen so that the original and the enhanced SSDV techniques lead to the same displacement amplitude at the resonance frequency. The original and enhanced SSDV harmonic responses are shown to be very close. The original SSDV technique is slightly more effective out of the resonance frequency because in the case of the enhanced SSDV, V_{cc} decreases when moving away from the resonance frequency.

Experimental displacement and voltage waveforms are

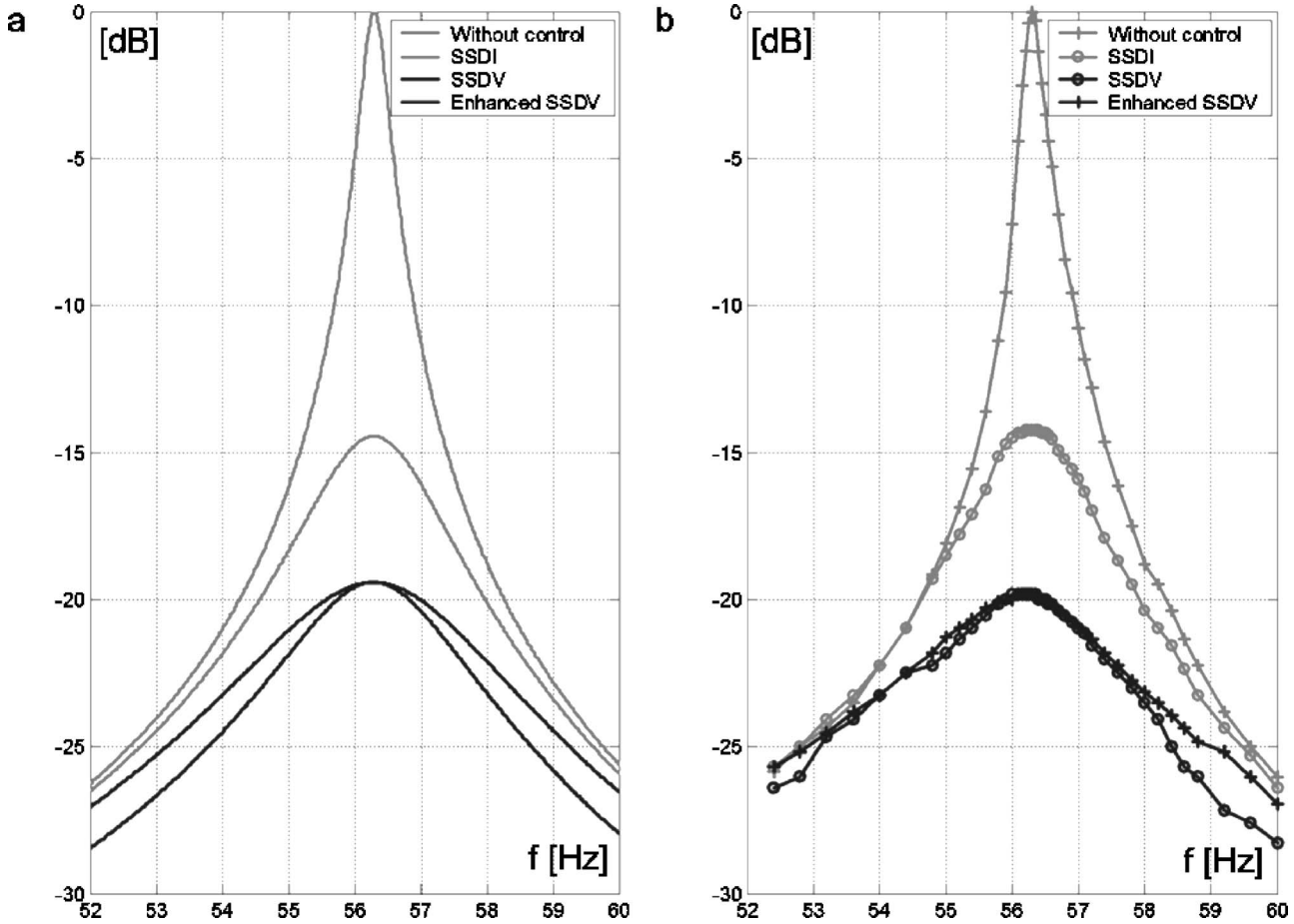


FIG. 9. Theoretical (a) and experimental (b) normalized harmonic responses for the different techniques.

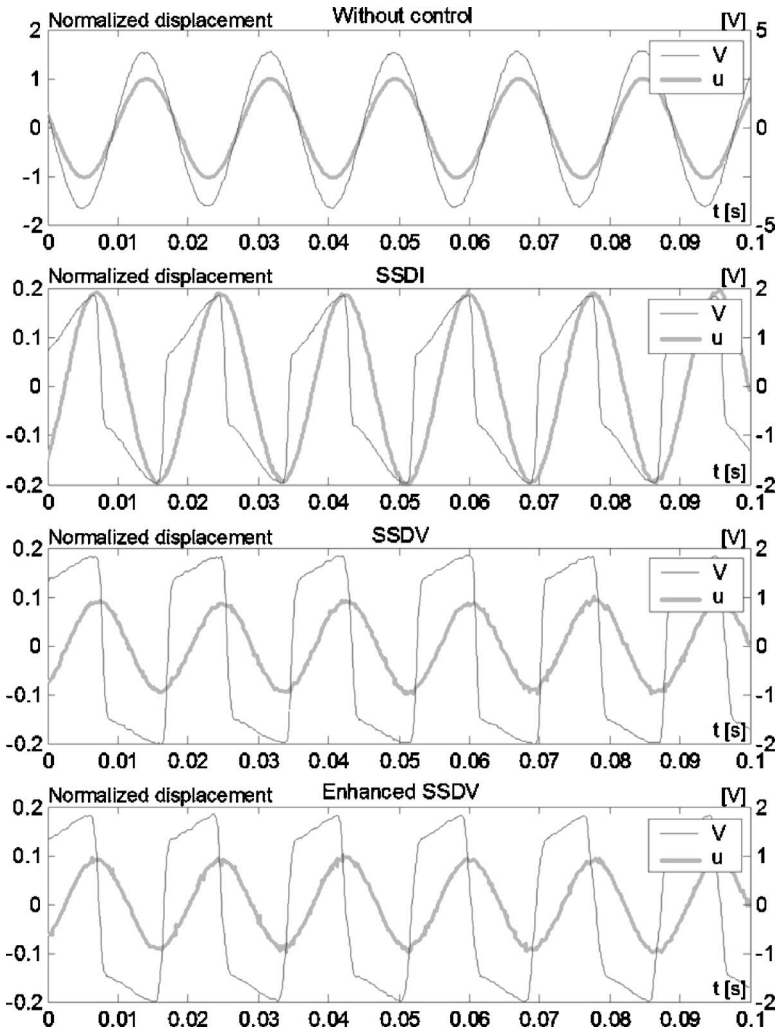


FIG. 10. Voltage and normalized displacement waveforms for the different technique.

plotted in Fig. 10. The voltage inversion induces a discontinuity in the force applied by the piezoelectric elements. This discontinuity leads to a small acceleration gap of the structure motion, but the displacement remains continuous, due to inertial effect. Moreover, the shape of the voltage induces excitation of the structure on odd multiples of the resonance frequency. In this particular case, the structure is highly resonant and no resonance frequency corresponds to odd multiples of the first bending mode. The displacement remains then almost sinusoidal (other lines of the displacement spectrum are at least −50 dB under the first resonance frequency line).

Figure 11 shows both theoretical and experimental damping for each technique as a function of the normalized driving force amplitude. The structure is driven at its resonance frequency; V_{cc} and β are unchanged. The driving force

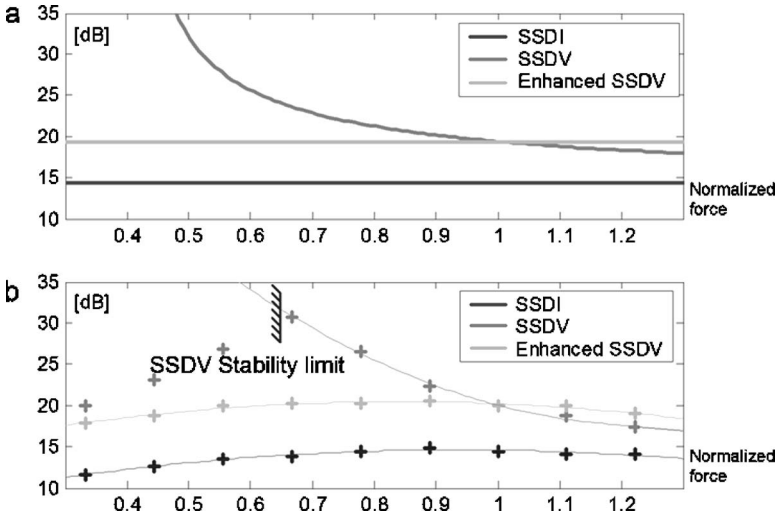


FIG. 11. Theoretical (a) and experimental (b) damping absolute value for the different techniques versus the normalized driving force $-V_{cc}=3$ V and $\beta=0.95$.

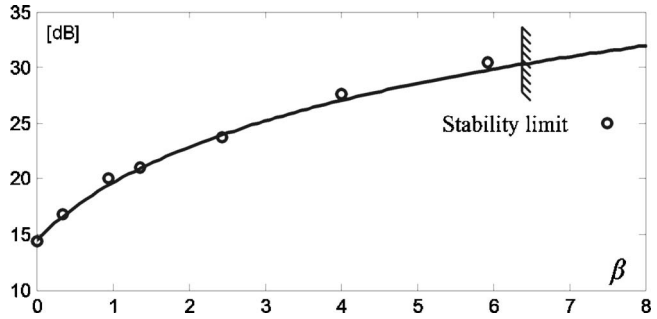


FIG. 12. Theoretical and experimental enhanced SSDV damping absolute value versus β.

is normalized so that 1 corresponds to the driving force of the previous harmonic analysis. SSDI and enhanced SSDV damping remain theoretically constant versus the driving force amplitude, which is confirmed experimentally. The original SSDV damping is logically a decreasing function of the driving force amplitude. The original SSDV damping is lower than the enhanced SSDV damping for normalized force amplitude greater than 1 and greater for normalized force amplitude lower than 1. The original SSDV technique leads to stability problems when the amplitude of the driving force is too low. In this case, the displacement higher harmonics cannot be neglected any more compared to the fundamental. New displacement extrema are induced which destabilize the switching control and lead to a lack of effectiveness compared to theoretical results. Actually, even if there were no higher harmonics excitations, stability problems would occur when the external driving force become lower than the force induced by the piezovoltage, i.e., when Eq. (32) is verified.

For $\beta=0.95$, the enhanced SSDV technique is stable whatever the amplitude of the driving force. Actually, β can be raised to increase the damping performance. Figure 12 shows the enhanced SSDV damping at the resonance frequency versus β. It is shown that the damping reaches -30 dB for $\beta=6$. It must also be highlighted that when $\beta=6$, the damping is almost -30 dB whatever the amplitude of the driving force.

The enhanced SSDV technique is inherently stable whatever the amplitude of the excitation because the voltage source level is continuously adapted in order to fit the piezoelectric generated control force to the mechanical excitation. Parameter β can be considered as a feedback gain allowing us to adjust the amplitude of the control force to the vibration amplitude. Theoretically, this closed loop process is stable for any value of β. However, as usual, a large value of the feedback gain limits the modulus margin and therefore instabilities may be observed when perturbations due to noise or harmonics are important compared to the excited mode. In the experimental case, noises and electronic implementation imperfections reach the margin modulus for β greater than 6. It is remarkable that under this practical limit, the system is stable and the damping is constant for any exciting force amplitude.

IV. EXPERIMENTS UNDER WIDEBAND EXCITATION

A. Experimental setup

The setup considered here is shown in Fig. 13. It is composed of a composite board with builtin piezoelectric elements (Qiu, 2003). This *smart board* is clamped on the extremity of a rectangular pipe. The piezoelectric elements are symmetrically located along the four clamped sides, where the strain is a maximum for almost all the modes. This geometry has been chosen because it exhibits a larger modal density than the previous cantilever beam. Actually eight lines almost regularly spaced are obtained on a 1-kHz frequency band. A speaker driven by a white noise electrical signal is used to excite simply and simultaneously the various modes. Practically the speaker is installed in a 60-cm pipe section, closed on one end by a rigid panel and by the equipped panel on the other end.

White noise is generated inside the pipe by the speaker and can only be transmitted to the outside via vibration of the composite board. This experimental setup is used to show that the SSDV technique is also effective for wideband excitation of a more complex and realistic structure than a cantilever beam. All piezoelectric patches are electrically connected together and then several vibration modes exhibit electromechanical couplings. The goal of this experiment is to show that, for those coupled mode, the SSD technique has multimodal vibration damping capabilities.

Most of the piezoelectric surface is connected to the nonlinear switching device, but a small part is used as a vibration sensor. Practically, the switch S_W is an electronic switch driven by a digital signal processor (DSP) which detects the main extrema of the sensing voltage V_S. The variable continuous voltage source V_{cc} is actually the output of a basic operational amplifier driven by the DSP. The amplitude

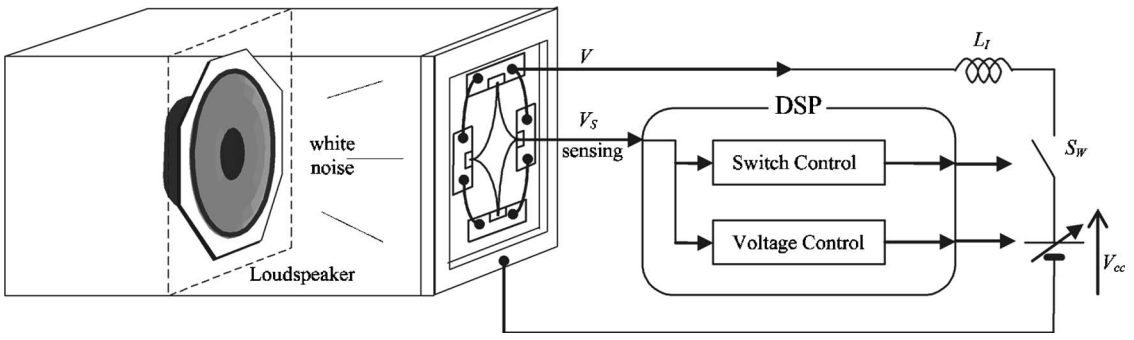


FIG. 13. Experimental setup.

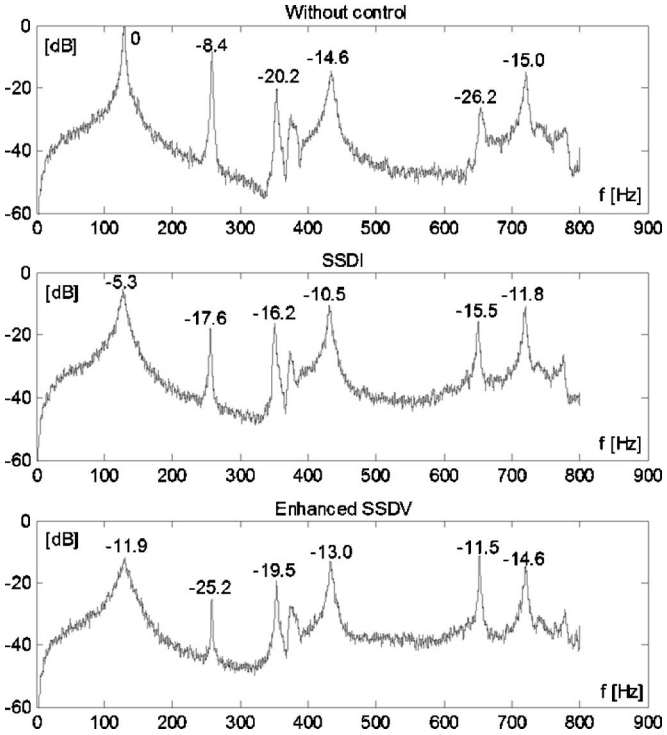


FIG. 14. Spectrum of sensing voltage temporal derivative without control, using the SSDI technique and using the enhanced SSDV technique.

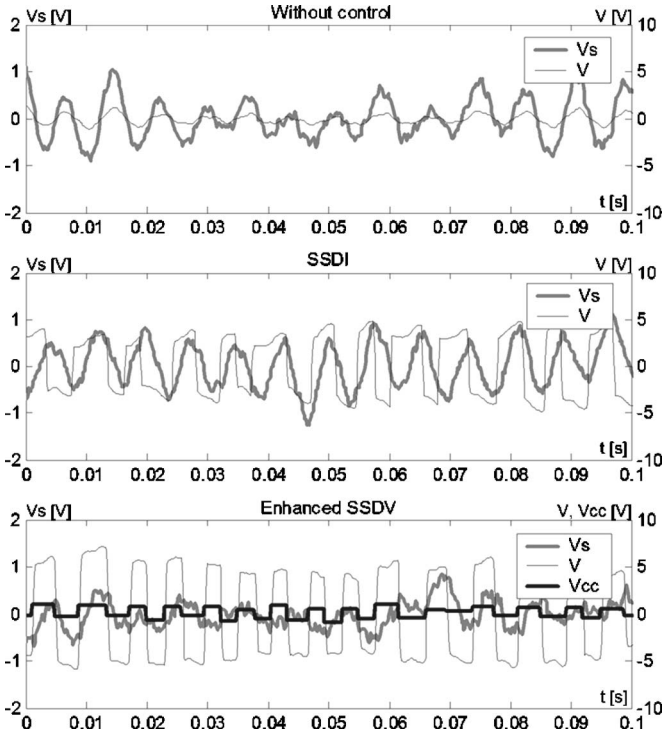


FIG. 15. Voltage and strain waveforms without control, using the SSDI technique and using the enhanced SSDV technique.

of V_{cc} is simply set to be proportional to the amplitude of V_S. The optimal gain coefficient β is fitted only one time. It is 2.2 in this experimental case.

In this first approach, only the coupled modes are considered and, thus, the sensing voltage V_S can be used to monitor the board vibrations on these particular modes.

B. Results

The vibration generated by the noise is related to the harmonic response of the board, and the voltage V_S, which is used for the monitoring of the board vibration for the coupled mode, is also related to the coupling coefficient corresponding to each mode.

Three cases are considered: vibration without control, SSDI control, and enhanced SSDV control. The spectrum of the temporal derivative of V_S is plotted in Fig. 14 in the three cases. The voltage temporal derivative is used instead of the voltage itself for clarity reasons, in order to emphasize the effect of the SSD techniques on higher frequencies. Without control, the first mode (130 Hz) exhibits the highest line, principally because it is the highest coupled mode. Characteristics of the first mode are given in Table V. The other modes are not as highly coupled and their squared coupling coefficients are between 0.1% and 0.5%. The uncontrolled voltage V_S is roughly harmonic, since the first mode is at least 8 dB higher than the other lines.

TABLE V. Characteristics of the board first mode (composite board).

Resonance frequency	130 Hz
Mechanical quality factor	30
Squared coupling coefficient k^2	1%

It is shown that using the SSDI technique leads to a −5.3 and −9.2-dB attenuation of the two main lines (the two first lines) of the spectrum without control. Other lines are increased, especially the 651-Hz one, but remain lower than the first line corresponding to the resonance frequency.

Using the enhanced SSDV technique leads to a −11.9 and −18.8-dB attenuation of the two main lines without control. It can be noticed that except for the 651-Hz line, which is greatly increased, the other lines remain almost constant. Actually, all lines are under −11.5 dB and the 651-Hz line is the most important line. It should also be noticed that the spectrum corresponds to the voltage temporal derivative and that the most important line would still be the first resonance frequency considering the voltage spectrum.

Actually the SSD techniques mainly control the first resonance frequency and its corresponding nonlinearity (the second line, which is twice the first resonance frequency). The SSD techniques lead then to a leveled down spectrum, since the first mode is mainly addressed. The switch frequency mainly corresponds to the first resonance frequency of the board. The sharp edges of the piezovoltage, resulting from the nonlinear processing, generate then excitations on odd multiples of the first resonance frequency. The 651-Hz line magnitude is increased because it corresponds to both a resonance frequency of the structure and an odd multiple of the first resonance frequency. The global damping effect of the SSD techniques can be evaluated by Eq. (44). The SSDI global damping effect is −2.3 dB whereas it is −6.1 dB using the enhanced SSDV approach:

$$A_{\mathrm{dB}} = 10 \log \left(\frac{(\int V_s^2 \, dt)_{\text{with control}}}{(\int V_s^2 \, dt)_{\text{without control}}} \right). \tag{44}$$

Figure 15 gives a representation of the sensing voltage V_S

and the processed voltage V for a small part of the whole recorded time. In the case of the enhanced SSDV technique the continuous voltage V_{cc} is also plotted. It clearly shows the voltage magnification due the SSDI technique and its enhancement using the voltage source.

The noise transmission control is not fully addressed in the first approach presented here. Further work will consist in using the SSD technique on several sets of piezoelectric patches, each coupled to a set of vibration modes. The control of the panel vibration will then reduce the noise transmitted through the board.

V. CONCLUSION

The enhanced SSDV technique allows us to tune the mechanical braking force resulting from the SSDV process to the vibration level. The continuous voltage source that amplifies the voltage inversion is proportional to the vibration amplitude. With the original SSDV technique, the absolute value of the continuous voltage remains constant which induces stability problems when the vibration amplitude tends to zero. Current work aims at clarifying the stability problems of the original SSDV technique. Several phenomena must be taken into account which reduce the system stability: higher harmonics excitation due to the sharp edges of the piezovoltage, technical difficulties to detect extrema for very low vibration level, overbalance of the external force when switching on too high voltage source, and practical delay between extrema occurrences and the moment they are detected.

These stability problems are bypassed using the enhanced SSDV. In the case of a single frequency excitation, it was shown that the enhanced SSDV technique gives a drastic constant and stable damping whatever the amplitude of the driving force (up to -30 dB when $\beta=6$). Experimental results also show that the enhanced SSDV technique is able to simultaneously control the different lower frequency modes of a composite plate. Ongoing work also aims at qualifying the panel sound transmission coefficient and its control using SSDI, SSDV, and enhanced SSDV techniques. Due to their simplicity and relative low cost, these SSD techniques could prove to be competitive for noise isolation.

ACKNOWLEDGMENTS

The authors would like to thank the International COE of Flow Dynamics, Tohoku University, for its support, in particular for the experimental part of this work.

Clark, W. W. (**2000**). "Vibration control with state-switching piezoelectric materials," J. Intell. Mater. Syst. Struct. **11**, 263–271.

Cunefare, K. A. (**2002**). "State-switched absorber for vibration control of point-excited beams," J. Intell. Mater. Syst. Struct. **13**, 97–105.

Larson, G. D., Rogers, P. H., and Munk, W. (**1998**). "State switched transducers: A new approach to high-power low frequency, underwater projectors," J. Acoust. Soc. Am. **103**(3), 1428–1441.

Lefeuvre, E., Guyomar, D., Petit, L., Richard, C., and Badel, A. "Semi-passive structural damping by synchronized switching on voltage sources," J. Intell. Mater. Syst. Struct. (in press).

Qiu, J., Tani, J., Haraguchi, M., and Yoshida, M. (**2003**). "Vibration Control of a Plate Using Self-Sensing Actuator," Proc. Mechanical Engineering Congress 2003 Japan, JSME **6**, 63–64.

Richard, C., Guyomar, D., Audigier, D., and Bassaler, H. (**2000**). "Enhanced semi passive damping using continuous switching of a piezoelectric device on an inductor," Proceedings. SPIE Smart Structures and Materials Conf.: Passive Damping and Isolation, Vol. **3989**, pp. 288–299.

Richard, C., Guyomar, D., Audigier, D., and Ching, G. (**1998**). "Semi-passive damping using continuous switching of a piezoelectric device," Proceedings of the SPIE Smart Struct. and Mat Conf., Passive Damping and Isolation, San Diego, Vol. **3672**, pp. 104–111.

Measuring the free field acoustic impedance and absorption coefficient of sound absorbing materials with a combined particle velocity-pressure sensor

R. Lanoye,[a] G. Vermeir, and W. Lauriks
Laboratorium voor Akoestiek en Thermische fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200D, BE-3001 Heverlee, Belgium

R. Kruse and V. Mellert
Institut für Physik, Carl von Ossietzky Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 20 October 2005; revised 17 February 2006; accepted 19 February 2006)

Acoustic surface impedance of sound absorbing materials can be measured by several techniques such as the impedance tube for normal impedance or the Tamura method for normal and oblique surface impedance. *In situ*, the acoustic impedance is mostly measured by use of impulse methods or by applying two-microphone techniques. All these techniques are based on the determination of the sound pressure at specific locations. In this paper, the authors use a method which is based on the combined measurement of the instantaneous sound pressure and sound particle velocity. A brief description of the measurement technique and a detailed analysis of the influence of the calibration, the source type, the source height, the sound incidence angle, and the sample size are included. [DOI: 10.1121/1.2188821]

PACS number(s): 43.58.Bh, 43.20.Ye, 43.20.El [LLT] Pages: 2826–2831

I. INTRODUCTION

To model the performance of acoustic absorbing materials in practice, researchers are interested in the reflection coefficient of those materials for different angles of sound incidence. The acoustic surface impedance and the absorption coefficient of absorbing materials can be measured with the impedance tube[1] for normal sound incidence or the Tamura method[2,3] for normal and oblique incidence of sound. For *in situ* measurements, common techniques are the impulse method[4] or the two-microphone technique.[5] Allard *et al.*[6] present a method for measurements around grazing incidence *in situ*. All these techniques make use of the measurement of local sound pressure.

As mentioned before, a proven technique to determine the angle dependency of the surface impedance and the absorption coefficient is described by Tamura. In this method, sound pressure is measured in two planes parallel to the surface of interest and a two-dimensional Fourier transformation is used to calculate the angle dependent surface impedance. Thus, no assumptions about the exact nature of the wave field are made. However, a drawback of the method is the rather complicated measurement setup together with the need for large test samples and the time-consuming measurement procedure.

In this paper, a new method for determining the surface impedance and the absorption coefficient of an absorbing material in the free field is investigated. A brief description of the measurement technique and a detailed analysis of the influence of the calibration, the source type, the source height, the sound incidence angle, and the sample size is included. All time-varying quantities should obey the time dependence $e^{-i\omega t}$ with $i=\sqrt{-1}$.

II. MEASUREMENT DEVICE

A. Working principle of the particle velocity transducer

The transducer consists of two very closely spaced heated wires. The measurement principle is based on the detection of the temperature difference between these two resistive sensors.[7] A traveling acoustic wave causes a time varying heat transfer from one sensor element to the other. The subsequent temperature difference results in a time-dependent difference between both electrical resistance values, which quantifies the particle velocity in a linear matter. However, due to the effects of diffusion and heat capacity, the sensitivity of the particle velocity sensor decreases with frequency.[8] In the *p-u* measurement probe, commercially known as "Microflown," a particle velocity sensor (u) and a miniature pressure microphone (p) are placed close together. In that way, local field impedance can directly be measured. This combination was shown to be able to measure the absorption coefficient in an impedance tube accurately.[9]

B. Calibration methods

Depending on frequency, a calibration can be performed in a standing wave tube or alternatively a free field calibration can be used. The standing wave tube calibration is used for lower frequencies (100 Hz–4 kHz). The free field calibration is applicable in the frequency range 200 Hz–20 kHz.

[a]Electronic mail: reinhilde.lanoye@bwk.kuleuven.be

 0001-4966/2006/119(5)/2826/6/$22.50

1. Standing wave tube calibration

In a standing wave tube, the sound pressure at a distance x from the rigid ending is given by

$$p(x,t,\omega) = A(e^{ikx} + e^{-ikx})e^{-i\omega t} \tag{1}$$

at frequencies below the cut-off frequency.

The particle velocity at the same position is given by

$$u(x,t,\omega) = \frac{1}{i\omega\rho_0}\frac{\partial p(x,t,\omega)}{\partial x} = \frac{A}{\rho_0 c}(e^{ikx} - e^{-ikx})e^{-i\omega t}. \tag{2}$$

The calibration of the pressure and particle velocity sensor as well as their phase difference is done separately to allow for the use of the combined probe (p-u probe) not only as an impedance sensor but also for intensity and particle velocity measurements.

A reference microphone is mounted at the rigid reflecting ending of the standing wave tube with a length of 76 cm and a diameter of 4.7 cm. At the other end, a loudspeaker is placed. If the p-u probe is set at a distance l from the reflecting surface (l lies typically in the range of 50 cm in these measurements), the ratio between the pressure measured by the p-u probe and the pressure at the reflecting ending is given by

$$\frac{p_{\text{probe}}(\omega)}{p_{\text{ref}}(\omega)} = \cos(kl). \tag{3}$$

The ratio between the particle velocity measured by the p-u probe and the pressure at the reflecting ending equals

$$\frac{u_{\text{probe}}(\omega)}{p_{\text{ref}}(\omega)} = \frac{i}{\rho_0 c}\sin(kl). \tag{4}$$

To calibrate the phase difference between the particle velocity sensor and the miniature pressure microphone in the combined probe, the ratio of particle velocity and pressure is measured. Theoretically, the ratio between both at the same point in the standing wave tube is

$$\frac{u_{\text{probe}}(\omega)}{p_{\text{probe}}(\omega)} = \frac{i}{\rho_0 c}\tan(kl) \tag{5}$$

and thus, the phase difference has to be equal to

$$\text{phase}\left\{\frac{u_{\text{probe}}(\omega)}{p_{\text{probe}}(\omega)}\right\} = \pm 90^\circ. \tag{6}$$

The difference between the measured phase and the theoretical value is used for calibration of the phase.

To calculate the frequency dependent sensitivity of the pressure sensor (sens_p) and particle velocity sensor (sens_u), the measured (voltage) transfer function is multiplied by the reference microphone sensitivity and, for the particle velocity sensor, additionally with $\rho \cdot c$. Following from Eqs. (3) and (4), the local maxima of these scaled transfer functions mark the sensitivity of the sensors at the respective frequency. A spline interpolation is done to obtain calibration values for the whole frequency range. In Fig. 1, the scaled transfer function and the resulting sensitivity is shown for the particle velocity sensor.

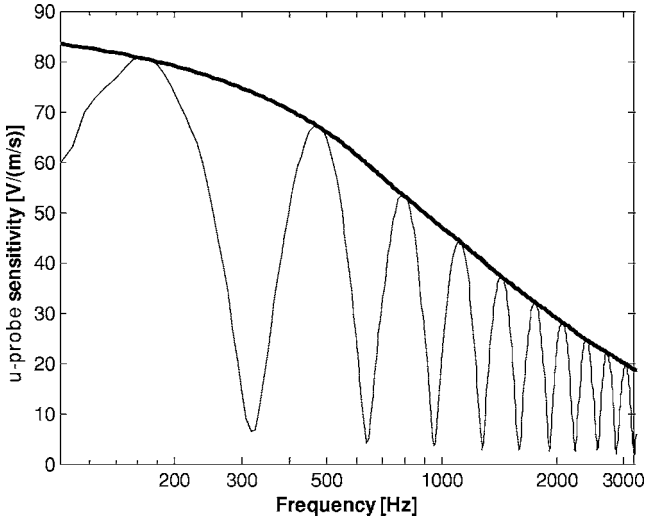


FIG. 1. Sensitivity of the Microflown element of the p-u probe. The thin line shows the measured ratio of the particle velocity at the p-u probe to the pressure at the rigid ending. The thick line marks the frequency dependent sensitivity in V/(m/s).

To obtain the phase correction function (phase_up), the local maxima and minima of the phase difference between the particle velocity sensor and the pressure microphone of the p-u probe are determined. Again, a spline interpolation is done to obtain calibration values for the whole frequency range. Following Eq. (6), the mean of the maximum and minimum values marks the deviation from ideal (zero phase) behavior and is used as correction function as shown in Fig. 2.

The preceding calibration method has the advantage of not requiring a precise value for l and c.

A correction function (CF) can be defined for the standing wave tube calibration. Later measurements have to be multiplied with this correction function to obtain calibrated measurement results. CF is defined in

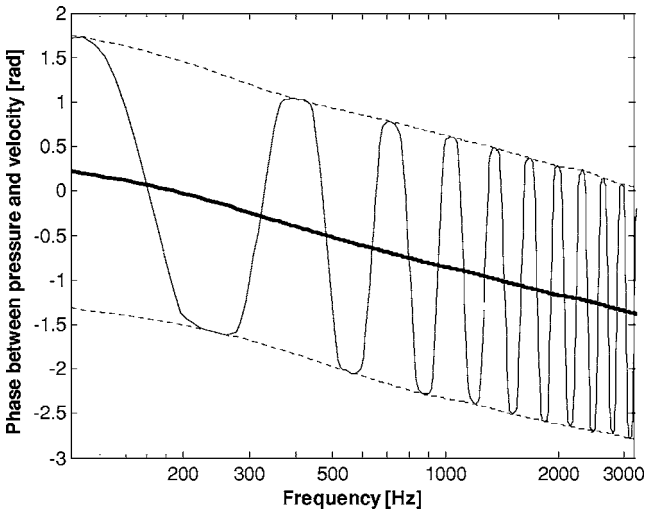


FIG. 2. Phase difference between the particle velocity sensor and the pressure element of the p-u probe. The thin line shows the measured phase between the particle velocity and the pressure at the p-u probe. The thick line marks the frequency dependent phase error.

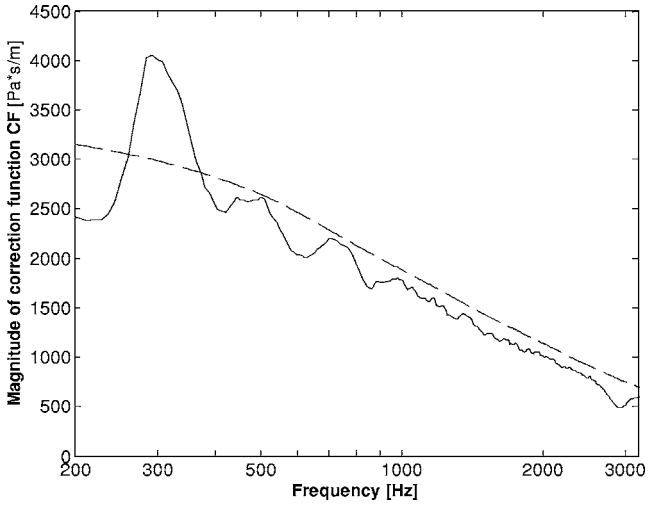


FIG. 3. Magnitude of the correction function (CF). Free field calibration (——) vs standing wave tube calibration (---).

$$\mathrm{CF} = \frac{\mathrm{sens\_}u}{\mathrm{sens\_}p} e^{i\cdot \mathrm{phase\_}up}. \tag{7}$$

Effects of damping along the tube could be shown to be negligible following an analysis based on Tijdeman's theory.[10]

2. Free field calibration

Free field calibration relies on the fact that the field impedance in a free field at a distance r is given by

$$Z_0(r,\omega) = \rho_0 c \frac{ikr}{ikr-1} \tag{8}$$

if sound is emitted by a point source.

The ratio between pressure and particle velocity is measured in free field conditions and equals Z_m. The ratio of the theoretical free field impedance to the measured impedance is the correction function,

$$\mathrm{CF} = \frac{Z_0}{Z_m}, \tag{9}$$

which is subsequently applied for calibration of all p-u measurements to obtain the calibrated field impedance. In contrast to the standing wave tube calibration, pressure and particle velocity probe cannot be calibrated separately.

The magnitude of the CF is represented in Fig. 3 for both the standing wave tube calibration [Eq. (7)] and the free field calibration [Eq. (9)]. The calibration function of the free field calibration is measured with use of a loudspeaker in a box and a probe-loudspeaker distance of 1.18 m. The probe is placed above absorbing wedges.

In the figure, a difference between both correction functions can be seen, especially at low frequencies. These deviations are due to unwanted reflections in the semi-anechoic room, e.g., coming from the floor. So, one can conclude that free field calibration is too sensitive to reflections for use at low frequencies.

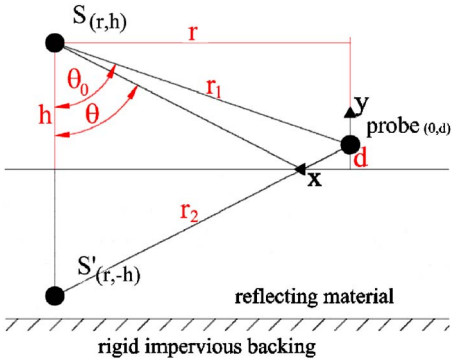


FIG. 4. (Color online) Mirror source model of the reflection of a spherical wave at an impedance plane.

III. CALCULATION TECHNIQUE

Two models are used to derive the surface impedance of a material from the measured field impedance at a certain single point above the surface. Both models assume a point source excitation of the air medium above the sample.

A. Mirror source model

This model assumes a plane-wave-like reflection. To include the reflection, a mirror source is placed behind the sample at the same distance as the physical source is placed from the sample, see Fig. 4.

This assumption is only valid if the source-sample distance is large compared to the wavelength.[11]

The angle dependent plane wave reflection coefficient $R_p(\omega,\theta)$, which can be calculated out of a measured impedance at a normal distance d above the surface, is given by

$$R_p(\omega,\theta) = e^{ik(r_1-r_2)} \frac{r_2}{r_1} \frac{Z(d,\omega)\dfrac{1-ikr_1}{-ikr_1}\cos\theta_0 - \rho_0 c}{Z(d,\omega)\dfrac{1-ikr_2}{-ikr_2}\cos\theta + \rho_0 c}. \tag{10}$$

r_1 equals the length of the direct path, r_2 is the length of the path mirror source-measurement point, and θ is the specular angle of incidence. θ_0 in this equation is the angle between the normal to the surface and the path which connects source and measurement point.

B. Asymptotic solution for locally reacting surfaces

The F-term solution of Nobile and Hayek[12] for the sound field above an impedance plane takes into account that a spherical wave is reflected differently from a plane wave. It assumes a locally reacting surface.

The velocity potential at a distance d above the impedance plane is given by

$$\phi(d) = \frac{e^{ikr_1}}{r_1} + Q\frac{e^{ikr_2}}{r_2} \tag{11}$$

in which Q denotes the spherical wave reflection coefficient

$$Q = R_p + (1-R_p)F. \tag{12}$$

R_p is the plane wave reflection coefficient, F the boundary loss function. This function F is determined by the geometry

TABLE I. Properties of the studied porous material.

Thickness (mm)	Density (kg/m^3)	Porosity (–)	Tortuosity (–)	Flow resistivity (Ns/m^4)	Viscous length (μm)	Thermal length (μm)
30	10	>0.95	1.09	10500	100	150

of the problem and the impedance of the absorbing surface. The pressure is given by

$$p(d) = \rho \frac{\partial \phi}{\partial t} = -i\rho\omega\phi, \tag{13}$$

the particle velocity by

$$u(d) = -\operatorname{grad} \phi(d). \tag{14}$$

This gradient is estimated by a finite difference approach,

$$u(d) = [\phi(d+\varepsilon) - \phi(d-\varepsilon)]/2\varepsilon \tag{15}$$

with a small ε.

With a minimum search, the surface impedance which minimizes the difference between the measured field impedance and the field impedance predicted by the model is found. The resulting surface impedance serves to calculate the plane wave absorption coefficient of the material.

If the source-sample distance and the measurement frequency are large enough, both calculation methods will be equivalent.

IV. MATERIAL DESCRIPTION

In this study, the measurement material is a 3 cm thick open-cell Melamine foam. The properties of this foam, given in Table I, are measured according to the appropriate methods. Descriptions of the methods to measure the characteristic lengths are given by Leclaire *et al.*[13] The techniques to measure porosity and tortuosity are described by Fellah *et al.*[14] The method to measure flow resistivity is given in ISO-9053.[15]

This material is assumed to behave as an equivalent fluid. To simulate the surface impedance and the absorption coefficient of this material, the theory of Johnson *et al.*[16] and Champoux and Allard[17] is used.

V. MEASUREMENT PROCEDURE

The measurements described in the following were performed in a semi-anechoic chamber with a lower limiting frequency of 150 Hz. The air in this semi-anechoic room was excited with two different sound sources, both driven with bandwith limited pseudorandom noise between 100 Hz and 6.4 kHz. The considered sound sources are a Monacor KU-156 compression driver with a 10 cm tube of $\frac{1}{2}$ in. diameter mounted on it ("point source") and a 20 cm woofer in a closed box. A Stanford Research System SR780 network analyzer was used for all measurements and postprocessing was done with MATLAB.

Samples of different sizes were placed on the rigid floor and the *p-u* probe was mounted above the surface at a height between 1.5 and 3 cm at a radial distance to the source equaling a sound incidence angle of up to 60°. The orientation of the *p-u* probe was carefully checked to agree with the normal of the sample surface.

VI. SENSITIVITY ANALYSIS

In the following, a detailed analysis of the influence of some measurement parameters is given. For the convenience of the reader and to provide a compact presentation of the results, the results are presented as absorption coefficients. Measurements are performed on a 1.8×1.2 m^2 sample, unless stated otherwise.

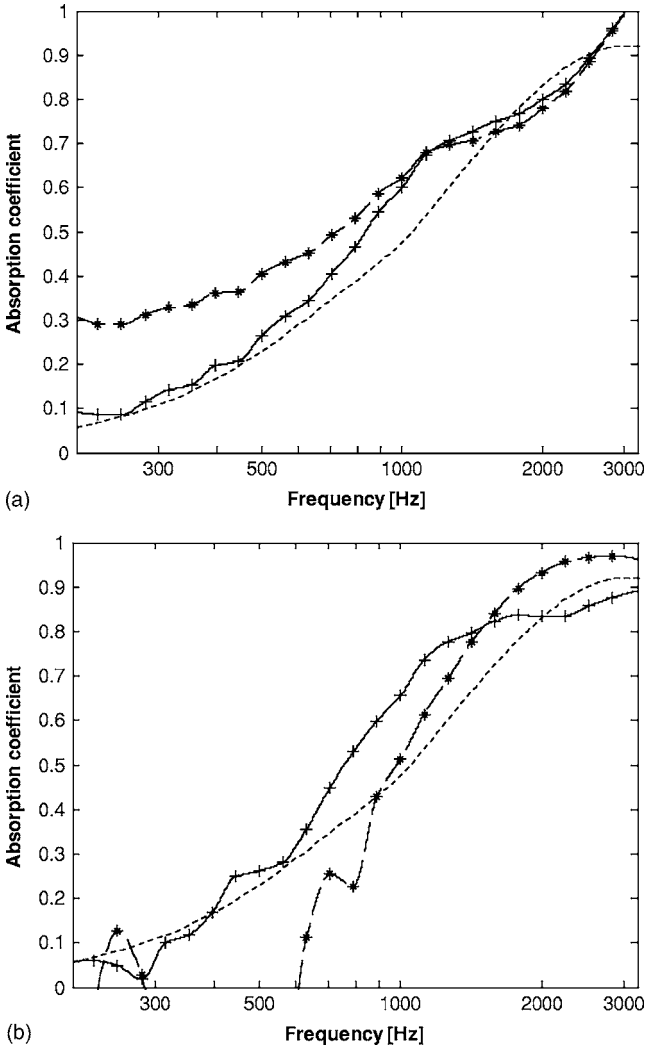
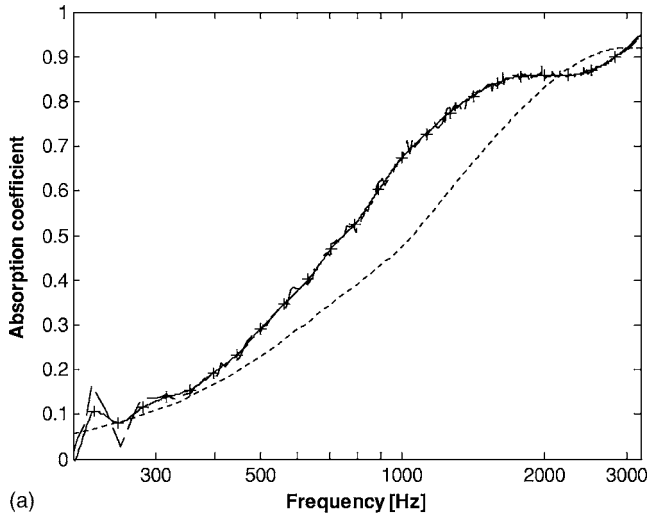


FIG. 5. Absorption coefficient of 3 cm Melamine foam for normal incidence. (a) Comparison of the mirror source model (*) and the asymptotic solution (+). Simulation results (···). (b) Comparison of the free field calibration (*) and the standing wave tube calibration (+). Simulation results (···).



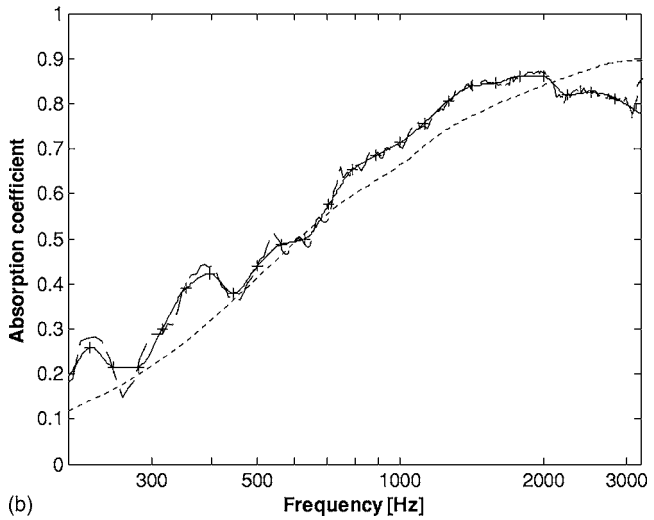


FIG. 6. Absorption coefficient of 3 cm Melamine foam. Measurement (---) with a woofer in a closed box at 1.21 m height. 1/6th octave average marked with (+). Simulation results (···). (a) Normal incidence, (b) 45° incidence.

A. Comparison of calculation models

In Fig. 5(a), a comparison is made between both above-described calculation methods. The same field impedance measurement, calibrated with a standing wave tube calibration, is used for this comparison. At low frequencies, the mirror source model differs significantly from the simulation. As shown in the experiments, the best agreement is found for the asymptotic model, which is further used as the reference.

B. Comparison of calibration methods

As can be seen in Fig. 5(b), a standing wave tube calibration allows more accurate measurements at low frequencies than a measurement with a free field calibration of the sensor probe. The surface impedance and absorption coefficient were calculated with the asymptotic solution.

C. Influence of source type

Comparison measurements were done with two different, above-described, sound sources. It could be concluded that the measurements with the loudspeaker in a box are the most accurate, especially at low frequencies, and this source will thus be used in further research. The good result for low frequencies is due to the high sound pressure that can be reached with a loudspeaker in a box. Therefore, the particle velocity at the surface of a material with high impedance is still measurable. This source has also the advantage of increasing directivity at higher frequencies and therefore avoids unwanted reflections.

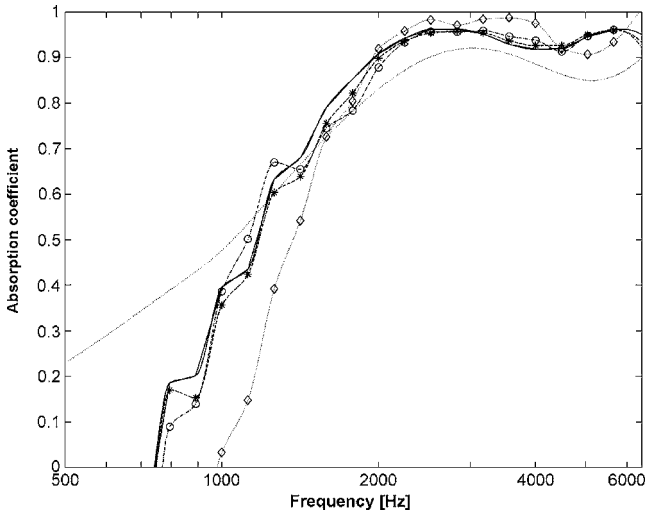


FIG. 7. Absorption coefficient of 3 cm Melamine foam for normal incidence: Influence of sample size. The size were 1.2×1.2 m^2 (---), 0.6×0.6 m^2 (*), 0.3×0.3 m^2 (°), and 0.15×0.15 m^2 ($\diamond$). Simulation results (···).

D. Influence of source height

The source was placed at different heights between 0.85 and 1.71 m above the absorbing material. If the solution of Nobile and Hayek is used to calculate the surface impedance, no significant differences can be found between the calculated absorption curves for different source heights. When the mirror source model is used, the solution below 1.2 kHz deviates more and more from the simulation when the source is placed closer to the surface.

E. Influence of incidence angle

Compared to the standing wave tube, one of the main advantages of this measurement technique is the possibility to measure the surface impedance for an arbitrary angle of sound incidence. The technique is also faster than the Tamura method. Measurements for a number of incidence angles were performed. The results for normal incidence and an angle of incidence of 45° are, respectively, shown in Figs. 6(a) and 6(b). The results for normal incidence are the best, since the phase between particle velocity and pressure is larger at normal incidence than at oblique incidence and hence the influence of the phase mismatch after calibration is smaller.

F. Influence of sample size

In this specific study, a point source was used. Knowing the required sample size is very important to know the *in situ*

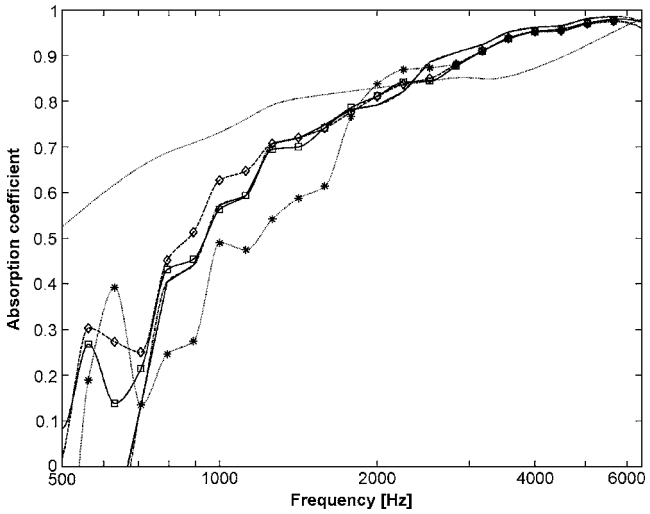


FIG. 8. Absorption coefficient of 3 cm Melamine foam for 60° incidence angle: Influence of sample size. The size were 2.4×1.2 m^2 (□), 1.8×1.2 m^2 (◇), 1.2×1.2 m^2 (---), and 0.6×0.6 m^2 (*). Simulation results (···).

applicability of the measurement device. The absorption coefficient from 200 Hz up to 6.4 kHz is measured. Results of these measurements are plotted in Figs. 7 and 8. One can clearly see that a larger sample size is necessary for more oblique angles of incidence. For 60°, there is even a sample size of 1.8×1.2 m^2 necessary to obtain good results. The poor quality of the measurements at low frequencies is due to the free field calibration, but thanks to this calibration method, one can plot higher frequency results.

VII. CONCLUSION

A new method for the measurement of the surface impedance in free field of a layer of absorbing material has been investigated in this paper. When using this method, a simultaneous measurement of the particle velocity and the pressure above the absorbing material is performed, so the field impedance close to the surface can be measured directly. The asymptotic solution for the calculation of the surface impedance has been shown to give the best results.

By means of the presented sensitivity analysis, it can be concluded that reasonably accurate and fast measurements can be performed with this new method, provided that the particle velocity-pressure transducer is adequately calibrated. Also oblique surface impedance measurements are possible with this technique. For normal to slightly oblique incidence, a rather small absorbing surface is sufficient and so the method could be applied for *in situ* analysis of local sound absorption.

ACKNOWLEDGMENT

R.L. is a Research Assistant of the Fund for Scientific Research—Flanders (Belgium).

[1]ISO 10534, "Acoustics—Determination of sound absorption coefficient and impedance in impedance tubes—, Part 1 and 2," 1998.
[2]M. Tamura, "Spatial Fourier-transform method for measuring reflection coefficients at oblique incidence. I. Theory and numerical examples," J. Acoust. Soc. Am. **88**, 2259–2264 (1990).
[3]M. Tamura, "Spatial Fourier-transform method for measuring reflection coefficients at oblique incidence. II. Experimental results," J. Acoust. Soc. Am. **97**, 2255–2262 (1995).
[4]E. Mommertz, "Angle-dependent in-situ measurement of reflection coefficients using a substraction technique," Appl. Acoust. **46**, 251–263 (1995).
[5]J. F. Allard and Y. Champoux, "In situ two-microphone technique for the measurement of the acoustic surface impedance of materials," Noise Control Eng. J. **32**, 15–23 (1989).
[6]J. F. Allard, M. Henry, V. Gareton, G. Jansens, and W. Lauriks, "Impedance measurements around grazing incidence for nonlocally reacting thin porous layers," J. Acoust. Soc. Am. **113**, 1210–1215 (2003).
[7]F. Jacobsen and H.-E. de Bree, "A comparison of two different sound intensity measurement principles," J. Acoust. Soc. Am. **118**, 1510–1517 (2005).
[8]J. W. van Honschoten, "Modelling and optimisation of the Microflown," dissertation, University of Twente, 2004.
[9]Y. Liu and F. Jacobsen, "Measurement of absorption with a *p-u* sound intensity probe in and impedance tube," J. Acoust. Soc. Am. **118**, 2117–2120 (2005).
[10]H. Tijdeman, "On the propagation of sound waves in cylindrical tubes," J. Sound Vib. **39**, 1–33 (1975).
[11]I. Rudnick, "The propagation of an acoustic wave along a boundary," J. Acoust. Soc. Am. **19**, 348–356 (1947).
[12]M. A. Nobile and S. I. Hayek, "Acoustic propagation over an impedance plane," J. Acoust. Soc. Am. **78**, 1325–1336 (1985).
[13]P. Leclaire, L. Kelders, W. Lauriks, C. Glorieux, and J. Thoen, "Determination of the viscous characteristic length in air-filled porous materials by ultrasonic attenuation measurements," J. Acoust. Soc. Am. **99**, 1944–1948 (1996).
[14]Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristégui, and J.-Y. Chapelon, "Measuring the porosity and the tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **113**, 2424–2433 (2003).
[15]ISO-9053:1991, "Acoustics-materials for acoustical applications-determination of air-flow resistance."
[16]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).
[17]Y. Champoux and J. F. Allard, "Dynamic tortuosity and bulk modulus in air saturated porous media," J. Appl. Phys. **70**, 1975–1979 (1991).

Accuracy of an acoustic location system for monitoring the position of duetting songbirds in tropical forest

Daniel J. Mennill[a)]
Department of Biological Sciences, University of Windsor, 401 Sunset Avenue, Windsor, Ontario, Canada N9B3P4

John M. Burt[b)]
Psychology Department, University of Washington, Seattle, Washington 98195

Kurt M. Fristrup[c)]
Natural Sounds Program Center, National Park Service, 1201 Oakridge Drive, Suite 100, Fort Collins, Colorado 80525

Sandra L. Vehrencamp[d)]
Lab of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, New York 14850

(Received 1 September 2005; revised 14 February 2006; accepted 14 February 2006)

A field test was conducted on the accuracy of an eight-microphone acoustic location system designed to triangulate the position of duetting rufous-and-white wrens (*Thryothorus rufalbus*) in Costa Rica's humid evergreen forest. Eight microphones were set up in the breeding territories of 20 pairs of wrens, with an average intermicrophone distance of 75.2±2.6 m. The array of microphones was used to record antiphonal duets broadcast through stereo loudspeakers. The positions of the loudspeakers were then estimated by evaluating the delay with which the eight microphones recorded the broadcast sounds. Position estimates were compared to coordinates surveyed with a global-positioning system (GPS). The acoustic location system estimated the position of loudspeakers with an error of 2.82±0.26 m and calculated the distance between the "male" and "female" loudspeakers with an error of 2.12±0.42 m. Given the large range of distances between duetting birds, this relatively low level of error demonstrates that the acoustic location system is a useful tool for studying avian duets. Location error was influenced partly by the difficulties inherent in collecting high accuracy GPS coordinates of microphone positions underneath a lush tropical canopy and partly by the complicating influence of irregular topography and thick vegetation on sound transmission. [DOI: 10.1121/1.2184988]

PACS number(s): 43.60.Fg, 43.60.Rw, 43.60.Bf [JAS] Pages: 2832–2839

I. INTRODUCTION

An acoustic location system consists of an array of microphones that facilitates spatial monitoring of animals based on the sounds they produce. Acoustic location systems offer special research advantages because animals' natural movement patterns are not inhibited by the passive monitoring process and animals can be located even in environments characterized by visual obstructions (McGregor *et al.*, 1997). Furthermore, the signals exchanged between animals are recorded as part of the acoustic location process, facilitating detailed studies of communication processes between multiple individuals simultaneously (McGregor *et al.*, 1997; Burt and Vehrencamp, 2005). Acoustic location systems have been used extensively to monitor the behavior of many marine organisms, with particular focus on large mammals (e.g., Stafford *et al.*, 1998; Mellinger and Clark, 2003; Clark and Clapham, 2004). In contrast, research employing terrestrial acoustic location systems is uncommon (but see Spiesberger, 1999; Hayes *et al.*, 2000; Mohl *et al.*, 2001; Wahlberg *et al.*, 2001, 2003; Bower and Clark, 2005; Burt and Vehrencamp, 2005). Here we test the accuracy of a passive acoustic location system for monitoring the position of duetting songbirds in neotropical forest.

Duets occur when two animals, usually breeding partners, coordinate their songs in overlapping or exactly alternating bouts of vocalizations. Duets are produced in many texa, including insects (Bailey, 2003), frogs (Emerson and Boyd, 1999), mammals (e.g., Geissmann, 2002), and birds (Hall, 2004). Duetting occurs in more than 220 species of birds and ranks among the most highly coordinated behaviors known in the animal kingdom (Thorpe, 1963; Hall, 2004). However, much controversy and little consensus surrounds the ecological and evolutionary significance of duetting, as well as the mechanics of vocal coordination by duetting partners. This is, in part, because many duetting birds live in densely vegetated tropical forest where traditional tools for following and observing animals, such as visual tracking or radiotelemetry, are difficult. Nevertheless, spatial analyses of duetting may allow for an advanced understand-

[a)]Electronic mail: dmennill@uwindsor.ca
[b)]Electronic mail: quill@u.washington.edu
[c)]Electronic mail: kurt_fristrup@nps.gov
[d)]Electronic mail: slv8@cornell.edu

ing of the significance of this intriguing behavior. We were interested in studying duetting behavior in neotropical rufous-and-white wrens (*Thryothorus rufalbus*) from a spatial perspective. To that end, we developed an eight-microphone acoustic location system for spatial monitoring of the duetting behavior of birds in the humid forests of northwest Costa Rica. In this study we present a test of this system by quantifying the accuracy with which microphone array recordings allowed us to triangulate the position of loudspeakers broadcasting simulated rufous-and-white wren duets.

II. METHODS

A. Study species and study site

Rufous-and-white wrens live in forested and semi-forested habitat throughout much of Central America and northwestern South America. Males and females coordinate their songs to produce both antiphonal and polyphonal duets. The voices of the two sexes are similar, although male songs have significantly lower frequency characteristics and are significantly longer than female songs (Mennill and Vehrencamp, 2005). Rufous-and-white wrens are largely ground-dwelling songbirds (Ahumada, 2001) and typically remain in very low vegetation, occasionally ascending to higher perches to sing.

At Santa Rosa National Park in northwest Costa Rica (10°40′N, 85°30′W), our study population of rufous-and-white wrens inhabits the fragments of evergreen moist forest that remain within the mosaic of regenerating forests that constitute the Guanacaste Conservation Area. The study site is characterized by thick vegetation, including a rich vertical structure and a lush canopy dominated by guapinol (*Hymenea courbaril*), milk (*Brosiumu utile*), and chicle (*Manilkara chicle*) trees. The study site also has diverse topography—the wrens' territories follow the sloped stream-sides and small escarpments that are associated with the park's evergreen moist forest. These biotic and abiotic features of rufous-and-white wren territories provide a challenging model system in which to test an acoustic location system: rich understory vegetation complicates sound transmission, thick forest canopy makes high-accuracy survey coordinates difficult to obtain, and topographic diversity complicates both sound transmission and satellite reception. Therefore, if an acoustic location system can be used to accurately monitor the position of duetting birds in this habitat, it will be useful for monitoring animals in many other habitats as well.

B. Field methods

Our acoustic location system consisted of an array of eight omni-directional microphones which were connected by 1200 m of microphone cable to a centrally located computer. Microphone input was digitized using a multi-channel data acquisition card (model: National Instruments DAQ-6260). The microphones were housed in rain guards made of PVC tubing and speaker cloth, mounted on top of 3-m wooden poles via 30-cm shelf brackets. Stakes were elevated off the ground and attached to trees with bungee cords (Fig. 1). We attempted to space microphones evenly throughout each rufous-and-white wren territory, mounting microphones opportunistically where vegetation allowed. Within each array, the average between-microphone distance was 75.2±2.6 m (calculated pairwise for all microphones in each array), the maximum distance was 137.6±5.9 m, and the minimum distance was 31.5±1.4 m (mean±SE; $n=20$ array configurations).

FIG. 1. Photograph of one of the eight microphones used to record playback of rufous-and-white wren duets. Microphones were placed on 3.0-m wooden poles and attached to trees within wren territories. The average between-microphone distance within each of the 20 array configurations was 75.2±2.6 m.

Between 18 April and 8 July 2003 and 2004 we used the eight-channel microphone array to record duets played through stereo loudspeakers within the breeding territories of 20 different pairs of wrens. Recording sessions were conducted between 0630 and 0845 h. The left and right loudspeakers (model: Sony SRS-A37) were separated by a distance of 10 to 20 m (average distance between loudspeakers: 16.3±0.5 m), a distance within the observed range of distances between duetting rufous-and-white wrens. The left loudspeaker broadcast one bird's duet contribution and the right loudspeaker broadcast the partner's duet contribution. The loudspeakers were oriented upwards and mounted on poles at a height of 1.0 m. Stimuli were broadcast from a portable CD player (model: Sony D-SJ301). The volume of broadcast songs was held constant across all trials, at a natural sound pressure level of 80 dB at 1 m (measured at a horizontal distance from the upwards-oriented loudspeaker using a Realistic 33-2050 sound level meter on slow setting).

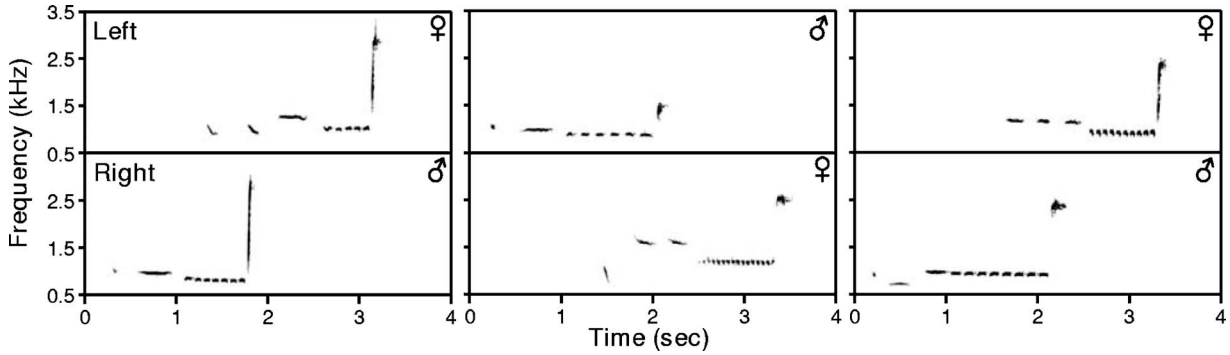


FIG. 2. Sound spectrograms of rufous-and-white wren duet stimuli broadcast within the acoustic location system. Spectrograms depict stereo sound files with the left channel on top and the right channel below. Examples of three different pairs' duets are shown.

The playback stimuli were duets recorded from the 20 resident pairs. We recorded stimuli as monaural WAV files (16 bit sampling, 22 050 Hz sample rate) with a directional microphone (model: Sennheiser MKH-70) and a solid state digital recorder (model: Marantz PMD-670) during spontaneous bouts of early-morning singing. The male and female duet components were isolated from each another using the "frequency cursor filter" function of Syrinx-PC (J. Burt, Seattle, WA). The amplitudes of both the male and female songs were then subtly adjusted to the same maximum amplitude using the "amplitude normalize" feature of CoolEdit 2000 (Syntrillium, Phoenix, AZ). Male and female songs were then pasted into the left and right channels of 16-bit stereo WAV files (Fig. 2). We alternated whether the male or female song was broadcast through the left or right channel from one playback session to the next. Rufous-and-white wren duets occur in several different varieties: the male may begin singing and the female may join in part way through the male song or immediately after it, the sex roles may be reversed, or both birds may begin singing at the same time. Male-initiated duets are the most common variety (Mennill and Vehrencamp, 2005) and for this reason we used male-initiated duets in this test of location accuracy. Within each array, we recorded the broadcast of six duets, as well as six male solos and six female solos for comparison. The solo stimuli were identical to the male and female contributions to the duet.

Traditional survey methods are difficult given the thick vegetation at our study site. We surveyed microphone and loudspeaker positions using an Ashtech ProMark II Global Positioning System (GPS; Thales Navigation, Santa Clara, CA). This system consists of twin receiver units with external antennae which collect data in tandem at two positions, allowing for differential correction based on the data collected at the two sites simultaneously. We mounted the GPS antennae on tripods at a height of 1.70 m. In an attempt to overcome the error introduced by GPS sampling underneath a thick forest canopy, we sampled points extensively and repeatedly, sampling each site two to six times between 2003 and 2005. Combined over three years of GPS sampling, we sampled the position of each microphone for 85.9 ± 1.3 min ($n=160$ microphone positions) and the position of each loudspeaker for 74.7 ± 1.6 min ($n=40$ loudspeaker positions). We used Ashtech Solutions Software (Thales Navigation, Santa Clara, CA) to calculate an average Universal Transverse Mercator (UTM) coordinate from each loudspeaker and microphone position. We report error around these coordinates as the standard error given as output by the Ashtech Solutions Software following postprocessing adjustment.

C. Array analysis

In the laboratory, we used Syrinx-PC to browse eight-channel spectrograms of field recordings. We used the time and frequency cursors of Syrinx-PC to draw a selection box around non-overlapping sections of male and female duet contributions. As is typical of rufous-and-white wren duets (Mennill and Vehrencamp, 2005), our broadcast stimuli involved duets where the terminal section of the male song was overlapped by the beginning section of the female song. Therefore, we selected only the introductory portion of male duet contributions and the terminal portion of the female duet contributions. For our analyses of the solo songs, in contrast, we were able to select the entire male or female song.

We used software written in MatLab (Mathworks Inc., Natick, MA) to estimate the location of the loudspeakers that broadcast male and female songs. The localization process consists of two components: computing cross-correlation functions and searching for the best location estimate. The correlation component begins by copying the sound data from the selected channel and applying a FIR filter to the data to restrict the signal energy to the frequency limits of the selection box. The digital samples that fall outside of the temporal bounds of the selected region are replaced with zeroes. This edited signal is cross correlated against the remaining seven channels. These correlations are resistant to the presence of other sounds within the data segment because the extent of the selection is used to mask out all the other signals in the selected channel. Twenty-one additional correlation functions are computed between all pairs of microphones that do not include the selected channel. For these correlations, it is advantageous to utilize the channels with the strongest signals to provide the masked template that is correlated against the remaining channels. The amplitudes of peaks in the initial seven correlations are used to order the seven channels in terms of estimated signal strength. The remaining correlations are computed as follows: the channel

with the highest correlation peak is correlated with the remaining six channels, the channel with the second highest correlation peak is correlated with the remaining five channels, etc. The locations of the correlation peaks are used to apply masking, to help reduce the impact of other signals that occur in the data segment. The software can deal with multiple correlation peaks that complicate the masking process, but this complication was not an issue in this study.

Once the correlation functions are computed, the estimated location of the selected sound is obtained by a stochastic search which involves the GPS coordinates of the eight microphones. For each candidate source location, the predicted time delays are computed using the Euclidean distances between the source location and the microphone locations, combined with the speed of sound. The values of the correlation functions associated with the predicted delays are extracted and these 28 values are summed. The location with the maximal sum of these correlation values provides the estimated location. This optimization criterion is algebraically identical to finding the point in space that produces the highest energy output from a delay-and-sum beamforming system (Johnson and Dudgeon, 1993). Our array analysis process assumes that the speakers lie in a plane (two-dimensional localization) and deviations from this assumption contribute to location errors. For all analyses we fixed the speed of sound at 348 m/s (corresponding to air temperature of 25 °C and 90% relative humidity, as is typical at our study site at the time of recording; see below). We evaluate how variation in air temperature and humidity—and therefore speed of sound—influenced position estimates by plotting temperature and humidity against location accuracy.

D. Location accuracy

To test the accuracy of the acoustic location system, we compared the loudspeakers' GPS coordinates to the positions estimated by the array analysis software. We report the difference as "location accuracy;" the Euclidean distance between the GPS coordinates of the loudspeaker and the array-estimated coordinates of the sound source (in meters).

To evaluate environmental variables that may have influenced the accuracy of our array-based location estimates, we compared location accuracy measurements across the 20 arrays to four features of the acoustic environment: (1) air temperature at the time of the recording, (2) relative humidity of the air at the time of the recording, (3) recording date (a rough measure of leaf phenology), and (4) variance in the height of the eight microphones within each array (a rough measure of topographic variation within each recording area). We measured temperature and humidity with a data logger (model: Hobo Pro, Onset Technology, Pocasset, MA), which was hung at a height of 1.0 m near the center of each array, and sampled the air condition every 5 min. We also compare our error measurements to two features of the microphone arrays themselves: (5) the density of microphones (measured as the average distance between all pairs of microphones in each array) and (6) the distance between the loudspeaker and the nearest microphone [a feature which has been shown to influence recording accuracy in a study of temperate microphone arrays (McGregor *et al.*, 1997)].

E. Statistical methods

All analyses were conducted in JMP 5.0 (SAS Institute, Cary, NC). We compare the location error of female versus male songs broadcast as solos versus duets with two-factor ANOVA. We test for associations between accuracy and environmental variables with Pearson correlations. All tests are two-tailed. All values are reported as mean±SE.

III. RESULTS

A. GPS accuracy

Using a global position system (GPS) to survey the positions of eight microphones within each of the 20 microphone arrays, we achieved a latitudinal accuracy of 1.26±0.05 m and a longitudinal accuracy of 1.36±0.05 m, for an average accuracy of 1.31±0.05 m (n=160 microphone positions). For loudspeaker positions, we achieved a latitudinal accuracy of 2.49±0.09 m and a longitudinal accuracy of 2.68±0.09 m, for an average accuracy of 2.59±0.09 (n=40 microphone positions; lower accuracy in loudspeaker coordinates arose from shorter sampling periods). The accuracy of all GPS coordinates was constrained by the thick canopy, as has been found in other studies involving GPS-based research in tropical forest environments (e.g., Phillips *et al.*, 1998; Dominy and Duncan, 2001).

B. Location accuracy

Using the eight-channel acoustic location system, we triangulated the position of loudspeakers broadcasting rufous-and-white wren song with an average location error of 2.82±0.26 m (n=40 loudspeakers, average location error calculated across all stimuli broadcast from each loudspeaker; Fig. 3). It is noteworthy that this value is fractionally larger than the estimated GPS location error for the speaker locations, suggesting that the combined effects of all errors in the acoustic location system are smaller than the GPS survey error for the speakers. The location error was similar for locations based on the broadcast of male solos, male duet contributions, female solos, and female duet contributions (Fig. 4; two-factor ANOVA: whole model $F_{3,76}$ =0.4, p=0.73) despite the significantly lower frequency characteristics of male songs (see Mennill and Vehrencamp, 2005). The six songs broadcast from each loudspeaker generated tightly clustered position estimates, demonstrating that our position estimates were repeatable.

Array recordings accurately estimated the distance between a pair of loudspeakers simulating the male and female duet contributions of a pair of rufous-and-white wrens. The discrepancy in the estimated distance between the male and female loudspeakers, as measured by GPS versus the acoustic location system, was 2.12±0.42 m (n=20).

C. Factors influencing location accuracy

None of the variables we measured showed a relationship with location accuracy (Fig. 5). Location accuracy did

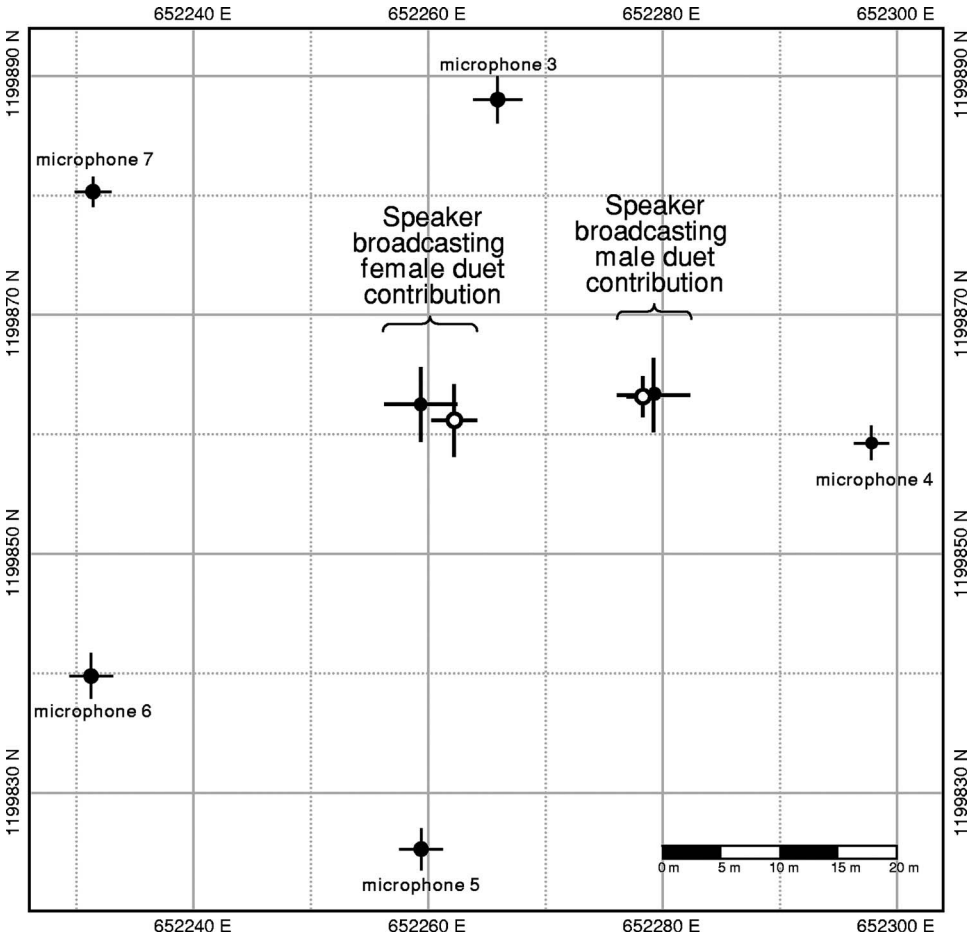


FIG. 3. Map of a portion of a microphone array, showing five microphones and two loudspeaker positions. Stereo loudspeakers were set up near the center of the area enclosed by the microphones and stereo duet stimuli were broadcast through the left and right channels. The solid circles show positions of loudspeakers and microphones measured by GPS. The open circles show the position of the loudspeakers estimated using the acoustic location system. Horizontal and vertical lines show standard error.

not vary significantly across a temperature range of 22.1 to 25.2 °C (r=0.07, p=0.67, n=40), a range typical of the early to mid morning at our study site. Location accuracy did not vary significantly across a humidity range of 58.3% to 99.2% (r=0.12, p=0.45, n=40). The acoustic location system was remarkably robust and functioned effectively even at the high humidity levels within this tropical forest.

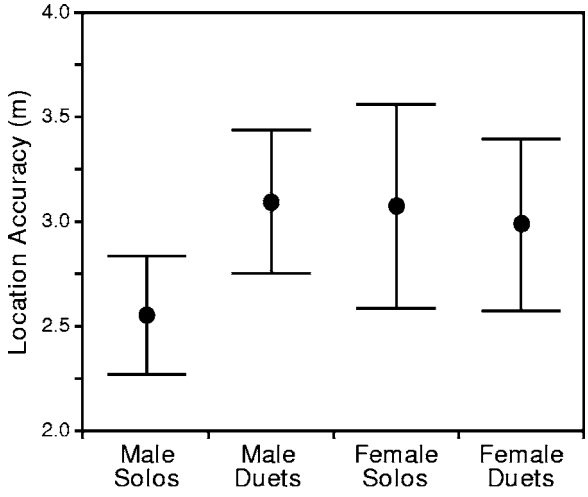


FIG. 4. Location accuracy for n=40 loudspeakers which broadcast rufous-and-white wren vocalizations and were recorded within 20 microphone arrays. Location accuracy was similar when loudspeakers broadcast male and female vocalizations, whether as solos or duets. Points show means and vertical bars show standard error.

Location accuracy did not change with the time of year (r =0.02, p=0.92, n=40), indicating that leaf phenology had little influence on the accuracy of the acoustic location system. This is not surprising, as the leaf structure of the study site remains similar throughout most of the year, varying primarily with bursts of leaf production in the canopy, as is the case for many tropical forests (van Schaik *et al.*, 1993). Despite the variation in topography at the study site, location accuracy did not show a significant relationship with topographic variation, measured as the variance in the heights of the microphones (r=0.18, p=0.26, n=40). Similarly, location accuracy did not vary significantly with microphone density (r=0.04, p=0.81, n=40) or the distance between the loudspeaker and the center of the array (r=0.09, p=0.59, n =40).

IV. DISCUSSION

The acoustic location system provided accurate estimates of the position of loudspeakers broadcasting rufous-and-white wren duets in neotropical humid forest. Triangulating the position of loudspeakers based on the delay in time with which eight microphones recorded playback stimuli, we achieved an average location error of 2.82 m. We estimated the distance between the loudspeakers broadcasting male and female duet components to within 2.12 m. Therefore, our test of acoustic location system accuracy establishes microphone array technology as a useful tool for spatial analysis of

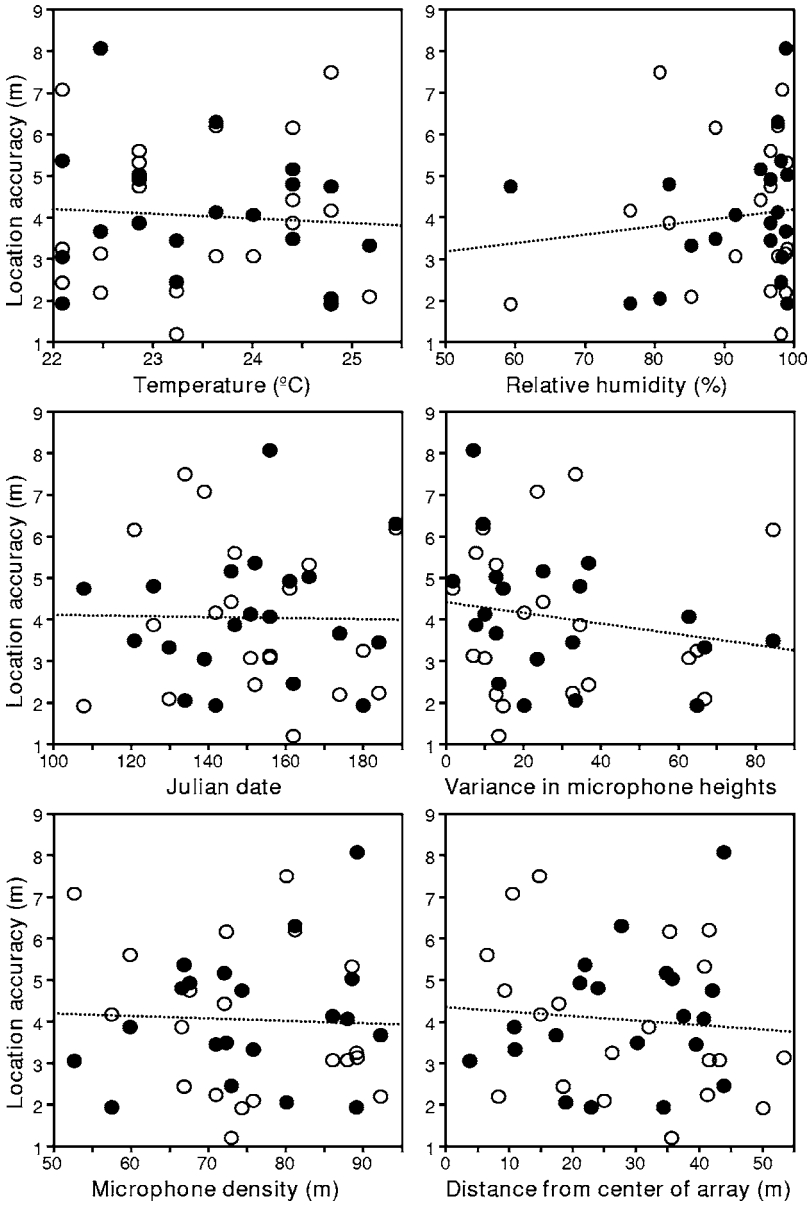


FIG. 5. Location accuracy did not show a strong relationship with four features of the recording environment or two features of the array configuration: air temperature, the relative humidity of the air, Julian date (a rough measure of leaf phenology), variance in microphone height (a rough measure of topographic variation), the density of microphones within the array (measured as the average distance between microphones), or the distance between the loudspeaker and the center of the array. Open circles show $n=20$ loudspeakers which broadcast female duet contributions and closed circles show $n=20$ loudspeakers which broadcast male duet contributions.

the vocal behavior of animals in tropical forests, even within areas characterized by dense vegetation, thick canopy cover, and irregular topography.

When participating in vocal duets, the distance between male and female rufous-and-white wrens varies tremendously (D. J. Mennill, unpublished data). Because the error of our acoustic location system is small, the accuracy of this acoustic location system will be revealing for monitoring duetting behavior in these neotropical wrens. Recording natural duets with the acoustic location system will allow us to test hypotheses for the function of avian duets by evaluating how close together birds are when they duet, whether birds approach one another during a bout of duetting, and where birds sing duets relative to territory boundaries and other habitat features. Whereas single-channel recordings are predicated on the idea of monitoring one individual at a time, a multi-microphone system may be used to monitor multiple individuals simultaneously, and therefore to understand behaviors such as duetting. Moreover, because multi-microphone recordings permit spatial analysis of animal vocal behavior, acoustic location systems can be used to monitor the movement and spacing of animals in habitats where traditional techniques are impractical.

Error in our ability to triangulate exactly the position of the loudspeakers arose from at least two sources. First, the accuracy of our acoustic location system is dependent on the accuracy with which we surveyed microphone positions. Our GPS coordinates of microphone positions were accurate to only 1.31 m despite extensive surveying across multiple years. Tropical forest ranks among the most challenging environments for collecting high-accuracy GPS coordinates because of the presence of dense canopy vegetation (Blake *et al.*, 2001; Phillips *et al.*, 1998; Dominy and Duncan, 2001). When surveying areas near our study site that lacked a dense canopy cover, we found that our GPS was capable of achieving submeter accuracy, supporting the idea that it is the canopy layer within the study site that made high-accuracy GPS coordinates difficult to obtain, not satellite geometry or any other feature of the tropics *per se*. When future research questions require more precise GPS coordinates for points

within canopied tropical forest, raising GPS antennae high into the canopy on telescoping poles may facilitate submeter level accuracy.

The second source of error in our location accuracy probably stemmed from the influence of dense vegetation and diverse topography on sound transmission within the study site. Our acoustic location system triangulates the position of animals in two dimensions, effectively treating the recorded area as a flat plane. In reality, the study site has many hills and valleys which are likely to influence patterns of sound reverberation for songs transmitted in this environment. We found no relationship between the variance in the height of the eight microphones in each array (a rough measure of the topographic variation within each array) and location accuracy. However, we do not interpret this as evidence that topography had no influence on the acoustic location system, but rather that topography had equivalent effects across all 20 array configurations. There are no analytical software tools yet developed to account for both three-dimensional transmission of sound together with a variable habitat surface. This is an important area for future enhancement in acoustic location system accuracy.

Both McGregor *et al.* (1997) and Bower and Clark (2005) evaluated the accuracy of a four-microphone array for triangulating the position of songbirds. By recording a variety of sounds in an open meadow and an open woodland in Europe, McGregor *et al.* (1997) evaluated the accuracy of an acoustic location system in two temperate environments. They report an average location error of 3.68 ± 0.29 m in open meadow and 6.76 ± 0.56 in open woodland, although accuracy improved substantially when the broadcast sound was closer to the center of their four-element arrays (McGregor *et al.* 1997). Bower and Clark (2005) recorded songbirds in a regrown farm field in North America with a similar array of four microphones. They report an average location error of 0.82 ± 0.29 m for birds recorded near the center of their four-element array and 2.13 ± 1.30 m for birds recorded just outside of their array. Both McGregor *et al.* (1997) and Bower and Clark (2005) report a relationship between location error and the distance between the loudspeaker and the center of the array. We did not find this effect in our analyses, however all of our stimuli were recorded from positions near the center of the 20 arrays, whereas these other analyses included sounds recorded outside the area bounded by their four microphones. We expect that location accuracy would decline if duets were recorded from outside of the area encompassed by the array of microphones. McGregor *et al.* (1997) and Bower and Clark (2005) recorded only on windless days in environments which induced little reverberation, whereas we conducted recordings through a variety of inclement weather conditions, including substantial winds and persistent dripping water, and in an environment which was likely to induce much reverberation. Our analyses provide a realistic test of the abilities of an acoustic location system and demonstrate that an acoustic location system can be used in suboptimal recording conditions.

Our understanding of the behavior and ecology of tropical animals remains poorly developed, despite the tremendous diversity of animal life in the tropics and the urgency for tropical research in light of the conservation threats to tropical forests (Kroodsma *et al.*, 1996). Acoustic location technology is a highly valuable tool for enhancing our understanding of the ecology and evolution of tropical animals. Microphone array recordings may be useful for studying extremely secretive animals or animals that are difficult to observe because of dense vegetation, large home ranges, or nocturnal activity. Microphone arrays also offer a valuable tool for studying network-level communication and have been used to study neighborhoods of song sparrows (Bower, 2000) and banded wrens (Burt and Vehrencamp, 2005). We believe that acoustic location technology may be used to study many other systems and provide ecologists and ethologists with a valuable and versatile tool for tropical research.

ACKNOWLEDGMENTS

We thank a dedicated field team who assisted with the difficult task of operating microphone arrays and collecting GPS data: C. Calaustro, V. Connolly, S. Doucet, A. Lindo, D. Moseley, R. Richter, and S. Topp. We thank R. Blanco and the staff of the Guanacaste Conservation Area for logistical support in the field. We thank everyone at Cornell University's Bioacoustics Research Program for logistical support during the design and analysis stages, including K. Dunsmore, A. Gururajan, and K. Cortopassi for implementing and testing the localization algorithm. We thank J. Cuthbert, S. Doucet, L. Reed, K. Swiston, S. Topp, and an anonymous reviewer for comments on the manuscript. Financial support for this project was provided by grants from the National Geographic Society, the Natural Science and Engineering Research Council (NSERC) of Canada, the American Museum of Natural History's Frank M. Chapman Memorial fund, the Birmingham Audubon Society's Walter F. Coxe Research Fund, and the Explorers Club's Exploration Fund to D.J.M., and by a grant from the National Institutes of Health (R01-MH60461) to S.L.V.

Ahumada, J. A. (**2001**). "Comparison of the reproductive biology of two neotropical wrens in an unpredictable environment in northeastern Colombia," Auk **118**, 191–210.

Bailey, W. J. (**2003**). "Insect duets: underlying mechanisms and their evolution," Physiol. Entomol. **28**, 157–174.

Blake, S., Douglas-Hamilton, I., and Karesh, W. B. (**2001**). "GPS telemetry for forest elephants in Central Africa: results of a preliminary study," Afr. J. Ecol. **39**, 178–186.

Bower, J. L. (**2000**). "Acoustic interaction sduring naturally occurring territorial conflict in a song sparrow neighbourhood," Ph.D. thesis, Cornell University, Ithaca, NY.

Bower, J. L., and Clark, C. W. (**2005**). "A field test of the accuracy of a passive acoustic location system," Bioacoustics 15, 1–14.

Burt, J. M., and Vehrencamp, S. L. (**2005**). "Dawn chorus as an interactive communication network," in *Animal Communication Networks*, edited by P. K. McGregor (Cambridge U. P., New York), pp. 321–343.

Clark, C. W., and Clapham, P. J. (**2004**). "Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late Spring," Proc. R. Soc. London, Ser. B **271**, 1051–1057.

Dominy, N. J., and Duncan, B. (**2001**). "GPS and GIS methods in an African rain forest: Applications to tropical ecology and conservation," Conserv. Ecol. **5**, 537–549.

Emerson, S. B., and Boyd, S. K. (**1999**). "Mating vocalizations of female frogs: Control and evolutionary mechanisms," Brain Behav. Evol. **53**, 187–197.

Geissmann, T. (**2002**). "Duet-splitting and the evolution of gibbon songs," Biol. Rev. Cambridge Philos. Soc. **77**, 57–76.

Hall, M. L. (**2004**). "A review of hypotheses for the functions of avian duetting," Behav. Ecol. Sociobiol. **55**, 415–430.
Hayes, S. A., Mellinger, D. K., Croll, D. A., Costa, D. P., and Borsani, J. F. (**2000**). "An inexpensive passive acoustic system for recording and localizing wild animal sounds," J. Acoust. Soc. Am. **107**, 3552–3555.
Johnson, D. J., and Dudgeon, D. E. (**1993**). *Array Signal Processing: Concepts and Techniques* (Prentice-Hall, Englewood Cliffs, NJ).
Kroodsma, D. E., Vielliard, J. M., and Stiles, F. G. (**1996**). "Study of bird sounds in the neotropics: Urgency and opportunity," in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell U. P., Ithaca, NY), pp. 269–281.
McGregor, P. K., Dabelsteen, T., Clark, C. W., Bower, J. L., Tavares, J. P., and Holland, J. (**1997**). "Accuracy of a passive acoustic location system: Empirical studies in terrestrial habitats," Ethol. Ecol. Evol. **9**, 269–286.
Mellinger, D. K., and Clark, C. W. (**2003**). "Blue whale (*Balaenoptera musculus*) sounds from the North Atlantic," J. Acoust. Soc. Am. **114**, 1108–1119.
Mennill, D. J., and Vehrencamp, S. L. (**2005**). "Sex differences in the singing and duetting behavior of neotropical Rufous-and-white Wrens (*Thryothorus rufalbus*)," Auk **122**, 175–186.
Mohl, B., Wahlberg, M., and Heerfordt, A. (**2001**). "A large-aperture array of non-linked receivers for acoustic positioning of biological sound sources," J. Acoust. Soc. Am. **109**, 434–437.
Phillips, K. A., Elvey, C. R., and Abercrombie, C. L. (**1998**). "Applying GPS to the study of primate ecology: a useful tool?" Am. J. Primatol **46**, 167–172.
Spiesberger, J. L. (**1999**). "Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration," J. Acoust. Soc. Am. **106**, 837–846.
Stafford, K. M., Fox, C. G., and Clark, D. S. (**1998**). "Long-range acoustic detection and localization of blue whale calls in the northeastern Pacific Ocean," J. Acoust. Soc. Am. **104**, 3616–3625.
Thorpe, W. H. (**1963**). "Antiphonal singing in birds as evidence for avian auditory reaction time," Nature (London) **197**, 774–776.
van Schaik, C. P., Terborgh, J. W., and Wright, J. S. (**1993**). "The Phenology of tropical forests: Adaptive significance and consequences for primary consumers," Annu. Rev. Ecol. Syst. **24**, 353–377.
Wahlberg, M., Mohl, B., and Madsen, P. T. (**2001**). "Estimating source position accuracy of a large aperture hydrophone array for bioacoustics," J. Acoust. Soc. Am. **109**, 397–406.
Wahlberg, M., Tougaard, J., and Mohl, B. (**2003**). "Localising Bitterns Botaurus stellaris with an array of non-linked microphones," Bioacoustics 13, 233–246.

Passive imaging of underground acoustic sources

Stephen J. Norton[a] and Bradley J. Carr
Geophex, Ltd., 605 Mercury Street, Raleigh, North Carolina 27603

Alan J. Witten[b]
School or Geology and Geophysics, University of Oklahoma, Norman, Oklahoma 73019

(Received 3 November 2005; accepted 28 February 2006)

The problem of imaging the intensity distribution of a subsurface acoustic source from passive measurements is considered. The source may be spatially extended, and is assumed to be random and spatially incoherent. Two passive imaging algorithms are examined and shown to be closely related. The first algorithm is based on time-domain migration in which signals recorded by an array of receivers are backpropagated to multiple source locations. The second algorithm is based on backprojecting cross-correlating signals computed between all pairs of receiving elements. When the latter algorithm is applied in two dimensions, the correlation peak can be shown to be the line-integral of the source intensity along a hyperbolic path whose foci are the receiving points. This suggests a filtered-backprojection procedure in which the backprojection is performed along multiple hyperbolic paths. The algorithm is easily generalized to three dimensions, in which the cross-correlated signals are backprojected over hyperboloids. Images using simulated and field data are presented as an illustration. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188667]

PACS number(s): 43.60.Pt, 43.60.Rw, 43.60.Lq [EJS] Pages: 2840–2847

I. INTRODUCTION

We consider the problem of reconstructing images of subsurface acoustic sources generated either artificially within an underground facility or by the scattering of ambient seismic noise impinging upon a buried structure. One application identified as important by the military is source location within a buried facility based on the passive monitoring of acoustic emissions from both stationary and moving equipment within such a facility. Another potential objective is to identify the type of source on the basis of its noise spectrum, although source characterization is beyond the scope of this paper; rather, our goal is to examine source imaging algorithms, that is, to show how an intensity map, or image, of an extended source distribution can be reconstructed using signals passively recorded by a geophone array deployed above ground.

Several approaches to passive seismic imaging have been studied in recent years. A method investigated by Shuster *et al.*[1] and Wapenaar *et al.*,[2–5] termed "seismic interferometry," exploits the interference of a direct arrival with a scattered or secondary arrival, which may, for example, originate from backscattering from the surface. This approach is an elaboration of an idea by Claerbout[6] relating transmission and reflection seismograms using autocorrelated data in which one component originates from a free surface reflection. If the subsurface source is impulsive or very broadband, then the cross-correlation between the direct and secondary arrivals can bear a strong resemblance to a reflection seismogram.

An essential ingredient in the above methods is the interference between a direct and one or more secondary arrivals. The passive imaging methods investigated in this paper differ in that they employ only the direct arrivals; no interference with a "reference wave" is needed. In particular, two imaging algorithms of this kind have been published,[7,8] and one aim of this paper is to elucidate the relationship between these two techniques. The earliest of these algorithms was derived for imaging acoustic emission sources in materials,[8] while the second algorithm was intended for seismic imaging.[7] It is important to note that these algorithms do not use time delays across an array to "triangulate" on the source, which is inherently a nonlinear procedure. Rather, the algorithms are truly imaging methods, capable of mapping a spatially extended source. The imaging algorithms are linear in the intensity, and thus the source need not be spatially localized. In order for the algorithms to be linear, the assumption of a spatially incoherent source is required, meaning that the signals emitted from spatially distinct points within an extended source are assumed to be statistically uncorrelated. It should be noted that passive imaging, as defined here, differs from the seismic source characterization problem, in which the parameters of an isolated source are to be estimated, such as the source location, the seismic moment, the event time, and so on. Our objective is to map a single parameter, the source intensity.

In one imaging algorithm,[7] termed "time-exposure acoustics," passive data acquired by an array of geophones are used to reconstruct a source image by backpropagating the signals into image space. This procedure bears some resemblance to time-domain migration. But, in contrast to the active source problem in which signals are transmitted into the ground, in the passive problem no time origin exists with respect to which the propagation delays can be referenced. That is, the signals are assumed emitted from the source in a continuous and random fashion. The essential idea described in Ref. 7 is to average with respect to this time origin. In

[a]Electronic mail: snorton@geophex.com
[b]Deceased.

 0001-4966/2006/119(5)/2840/8/$22.50

particular, multiple images are reconstructed, each with a different assumed time origin (the time at which the signal is emitted) and then these images are averaged on an intensity basis. This approach exploits the fact that the signal, either emitted from or scattered by a point in the subsurface, will create a wavefront that will manifest itself as a correlated signal across a receiving aperture on the surface. By coherently summing this signal and backpropagating to multiple source locations, one can develop an image of source intensity. As discussed in Ref. 7, image resolution is determined essentially by two factors: the source bandwidth and the array size.

The above method has been referred to as "time-exposure acoustics" (TEA) since the source is continuous and ever present, and, as a result, an image can be built up over time as data are continually accumulated. The underlying idea is to exploit the spatial coherence of signals that are correlated across an array to construct an image. When properly averaged, random fluctuations in an ensemble of images will diminish over time as new data are incorporated. Simulations in Ref. 7 illustrate this process.

In a second source reconstruction algorithm,[8] the source imaging problem is formulated as a tomographic reconstruction problem. Here, signals recorded by different pairs of geophones are cross correlated. The resultant correlation peak consists of an integration of the source intensity lying along a hyperbolic path [in a two-dimensional (2-D) problem] or on the surface of a hyperboloid (in a 3-D problem) in which the coordinates of the two geophones define the foci of the hyperbola or hyperboloid. In this formulation, the correlation peaks may be regarded as the line integral of source intensity over this hyperbolic path (in two dimensions) or as the surface integral of source intensity over a hyperboloid (in three dimensions). We shall see that this method is closely related to the TEA algorithm, although this relationship may not be immediately obvious.

In the analysis that follows, we shall assume a scalar wave (acoustic) problem. We shall also assume a constant velocity and neglect mode conversion effects at the Earth's surface. These assumptions are made for simplicity, but the essential ideas presented here can be generalized in principle to more complex (and realistic) earth models. In Secs. II and III, we analyze, respectively, the TEA and cross-correlation source imaging algorithms and examine their relationship. In Sec. IV we present images using simulated data and in Sec. V we show images derived from field data.

II. TIME-EXPOSURE ACOUSTICS (TIME-DOMAIN MIGRATION)

We first review the source imaging method known as time-exposure acoustics[7] and cast it in a form to facilitate later comparison to the cross-correlation algorithm.[8] Denote by $S(\mathbf{r},t)$ a random and spatially-varying subsurface source distribution. We assume that the source distribution is spatially incoherent, meaning that separate source points are statistically uncorrelated,[9] which we express by $\langle S(\mathbf{r},t)\, S(\mathbf{r}',t')^{*}\rangle = 0$ for $\mathbf{r} \neq \mathbf{r}'$, where the angle brackets denote an average with respect to time and * means complex conjugate.[10] Define the source intensity by $I(\mathbf{r}) = \langle |S(\mathbf{r},t)|^2 \rangle$. This is the quantity that we wish to image.

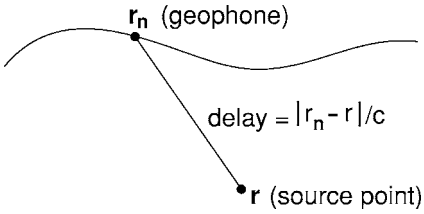


FIG. 1. The source point is located at $\mathbf{r}$, the geophone at $\mathbf{r}_n$, and the received signal denoted by $u(\mathbf{r}_n,t)$.

Suppose we have a two-dimensional array of N geophones above ground and let $u(\mathbf{r}_n,t)$ denote the signal recorded at the geophone location $\mathbf{r}_n$ (Fig. 1).

One could attempt to reconstruct an image, denoted $\hat{\mathcal{S}}_0(\mathbf{r})$, by backpropagating the received signals, $u(\mathbf{r}_n,t)$, to multiple source points, $\mathbf{r}$, and summing over the geophones. This is described by the formula

$$\hat{\mathcal{S}}_0(\mathbf{r}) = \sum_{n=1}^{N} |\mathbf{r} - \mathbf{r}_n| \bar{u}(\mathbf{r}_n, |\mathbf{r} - \mathbf{r}_n|/c), \tag{1}$$

where the overbar on $\bar{u}(\mathbf{r}_n,t)$ signifies a (possibly) filtered version of $u(\mathbf{r}_n,t)$, that is,

$$\bar{u}(\mathbf{r}_n,t) = u(\mathbf{r}_n,t) * b(t), \tag{2}$$

with $*$ denoting convolution and $b(t)$ a filter function. For example, $b(t)$ might be a whitening filter. In (1), the factor $|\mathbf{r}-\mathbf{r}_n|$ is included to compensate for signal loss due to geometric spreading of the wave originating from the point $\mathbf{r}$. If desired, we could replace this with a more general "gain" function, $a(|\mathbf{r}-\mathbf{r}_n|)$, to take into account a more complex model of signal attenuation (e.g., absorption in addition to geometric spreading). Note that in (1), the signal $\bar{u}(\mathbf{r}_n,t)$ is evaluated at the time $t=|\mathbf{r}-\mathbf{r}_n|/c$; this correctly describes backpropagation to the point $\mathbf{r}$ for a signal emitted from $\mathbf{r}$ at $t=0$. For a signal emitted at another time, say, t_0, then (1) should be replaced by

$$\hat{\mathcal{S}}(\mathbf{r}) = \sum_{n=1}^{N} |\mathbf{r} - \mathbf{r}_n| \bar{u}(\mathbf{r}_n, |\mathbf{r} - \mathbf{r}_n|/c + t_0). \tag{3}$$

The backprojection process is illustrated in Fig. 2 for a 2-D imaging problem. When imaging the point (x,z), backprojection is equivalent to coherently summing over a hyperbolic path in data space, (x_n,t), defined by $t=\sqrt{(x-x_n)^2+z^2}/c+t_0$ (the dashed line), where x_n is the coordinate of a geophone in a linear array. For a 3-D problem, when imaging the point $\mathbf{r}$, the summation is performed over a hyberboloid in data space, (r_n,t), defined by $t=|\mathbf{r}-\mathbf{r}_n|/c+t_0$.

Our approach is to average these images with respect to the "time origin," t_0, in a manner to be prescribed. For brevity, define

$$\bar{u}_n(\mathbf{r}) \equiv |\mathbf{r} - \mathbf{r}_n| \bar{u}(\mathbf{r}_n, |\mathbf{r} - \mathbf{r}_n|/c + t_0), \tag{4}$$

so that (3) is

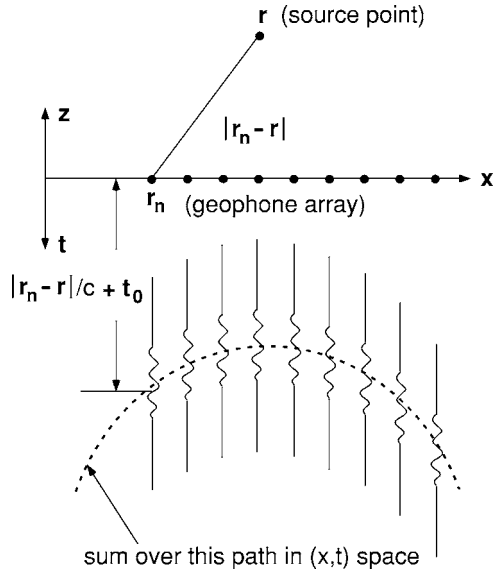


FIG. 2. When imaging the point (x,z) in two dimensions, backprojection is equivalent to coherent summation over the path $t=\sqrt{(x-x_n)^2+z^2}/c+t_0$ in (x_n,t) space, where x_n is the nth geophone coordinate in a linear array. When imaging the point (x,y,z) in three dimensions, backprojection is equivalent to coherent summation over the surface $t=\sqrt{(x-x_n)^2+(y-y_n)^2+z^2}/c+t_0$ in (x_n,y_n,t) space, where (x_n,y_n) is the nth geophone coordinate in a 2-D array.

$$\hat{\mathcal{S}}(\mathbf{r}) = \sum_{n=1}^{N} \bar{u}_n(\mathbf{r}). \tag{5}$$

Letting $\langle\cdot\rangle_{t_0}$ denote an ensemble average with respect to t_0, we see that $\langle\hat{S}(\mathbf{r})\rangle_{t_0}=0$ since the image $\hat{S}(\mathbf{r})$ is linear in the field $\bar{u}_n(\mathbf{r})$, which is a random variable with zero mean. This precludes averaging the images on an "amplitude" basis. We therefore consider averaging successive images on an intensity basis, i.e., computing as our image $\hat{\mathcal{I}}(\mathbf{r})=\langle|\hat{S}(\mathbf{r})|^2\rangle_{t_0}$. This, however, leads to a large "DC bias" in the image that builds up over time. An unbiased image estimator is obtained by subtracting the "DC component," as follows:

$$\hat{\mathcal{I}}(\mathbf{r}) = \left\langle \left| \sum_{n=1}^{N} \bar{u}_n(\mathbf{r}) \right|^2 \right\rangle_{t_0} - \left\langle \sum_{n=1}^{N} |\bar{u}_n(\mathbf{r})|^2 \right\rangle_{t_0}, \tag{6}$$

which can also be written as

$$\hat{\mathcal{I}}(\mathbf{r}) = \sum_{\substack{n=1 \\ n\neq m}}^{N} \sum_{m=1}^{N} \langle \bar{u}_n(\mathbf{r}) \bar{u}_m(\mathbf{r})^* \rangle_{t_0}. \tag{7}$$

Equation (7) shows that, if there is no correlation between distinct traces, $\bar{u}_n(\mathbf{r})$, then the image $\hat{I}(\mathbf{r})$ vanishes as desired.

In (6) and (7), the ensemble average is performed with respect to the time origin t_0. Let us rename this time origin t_k and define

$$\bar{u}_n^k(\mathbf{r}) \equiv |\mathbf{r}-\mathbf{r}_n| \bar{u}(\mathbf{r}_n, |\mathbf{r}-\mathbf{r}_n|/c + t_k). \tag{8}$$

We now average with respect to t_k, and (6) and (7) may be approximated by

$$\hat{\mathcal{I}}(\mathbf{r}) = \frac{1}{K}\sum_{k=1}^{K} \left\{ \left| \sum_{n=1}^{N} \bar{u}_n^k(\mathbf{r}) \right|^2 - \sum_{n=1}^{N} |\bar{u}_n^k(\mathbf{r})|^2 \right\} = \sum_{\substack{n=1 \\ n\neq m}}^{N} \sum_{m=1}^{N} \left(\frac{1}{K} \sum_{k=1}^{K} \bar{u}_n^k(\mathbf{r}) \bar{u}_m^k(\mathbf{r})^* \right), \tag{9}$$

where K is the number of averages.

Equation (9) represents our basic imaging algorithm. In the next section, we describe another algorithm based on a process of cross-correlation, filtering, and backprojection. To make the comparison between (9) and the cross-correlation algorithm easy, it will be convenient to rewrite (9) explicitly in terms of the correlation between the signals $\bar{u}(\mathbf{r}_n,t)$ and $\bar{u}(\mathbf{r}_m,t')$. Define the temporal correlation function

$$\Gamma_{\bar{u}\bar{u}}(\mathbf{r}_n,\mathbf{r}_m;t,t') = \langle \bar{u}(\mathbf{r}_n,t) \bar{u}(\mathbf{r}_m,t')^* \rangle_{t_0} \equiv \lim_{T\to\infty} \frac{1}{T} \int_0^T \bar{u}(\mathbf{r}_n,t+t_0) \bar{u}(\mathbf{r}_m,t'+t_0)^* \, dt_0. \tag{10}$$

The discrete-time version of this is

$$\Gamma_{\bar{u}\bar{u}}(\mathbf{r}_n,\mathbf{r}_m;t,t') = \frac{1}{K}\sum_{k=1}^{K} \bar{u}(\mathbf{r}_n,t+t_k) \bar{u}(\mathbf{r}_m,t'+t_k)^*, \tag{11}$$

where t_k is an indexed time sample. Substituting (8) into (9) and making use of (11), we have

$$\hat{\mathcal{I}}(\mathbf{r}) = \sum_{\substack{n=1 \\ n\neq m}}^{N} \sum_{m=1}^{N} |\mathbf{r}-\mathbf{r}_n||\mathbf{r}-\mathbf{r}_m| \times \Gamma_{\bar{u}\bar{u}}(\mathbf{r}_n,\mathbf{r}_m;|\mathbf{r}-\mathbf{r}_n|/c,|\mathbf{r}-\mathbf{r}_m|/c). \tag{12}$$

For a stationary process, $\Gamma_{\bar{u}\bar{u}}(\mathbf{r},\mathbf{r}';t,t')=\Gamma_{\bar{u}\bar{u}}(\mathbf{r},\mathbf{r}';t-t')$ and, in this case, (12) may be written as

$$\hat{\mathcal{I}}(\mathbf{r}) = \sum_{\substack{n=1 \\ n\neq m}}^{N} \sum_{m=1}^{N} |\mathbf{r}-\mathbf{r}_n||\mathbf{r}-\mathbf{r}_m| \Gamma_{\bar{u}\bar{u}}(\mathbf{r}_n,\mathbf{r}_m;\tau_{nm}), \tag{13}$$

where $\tau_{nm} \equiv |\mathbf{r}_n-\mathbf{r}|/c - |\mathbf{r}_m-\mathbf{r}|/c$ is the difference in the time delay between the source point, $\mathbf{r}$, and the two geophones locations, $\mathbf{r}_n$ and $\mathbf{r}_m$. This equation is now in a form that can be directly compared with the cross-correlation algorithm.

III. SOURCE IMAGING USING CROSS-CORRELATED SIGNALS

In this approach, signals between pairs of geophones are cross correlated. Here, the time-origin problem does not arise, because only relative delays are relevant; absolute delays with respect to a time origin do not play a role.

The signals arriving at two geophones from a particular source point will be delayed versions of each another with the delay proportional to the difference in pathlengths between the source and the two receivers. Cross correlating these two signals will yield a correlation peak with an offset, τ, equal to this relative delay. In particular, if the two geophones are located at $\mathbf{r}_n$ and $\mathbf{r}_m$ and the source location is $\mathbf{r}$,

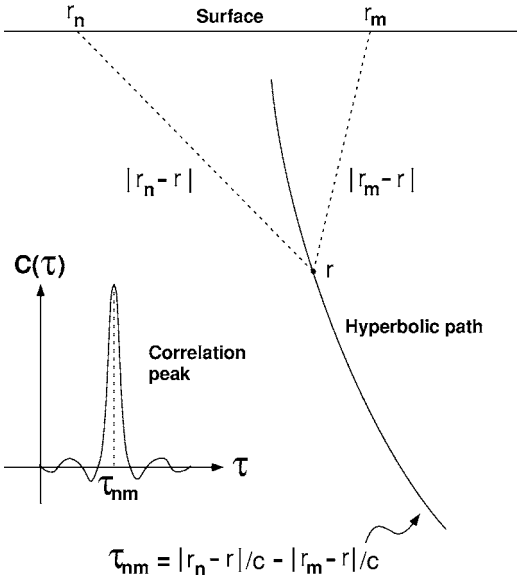


FIG. 3. The correlation peak at the delay τ_{nm} is the sum of contributions along the hyperbolic path whose foci are at the geophone locations $\mathbf{r}_n$ and $\mathbf{r}_m$.

then the correlation peak will occur at $\tau_{nm}=|\mathbf{r}_n-\mathbf{r}|/c-|\mathbf{r}_m-\mathbf{r}|/c$. The locus of points that contribute to this correlation peak (for a 2-D image) are all points that yield the same time difference τ_{nm}, that is, all points $\mathbf{r}$ that lie on a hyperbolic path whose foci are $\mathbf{r}_n$ and $\mathbf{r}_m$, as illustrated in Fig. 3.[11]

As a result, the cross correlation may be regarded as measuring the path integral of the source intensity along this hyperbolic path. The reconstruction problem is similar to the classic tomographic reconstruction problem that recovers an image from measurements of multiple overlapping line integrals of an unknown density function (in our case, source intensity). In this case, however, the line integrals are along different hyperbolas (defined by the receiver locations) rather than straight lines. This suggests a backprojection algorithm. That is, to reconstruct an image of the source intensity distribution, we filter and backproject the cross-correlated signals along these overlapping hyperbolic paths. In the 3-D version, the hyperbolic paths become surfaces (hyperboloids) and the filtered cross-correlated signals are backprojected over these surfaces. The optimal tomographic filter function will differ for the 2-D and 3-D problems.

We can justify such an approach as follows. Again, let $S(\mathbf{r},t)$ denote a random subsurface source distribution with an arbitrary spatial dependence. We assume that the source is spatially incoherent with some finite frequency bandwidth, ω_c. That is, we assume that $S(\mathbf{r},t)$ is a random process with a mutual coherence function, given by

$$\Gamma_{ss}(\mathbf{r},\mathbf{r}';t,t') \equiv \langle S(\mathbf{r},t)S(\mathbf{r}',t')^*\rangle = \sigma_s^2(\mathbf{r})\,\delta(\mathbf{r}-\mathbf{r}')C(t-t'). \tag{14}$$

Again, the angle brackets denotes a time average, $\delta(\mathbf{r}-\mathbf{r}')$ is a 3-D Dirac delta function, and $C(t-t')$ is a temporal correlation function whose frequency content is confined to $|\omega| < \omega_c$.[12] The delta function implies that the process is spatially incoherent (statistically uncorrelated from point to point). The spatial dependence $\sigma_s^2(\mathbf{r})$ may be regarded as a spatially varying source intensity. It is this function that we wish to image.

The random source generates an acoustic field, $u(\mathbf{r},t)$, that obeys the following time-dependent wave equation:

$$\nabla^2 u(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2 u(\mathbf{r},t)}{\partial t^2} = -S(\mathbf{r},t), \tag{15}$$

where c is the speed of sound. The time dependent Green's function obeys the wave equation,

$$\nabla^2 g(\mathbf{r},t|\mathbf{r}',t') - \frac{1}{c^2}\frac{\partial^2 g(\mathbf{r},t|\mathbf{r}',t')}{\partial t^2} = -4\pi\delta(\mathbf{r}-\mathbf{r}')\,\delta(t-t'), \tag{16}$$

where the Green's function is given by[13]

$$g(\mathbf{r},t|\mathbf{r}',t') = \frac{1}{|\mathbf{r}-\mathbf{r}'|}\delta[|\mathbf{r}-\mathbf{r}'|/c-(t-t')]H(t-t'), \tag{17}$$

and $H(t)$ is the unit step function. Using standard techniques, (15) and (16) can be combined to yield

$$u(\mathbf{r},t) = \int\int\int d^3\mathbf{r}' \int_{-\infty}^{t} dt'\; g(\mathbf{r},t|\mathbf{r}',t')S(\mathbf{r}',t'). \tag{18}$$

Substituting (17) into (18) and integrating with respect to t' then gives

$$u(\mathbf{r},t) = \int\int\int d^3\mathbf{r}' \frac{S(\mathbf{r}',t-|\mathbf{r}-\mathbf{r}'|/c)}{|\mathbf{r}-\mathbf{r}'|}. \tag{19}$$

We now compute the mutual coherence function of the received signals $u(\mathbf{r}_1,t_1)$ and $u(\mathbf{r}_2,t_2)$ at the two geophone locations $\mathbf{r}_1$ and $\mathbf{r}_2$:

$$\Gamma_{uu}(\mathbf{r}_1,\mathbf{r}_2;t_1,t_2) \equiv \langle u(\mathbf{r}_1,t_1)u(\mathbf{r}_2,t_2)^*\rangle. \tag{20}$$

Substituting (19) into (20), interchanging the orders of correlation and integration, and using (14), we obtain

$$\Gamma_{uu}(\mathbf{r}_1,\mathbf{r}_2;\tau) = \int\int\int \frac{\sigma_s^2(\mathbf{r})}{|\mathbf{r}_1-\mathbf{r}||\mathbf{r}_2-\mathbf{r}|} \times C(\tau-|\mathbf{r}_1-\mathbf{r}|/c+|\mathbf{r}_2-\mathbf{r}|/c)d^3\mathbf{r}, \tag{21}$$

where $\tau \equiv t_1 - t_2$.

This is a key equation. If the bandwidth of the source is broad, then the correlation function, $C(\tau)$, will be narrow. If we momentarily replace $C(\tau)$ with a delta function, i.e., let $C(\tau)\approx\delta(\tau)$ in (21), then $\Gamma_{uu}(\mathbf{r}_1,\mathbf{r}_2;\tau)$ becomes a (weighted) surface integral through the function $\sigma_s^2(\mathbf{r})$ over a hyperboloid whose foci are located at $\mathbf{r}_1$ and $\mathbf{r}_2$. In practice, $C(\tau)$ will have a finite width roughly equal to $1/\omega_c$. This implies a spatial width of approximately the speed of sound divided by the source bandwidth. That is, the "thickness" of the hyperboloid is roughly $\Delta r = 2\pi c/\omega_c$. This is an approximate theoretical upperbound on the image resolution.

As motivation for the reconstruction algorithm, we Fourier transform (21) with respect to τ, giving

$$\tilde{\Gamma}_{uu}(\mathbf{r}_1,\mathbf{r}_2;\omega)=\tilde{C}(\omega)\int\int\int\frac{\sigma_s^2(\mathbf{r})}{|\mathbf{r}_1-\mathbf{r}||\mathbf{r}_2-\mathbf{r}|}\times\exp[i\omega(|\mathbf{r}_1-\mathbf{r}|-|\mathbf{r}_2-\mathbf{r}|)/c]d^3\mathbf{r}, \quad (22)$$

where $\tilde{C}(\omega)$ is the Fourier transform of $C(\tau)$. The image is then obtained by multiplying (22) by $\tilde{b}(\omega)|\mathbf{r}'-\mathbf{r}_1||\mathbf{r}'-\mathbf{r}_2|\times\exp[-i\omega(|\mathbf{r}_1-\mathbf{r}'|-|\mathbf{r}_2-\mathbf{r}'|)/c]$, where $\mathbf{r}'$ is the image point, and integrating over the signal bandwidth and aperture. This gives the following formula for the image (after dropping the prime on $\mathbf{r}'$):

$$\hat{\sigma}_s^2(\mathbf{r})=\int\int_A d^2\mathbf{r}_1\,d^2\mathbf{r}_2\int_{-\omega_c}^{\omega_c}d\omega\,\tilde{b}(\omega)F_\omega(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r})\times\exp[-i\omega(|\mathbf{r}_1-\mathbf{r}|-|\mathbf{r}_2-\mathbf{r}|)/c]. \quad (23)$$

where

$$F_\omega(\mathbf{r}_1,\mathbf{r}_2;\mathbf{r})\equiv|\mathbf{r}-\mathbf{r}_1||\mathbf{r}-\mathbf{r}_2|\tilde{\Gamma}_{uu}(\mathbf{r}_1,\mathbf{r}_2;\omega). \quad (24)$$

Here A denotes the aperture (or receiving array), and $\tilde{b}(\omega)$ is a tomographic filter; in two dimensions, $b(\omega)=|\omega|$[14] and in three dimensions, $b(\omega)=\omega^2$.[15]

The justification for (23) is carried out by substituting (22) into (23) and interchanging orders of integration to obtain an expression of the form

$$\hat{\sigma}_s^2(\mathbf{r})=\int\int\int\sigma_s^2(\mathbf{r}')P(\mathbf{r},\mathbf{r}')d^3\mathbf{r}', \quad (25)$$

where $P(\mathbf{r},\mathbf{r}')$ is the image point-spread function. One can show that, in the limit of an infinite bandwidth and aperture, $P(\mathbf{r},\mathbf{r}')$ becomes delta like for $\mathbf{r}=\mathbf{r}'$, as desired. An analytical derivation of $P(\mathbf{r},\mathbf{r}')$ was derived in Ref. 8 using this formula for the special case of a 2-D circular geometry. In the latter geometry, the receiving points, $\mathbf{r}_1$ and $\mathbf{r}_2$, lie on the circumference of a circle enclosing the region to be imaged. It was shown in Ref. 8 that $P(\mathbf{r},\mathbf{r}')$ tends toward a 2-D Dirac delta function in the limit of an infinite bandwidth (i.e., as $\omega_c\to\infty$).

Equation (23) can be transformed into a backprojection formula by substituting

$$\tilde{\Gamma}_{uu}(\mathbf{r}_1,\mathbf{r}_2;\omega)=\int\Gamma_{uu}(\mathbf{r}_1,\mathbf{r}_2;t)e^{i\omega t}\,dt \quad (26)$$

into (24) and (23), which gives

$$\hat{\sigma}_s^2(\mathbf{r})=\int\int_A d^2\mathbf{r}_1\,d^2\mathbf{r}_2|\mathbf{r}-\mathbf{r}_1||\mathbf{r}-\mathbf{r}_2|\bar{\Gamma}_{uu}[\mathbf{r}_1,\mathbf{r}_2;(|\mathbf{r}_1-\mathbf{r}|-|\mathbf{r}_2-\mathbf{r}|)/c]. \quad (27)$$

In this formula, $\bar{\Gamma}_{uu}(\mathbf{r}_1,\mathbf{r}_2;t)$ is a filtered version of $\Gamma_{uu}(\mathbf{r}_1,\mathbf{r}_2;t)$, that is, $\bar{\Gamma}_{uu}(\mathbf{r}_1,\mathbf{r}_2;t)=b(t)*\Gamma_{uu}(\mathbf{r}_1,\mathbf{r}_2;t)$, where $*$ denotes convolution and $b(t)$ is the inverse Fourier transform of $\tilde{b}(\omega)$.

Equation (27) represents a filtered-backprojection algorithm along hyperbolic paths in two dimensions and over surfaces (hyperboloids) in three dimensions. The discrete version of (27) is given as follows. Suppose we have an array of N geophones on the surface at coordinates $\mathbf{r}_n$, $n=1,2,\ldots,N$. Then (27) reduces to

$$\hat{\sigma}_s^2(\mathbf{r})=\sum_{n=1}^{N}\sum_{\substack{m=1\\n\neq m}}^{N}|\mathbf{r}-\mathbf{r}_n||\mathbf{r}-\mathbf{r}_m|\bar{\Gamma}_{uu}(\mathbf{r}_n,\mathbf{r}_m;\tau_{nm}), \quad (28)$$

where $\tau_{nm}\equiv|\mathbf{r}_n-\mathbf{r}|/c-|\mathbf{r}_m-\mathbf{r}|/c$.

Note that (28) is identical in form to (13). The only difference is that, in (13), $\Gamma_{\bar{u}\bar{u}}$ is backprojected, whereas in (28), $\bar{\Gamma}_{uu}$ is backprojected. That is, in the former case, the cross-correlation is performed on the filtered signals, $\bar{u}$, whereas in the latter case, the filtering is performed after the cross-correlation. Simulations have shown that the distinction between these two cases is not particularly important in affecting image quality or resolution unless the emitted signals are extremely broadband (i.e., have a very large fractional bandwidth). Typically, for artificial sources (e.g., emissions from an underground facility), the fractional bandwidths are not large. In such cases, as simulations demonstrate, the application of an optimal "tomographic filter function" of the form $\tilde{b}(\omega)=|\omega|$ (for two dimensions) or $b(\omega)=\omega^2$ (for three dimensions) prior to backprojection does not significantly affect image quality.

IV. SIMULATED RECONSTRUCTIONS

To convey some idea of the quality and resolution of images that can be obtained using the reconstruction formula (28), we applied this equation to simulated data. We assumed three isolated point sources at the locations indicated in Fig. 4(a), each emitting continuous random noise over a bandwidth of 50 to 200 Hz. The signals were assumed recorded for a duration of one second at a sampling rate of 1 kHz using a geophone array with 21 elements spaced at 4 m intervals. A sound velocity of 500 m/s was assumed. The received signals were cross-correlated between all geophone pairs with a 12 m separation or greater. The cross-correlated signals were then filtered with the tomographic filter $|\omega|$ and backprojected, as prescribed by (28). The resultant image is shown in Fig. 4(b).

Note that the source images are somewhat elongated vertically. That is, the lateral resolution appears somewhat better than the longitudinal (depth) resolution. This is not surprising since lateral resolution is determined by the aperture size (array length) and the bandwidth of the noise source, whereas longitudinal resolution in a passive system is determined primarily by the system's ability to detect the curvature of the wave front over the aperture length. When the source depth is large compared to the aperture, the wave front curvature becomes less pronounced and a loss of longitudinal resolution is expected.

The images in Fig. 4(b) were computed under the assumption of a known sound velocity. It is interesting to investigate the effect of an incorrect velocity estimate on image quality. If the velocity is in error by no more than about 40% or 50% percent, the effect on the image is primarily a vertical shift in the placement of the source images with a relatively minor image distortion. Severe image degradation

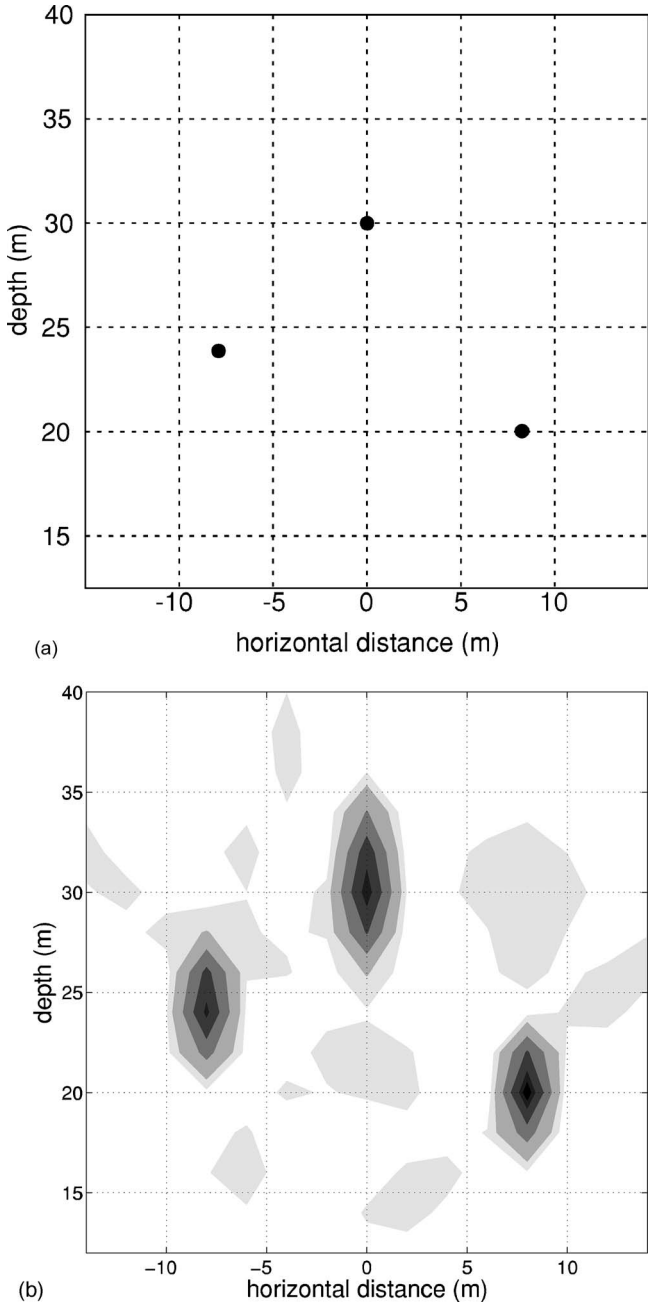


FIG. 4. (a) Three source locations assumed to emit continuous random noise with a bandwidth between 50 and 200 Hz. Depth is measured upward. The array consists of 21 elements on the surface with a 4 m spacing. The array runs along the lower boundary of the imaged region. (b) Image of the three sources using the imaging formula (28).

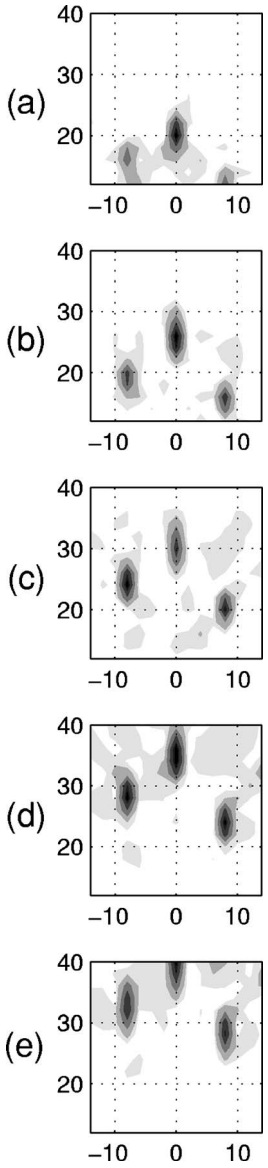


FIG. 5. Figure (a)–(e) are simulated images using an assumed velocity 80%, 90%, 100%, 110%, and 120% of the true velocity, respectively. We see that an underestimation of the velocity, for example, results in a displacement of the image closer to the array.

begins to occur as the velocity error exceeds about 50%. Figures 5(a)–5(e) show images reconstructed under the assumption of a velocity 80%, 90%, 100%, 110%, and 120% of the true velocity, respectively. Thus, an overestimation of the velocity, for example, tends to shift the placement of the sources vertically (i.e., to overestimate the source depth).

V. FIELD DATA

In this section, the two algorithms were compared using a set of field data generated by a hammer blow. The data were acquired with a geophone array that recorded surface waves generated by the hammer source at one location on the ground. The received signal bandwidth was determined to extend from about 5 to 50 Hz by spectral analysis. The resultant images are two dimensional and derived from surface waves only. Surface waves are dispersive and thus, in principle, not optimal for imaging since the algorithms are not designed to account for wave dispersion. Subsurface body waves in three dimensional are not dispersive and should, in theory, provide better images. Nevertheless, as shown below, quite good images of the point source were obtained using this data set.

The data were acquired using a seismograph connected to a 13 element array in the shape of an "L," as illustrated in Fig. 6, with a geophone spacing of 15 ft. In the figure, the area enclosed by the dotted square denotes the imaged region. In this experiment, signals were generated by striking a metal plate on the ground with a hammer at the point labeled "source" in Fig. 6. Data were continually recorded over a

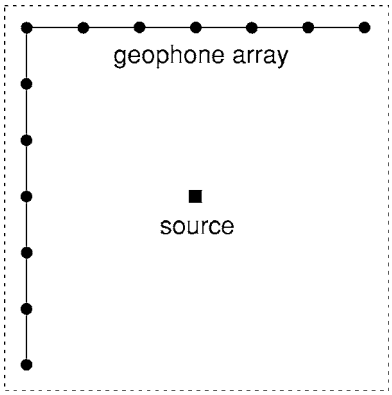


FIG. 6. Geophone array used to acquire field data. The dotted area is the imaged region shown in Figs. 7 and 8.

30 s interval while the ground was struck 12 times; no triggering of the seismograph was used. Figures 7(a) and 7(b) show the reconstructed image using the time-domain migration algorithm shown in both contour and pixel formats, where the latter is a 15 by 15 pixel image. A surface-wave speed of 600 ft/s placed the source in the correct location. Using the same data, the image obtained using the cross-correlation algorithm is shown in Figs. 8(a) and 8(b). These images were reconstructed without an "optimal" tomographic filter, since such filtering was found to make little difference in image quality. The images show that both algorithms perform similarly, as anticipated.

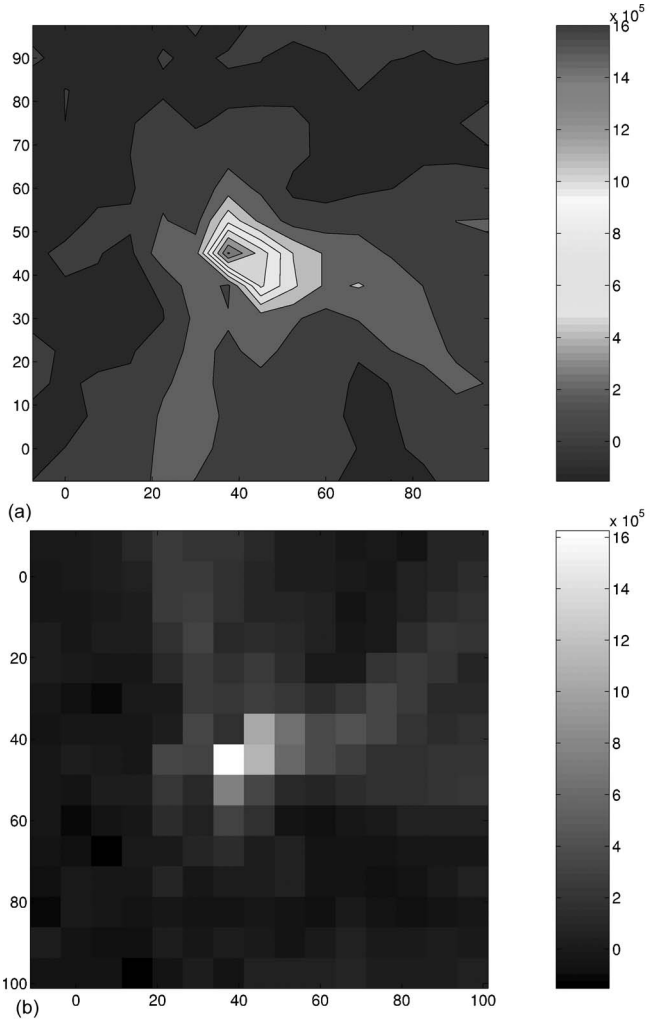


FIG. 7. Image obtained from field data using the time-domain migration algorithm. (a) Contour format; (b) pixel format. Units are in feet.

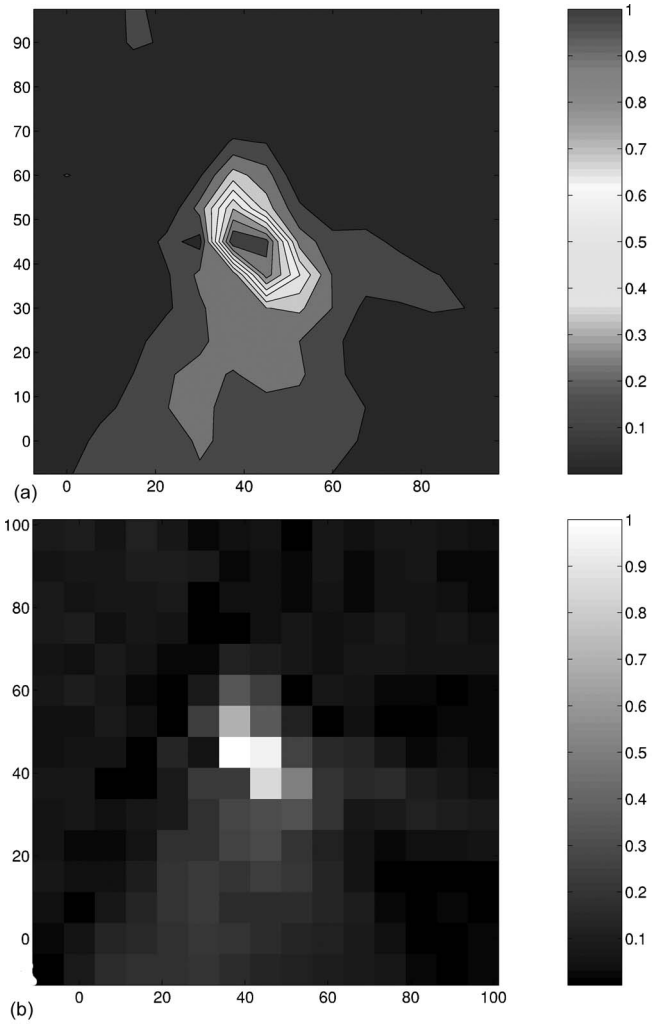


FIG. 8. Image obtained from field data using the cross-correlation algorithm. (a) Contour format; (b) pixel format. Units are in feet.

VI. CONCLUSION

Two passive imaging algorithms were examined, one based on time-domain migration[7] and the other on cross-correlation.[8] A key difference between active and passive imaging is that, in the latter case, no time origin exists with which to reference propagation delays. In the first algorithm examined, an image is obtained by averaging with respect to this time origin. The averaging must be done on an intensity basis, since the "amplitude image" has zero mean. After removal of the dc component of the intensity image, this averaging process was shown to reduce to a form essentially equivalent to the cross-correlation algorithm. The latter algorithm was developed for imaging acoustic emission sources in materials, but could also find applications in the imaging of underground sources. Such sources may arise due to scattering of ambient noise or from emissions generated by machinery inside a manmade facility. In either case, im-

age resolution is ultimately limited by the temporal correlation distance of the source, that is, the speed of sound divided by the source bandwidth.

We conclude by noting again that the algorithms and simulations described here have been idealized for simplicity. Two key idealizations are the assumption of a constant sound velocity and the neglect of mode conversion effects at the surface. For example, the algorithms do not account for the partial mode conversion of the source-generated body waves into surface waves at the air-earth interface. Surface waves can be treated in several ways: (1) by filtering using an optimally designed 2-D surface array, (2) by treating mode conversion as an incoherent noise source, which may tend to average out over time, or, finally, (3) by attempting to incorporate mode conversion into the propagation model. The latter approach could, in principle, be achieved by replacing the acoustic Green's function with an elastic-wave half-space Green's function. This would vastly increase the computational cost of any imaging algorithm but, as computer power continues to advance, may one day become feasible.

[1]G. T. Schuster, J. Yu, J. Sheng, and J. Rickett, "Interferometric/daylight seismic imaging," Geophys. J. Int. **157**, 838–852 (2004).

[2]K. Wapenaar, J. W. Thorbecke, and D. Draganov, "Relations between reflection and transmission responses of 3-D inhomogeneous media," Geophys. J. Int. **156**, 179–194 (2004).

[3]K. Wapenaar, "Synthesis of an inhomogeneous medium from its acoustic transmission response," Geophysics **68**, 1756–1759 (2003).

[4]K. Wapenaar, D. Draganov, J. van der Neut, and J. Thorbecke, "Seismic interferometry: a comparison of approaches," extended abstract, *74th Annual Meeting of the SEG*, Denver, 10–15 October 2004.

[5]D. Draganov, K. Wapenaar, B. Artman, and B. Biondi, "Migration methods for passive seismic data," extended abstract, *74th Annual Meeting of the SEG*, Denver, 10–15 October 2004.

[6]J. F. Claerbout, "Synthesis of a layered medium from its acoustic transmission response," Geophysics **33**, 264–269 (1968).

[7]S. J. Norton and I. J. Won, "Time exposure acoustics," IEEE Trans. Geosci. Remote Sens. **38**, 1337–1343 (2000).

[8]S. J. Norton and M. Linzer, "Backprojection reconstruction of random source distributions," J. Acoust. Soc. Am. **81**, 977–985 (1987).

[9]We note that the assumption of a spatially incoherent source may appear to fail if the "sources" are scatterers illuminated, for example, by an incident plane wave. However, if the direction of incidence is variable, then spatial incoherence should effectively hold when averaging over a time on the order of the reciprocal of the source bandwidth or greater.

[10]For generality, we allow the fields to be complex, because we may choose $u(\mathbf{r},t)$ to represent either a real quantity or its analytic signal, which is complex.

[11]This hyperbola is not to be confused with the hyperbola shown in Fig. 2, which is a hyperbolic path (or hyperboloid in three dimensions) in data space, that is, in $(\mathbf{r}_n,t)$ space. The hyperbola (or hyperboloid in three dimensions) of Fig. 3 resides in image space.

[12]Strictly speaking, the form of (14) implies a statistically stationary process, at least over the observation interval. In practice, this is a reasonable approximation when the observation interval is long compared to the reciprocal of the source bandwidth.

[13]H. W. Wyld, *Mathematical Methods for Physics* (Addison-Wesley, New York, 1977), Chap. 8.

[14]G. T. Herman, in *Image Reconstruction from Projections: Implementation and Applications* (Springer-Verlag, New York, 1979).

[15]S. J. Norton and M. Linzer, "Ultrasonic reflectivity imaging in three dimensions: Exact inverse scattering solutions for plane, cylindrical, and spherical apertures," IEEE Trans. Biomed. Eng. **BME-28**, 202–220 (1981).

Middle ear ossicles motion at hearing thresholds with air conduction and bone conduction stimulation

Stefan Stenfelt[a)]
Department of Signals and Systems, Chalmers University of Technology, Göteborg, Sweden

(Received 20 October 2005; revised 9 February 2006; accepted 9 February 2006)

Hearing threshold data with bone conduction and air conduction stimulation are combined with physiological and mechanical measurements of the middle ear ossicles vibration to compute the vibration level of the ossicles at threshold stimulation. By comparing the displacements of the stapes footplate with the two stimulation modalities and assuming the vibration of the stapes footplate to be the input to the cochlea when stimulation is by air conduction, the importance of middle ear ossicles inertia with bone conduction stimulation is evaluated. Given the limitations of the analysis, the results indicate that the inertia of the middle ear is not an important contribution to the perception of BC sound for frequencies below 1.5 kHz; it seems to contribute to perception of bone conducted sound between the frequencies 1.5 and 3.5 kHz. At frequencies above 4 kHz, the analysis failed since the input to the cochlea is probably not through the oval window with bone conduction stimulation. Comparison of basilar membrane vibration data verified the calculations for frequencies between 0.8 and 3.5 kHz. It was also found that the fluid flow at the round window, rather than at the oval window, reflects the stimulation of the basilar membrane with bone conduction stimulation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184225]

PACS number(s): 43.64.Bt [BLM] Pages: 2848–2858

I. BACKGROUND

Behavioral hearing thresholds are the stimulation level where a person starts to perceive a sound. The hearing thresholds for young healthy persons have been measured and their average was used to derive a standard for normal persons hearing ability. This was done for air conduction (AC) stimulation as well as for bone conduction (BC) stimulation (ISO 226, 2003; ISO 389-3, 1994).

It is normally assumed that an AC sound and a BC sound stimulates the basilar membrane and the hearing receptors similarly. This assumption comes from the ability to cancel a pure AC tone with a BC tone (von Békésy, 1932; Khanna *et al.*, 1976), the ability to produce distortion product otoacoustic emissions using AC and BC stimulation (Purcell *et al.*, 1998), and the similarity of resonance frequencies on the basilar membrane with AC and BC excitation (Stenfelt *et al.*, 2003a). Hence, it can be assumed that the basilar membrane motion at threshold levels is equal whether the stimulation is by AC or BC.

The major difference between AC and BC stimulation is that AC stimulation is basically a sound pressure in the ear canal that is transmitted to a sound pressure in the vestibule of the cochlea by the TM and middle ear ossicles, whereas BC stimulation involves several pathways. The most important physical phenomena involved in BC stimulation are believed to be (1) sound radiation into the ear canal, (2) inertial motion of the middle ear ossicles, (3) inertial motion of the fluid in the cochlea, and (4) compression and expansion of the bone encapsulating the cochlea (Stenfelt and Good, 2005a). The first two effects would stimulate the cochlea via the oval window and stapes footplate. Hence, this stimulation is, from the cochlea point of view, similar to AC stimulation.

It has previously been shown that the outer ear component (sound radiation into the ear canal) does not normally contribute to hearing by BC (Stenfelt *et al.*, 2003b). However, the inertial effects of the middle ear ossicles may contribute to the hearing perception when the stimulation is by BC. It has further been shown that the fluid flow at the oval window and round window is unequal for BC stimulation while the fluid flow for AC stimulation at the two windows is similar and with opposite phases (Stenfelt *et al.*, 2004). This indicates that the stimulation of the cochlea is more complicated when the stimulation is by BC than AC. The aim of this paper is to compare the displacement of the malleus umbo, stapes footplate, round window, and basilar membrane at human behavioral hearing thresholds for AC and BC stimulation. Such comparison can give insight to different contributing parts for BC stimulation.

II. METHODS

The data used in this paper are mainly taken from previously published results. The methods for each measurement are therefore not given in detail; the interested reader is encouraged to read those articles for a detailed description of the different methods used. However, a brief overview of the methods is given here for the reader to grasp some of the limitations of the current analysis.

The behavioral thresholds are taken from the standard (BC, ISO 389-3, 1994; AC, ISO 226, 2003). The other data are either taken from isolated human temporal bone specimens (middle ear ossicle, oval window, round window, and

[a)]Present address: Dept of Neuroscience and Locomotion, Division of Technical Audiology. Linköping University Hospital, Linköping, Sweden. Electronic mail: stefan.stenfelt@inr.liu.se

basilar membrane motion) using both AC and BC stimulation, or live humans and human cadaver heads for vibration transmission measurements in the skull (BC). It is only for the skull impedance measurement at the skin (Flottorp and Solberg, 1976) that live human data are used. It should be noted that all data are for the human and that the results are not necessarily applicable for animals. Also, the data derived for BC using isolated temporal bones are primarily stimulated in a single direction (same as the low-frequency piston-like motion of the stapes) while measurement of the cochlear promontory in whole heads with BC stimulation has shown vibration pattern with motions in all three space dimensions (Stenfelt and Good, 2005b). Further, the simplification by using isolated temporal bone specimens with BC stimulation may introduce artifacts in the results since the stimulation of the petrous part of the temporal bone encapsulating the cochlea can differ between isolated temporal bone specimen measurement and whole head measurements.

With AC stimulation, the vibrations of the middle ear ossicles, the round window, and the basilar membrane is measured with a laser Doppler vibrometer (LDV) (HLV 1000, Polytec, Waldbronn, Germany), while the sound pressure in-front of the TM is measured with a probe-tube microphone (ER-7C, Etymotic Research, Elk Grove Village, IL). When the stimulation is by BC, the input impedance at the stimulation position is measured with an impedance head (B&K type 8001, Brüel and Kjær, Nærum, Denmark), the vibration of the whole head is measured with either a three-axial (3D) accelerometer (Endevco model 7253C-10, Endevco, San Juan Capistrano, CA) or with the LDV while the vibrations of the middle ear ossicles, temporal bone, round window, and basilar membrane are measured with the LDV when the isolated temporal bone is rigidly attached to a shaker (B&K type 4810 mini-shaker).

III. AIR CONDUCTION STIMULATION

The minimum audible field (MAF) is defined as the lowest pressure of a sound field where a normal hearing human perceives a sound. The MAF is measured as the sound field pressure at the center of the head when the head is absent in the sound field. The sound pressure at the TM differs from the sound pressure at the center of the head: diffractions of the head as well as resonances of the ear canal alter the sound pressure. Figure 1 shows the MAF according to the standard ISO 226 (2003) as well the sound pressure at the TM at MAF with corrections for the head and ear canal according to Shaw (1974). The MAF is given for binaural hearing and the threshold sound pressure at the TM must be compensated accordingly. Killion (1978) estimated the monaural threshold sound pressure at the TM which is also shown in Fig. 1. The sound pressure thresholds from Killion (1978) are used in this study as reference for air conduction (AC) stimulation and are referred to as the minimum audible pressure (MAP).

The sound energy from the ear canal sound pressure is transmitted to the cochlea and the basilar membrane by the middle ear ossicles. This transmission is affected by several factors. Figure 2(a) shows the displacement of the malleus

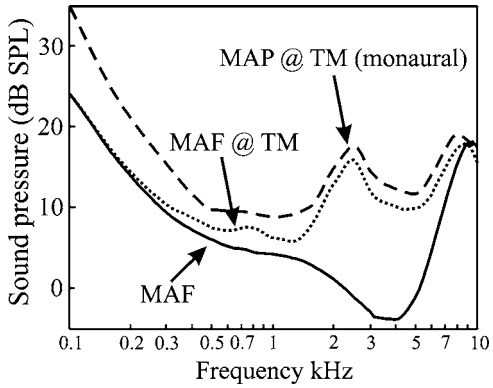


FIG. 1. The minimum audible field according to ISO 226 (2003) (solid line), the minimum audible field at the TM (dotted line), and the minimum audible pressure (dashed line, from Killion, 1978). The curves are interpolated between the frequencies given in the standard ISO 226 and in Killion (1978).

umbo and stapes footplate when the stimulation is a sound pressure of 80 dB SPL at the TM (ossicle displacement at 80 dB SPL, OD_{AC_80}). The magnitude data in Fig. 2(a) are taken from isolated temporal bone preparations (Hato *et al.*, 2003) and mathematically compensated for the artificial expansion of the middle ear cavity. When the ossicles motion was measured in Hato *et al.* (2003), post-mortem isolated temporal bone specimens were used. In order to execute the measurements, a hole was made through the mastoid air cell system and into the antrum where an enlarged opening of the facial recess into the middle ear cavity was made. This opening into the antrum was sealed airtight with a glass cover. However, instead of the natural approximately 0.5 cc tympanic cavity with adjacent antrum and air cells, the total volume of the large, artificially opened and sealed, middle ear cavity is approximately three to four times that of the original (1.5 to 2 cc).

Such alteration of the middle ear cavity affects the motion of the middle ear ossicles. Since subjective hearing thresholds are obtained with an intact middle ear and the vibration measurements of the ossicles in Hato *et al.* (2003) is conducted with an altered middle ear cavity, the results cannot be directly linked. In order to combine these two data sets, the effect from the alteration of the middle ear cavity is estimated using the middle ear model by Zwislocki (1962). Figure 2(b) shows the modeled ossicles motion with an intact middle ear cavity and antrum related to the modeled ossicles motion when the middle ear cavity is altered as in Hato *et al.* (2003). More details of the calculations deriving the correction curve in Fig. 2(b) are found in the appendix. It should be noted that the data in Fig. 2(a) is already corrected according to the curve in Fig. 2(b) meaning that Fig. 2(a) shows the ossicle displacement at 80 dB SPL input with an intact middle ear cavity.

The displacement of the stapes footplate and malleus umbo at hearing thresholds can be calculated by combining Figs. 1 and 2(a). The data used in Fig. 1 is the MAP and the threshold displacements in Fig. 3 is calculated as

$$OD_{MAP} = OD_{AC\_80} \times 10^{(MAP-80)/20}, \quad (1)$$

where OD means ossicle displacement.

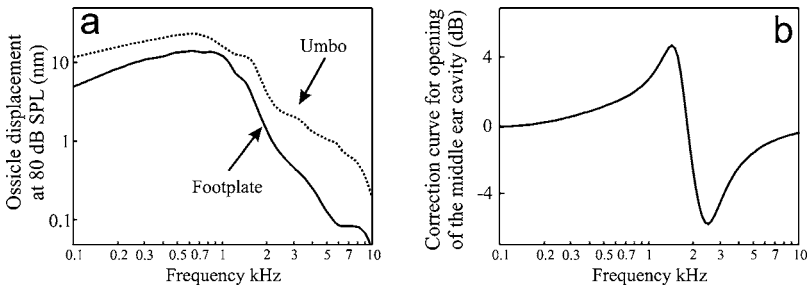


FIG. 2. (a) Displacement of the stapes footplate (solid line) and malleus umbo (dotted line) when the stimulation is a sound pressure of 80 dB SPL at the TM (OD_{AC_80}). The data is taken form Hato *et al.* (2003) and compensated for the effect of altering the space of the middle ear cavity. (b) The correction for the ossicle vibration caused by altering the middle ear volume. Using the middle ear model of Zwislocki (1962) the curve estimates the ratio of ossicle motion between an intact middle ear cavity and an open and resealed cavity as in the temporal bone experiments used to derive the ossicle vibration data (Hato *et al.*, 2003). A more thorough explanation for the calculations deriving the correction curve is given in the text and the Appendix .

IV. BONE CONDUCTION STIMULATION

When bone conduction (BC) threshold testing is conducted, a BC transducer with a circular surface area of 175 mm^2 is pressed against the skin covered mastoid with a static force of approximately 5.4 Newtons. With this setup, the dynamic force produced by the transducer at threshold levels is used as the reference for BC hearing. The reference equivalent threshold force levels (RETFL) given in the standard ISO 389-3 (1994) are shown in Fig. 4. Actually, the RETFL given in ISO 389-3 (1994) for BC audiometry are equal to the dynamic force levels produced by a BC transducer (Radioear B71) on a mechanical coupler (B & K type 4930 artificial mastoid) at electrical input levels directly corresponding to subjectively measured BC thresholds of young otologically normal listeners.

Although the transducer presses the skin onto the skull bone beneath it, the skin together with the skull act as a mass-spring system affecting the force transmission from the transducer to the skull bone. A lumped element impedance model of the vibration transmission through the skin and subcutaneous tissue to the skull bone is shown in Fig. 5(a). The parameters for the skin model (M_{SKIN}, C_{SKIN}, and R_{SKIN}) are taken from the skin covered mastoid impedance measurements made by Flottorp and Solberg (1976) and the skull impedance (Z_{SKULL}) is taken from mechanical point impedance measurements reported in Stenfelt and Good (2005b). Since

$$|Z_{SKULL}| \gg \left| \frac{1}{j\omega C_{SKIN}} + R_{SKIN} \right|$$

the force attenuation through the skin can be approximated with

$$\mathrm{Att}_{SKIN} = \frac{(1 - \omega^2 M_{SKIN} C_{SKIN}) + j\omega R_{SKIN} C_{SKIN}}{1 + j\omega R_{SKIN} C_{SKIN}}, \qquad (2)$$

where ω is the angular frequency and j is the imaginary constant.

The attenuation level of the force through the skin according to the model is shown in Fig. 5(b). The model in Fig. 5 predicts the force at the skull bone to be equal or better than at the skin position for frequencies up to about 4 kHz. At higher frequencies, the force attenuation caused by the skin increases with frequency and becomes approximately 8 dB at 10 kHz. The model prediction in Fig. 5 is in fair agreement with psychoacoustic measurements of the skin attenuation by Stenfelt and Håkansson (1999). However, the data in that study cannot be compared directly with the model predictions; the threshold data at the skin are obtained with the transducer placed at the mastoid behind the pinna whereas the BC thresholds for the skull bone attachment are obtained at the position of the bone-anchored hearing aid (www.entific.com), some 55 mm behind the ear canal opening. This difference in positions can affect the outcome. In an investigation of the human mechanical point impedance at the mastoid with both covering skin and direct skull bone

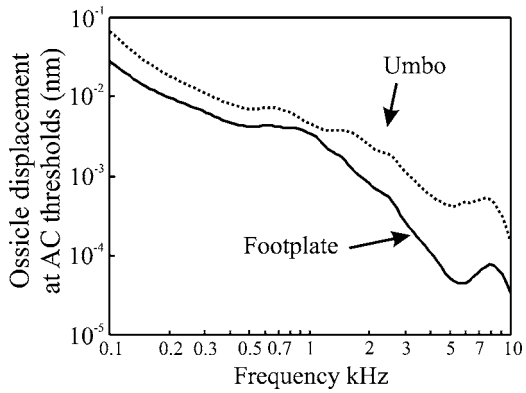


FIG. 3. Displacement of the stapes footplate (solid line) and malleus umbo (dotted line) when the stimulation is at AC behavior thresholds (OD_{MAP}). The result is derived from the ossicle displacement data in Fig. 2 and the minimum audible sound pressure (MAP) presented in Fig. 1.

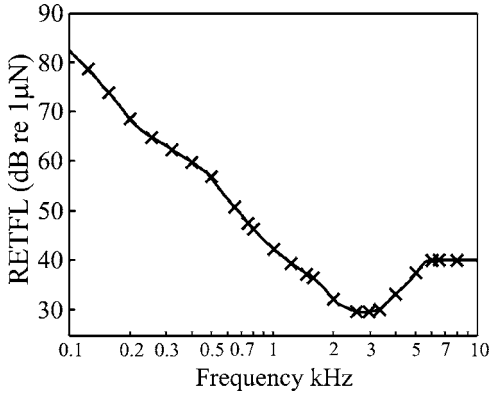


FIG. 4. Reference equivalent force threshold levels (RETFL) according to the standard ISO 389-3 (1994). The crosses indicate the frequencies and values stated in the standard. The solid line is the interpolated values used for the calculations.

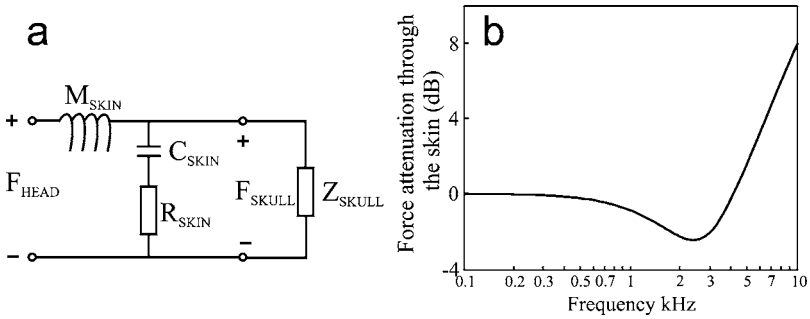


FIG. 5. (a) Lumped element impedance model of vibration transmission from a BC transducer through the skin to the skull bone. F_{HEAD} represents the dynamic force from the BC transducer and F_{SKULL} is the resulting force acting on the skull bone beneath the BC transducer. The impedance of the skull bone is modeled by Z_{SKULL} whereas M_{SKIN}, C_{SKIN}, and R_{SKIN} is the mass, compliance, and dissipation of the skin, respectively. (b) The resulting attenuation from F_{HEAD} to F_{SKULL} according to the model in (a) with skin parameters from Flottorp and Solberg (1976) and measured skull impedance from Stenfelt and Good (2005b).

attachment, Håkansson *et al.* (1986) derived a mathematical model for the impedance. Although Håkansson *et al.* (1986) used a slightly smaller contact area and greater static force between the transducer and skin than prescribed in the standard for BC threshold measurements, their computed model for the skin attenuation produces similar results as in Fig. 5(b). Further, the impedance data for direct skull attachment were, for corresponding positions, similar in Håkansson *et al.* (1986) using live humans with Stenfelt and Good (2005b) study using isolated cadaver heads; this indicates that, at least for the vibration input, isolated cadaver heads is a good substitute for live human heads conducting BC experiments.

The transmission of vibration between the skull bone surface at the mastoid and the cochlea was measured with a vibration transducer attached to the skull bone at the mastoid by means of a screw. With the stimulation at the mastoid, the vibration of the cochlear promontory was obtained with a three-axial accelerometer rigidly glued to the cochlear promontory in isolated cadaveric whole heads (Stenfelt and Good, 2005b). The excitation position is the same as the position applying the bone transducer for the RETFL data. To obtain the displacement of the cochlear promontory (PD) with BC stimulation at threshold, the data from Figs. 4 and 5 is combined with the skull vibration transmission from Stenfelt and Good (2005b) according to

$$PD_{RETFL} = \frac{PD_{1NEWTON} \times 10^{(RETFL-120)/20}}{Att_{SKIN}}, \quad (3)$$

where $PD_{1NEWTON}$ is the displacement of the cochlear promontory with 1 Newton stimulation according to Stenfelt and Good (2005b). PD_{RETFL} is shown in Fig. 6(b).

Although the cochlear promontory vibrates in all three directions of space, only the vibration in the direction in line with the low-frequency pistonlike stapes motion is used for the analysis here. This simplification of only using results from one direction introduces uncertainties that may produce errors in the analysis. However, the vibration of the cochlear promontory is some 5 dB greater in this measurement direction relative to the perpendicular directions when the stimulation is at the ipsilateral mastoid and, also, the relative middle ear ossicle motion is around 5 to 10 dB greater with the stimulation in this direction than with the stimulation in the perpendicular plane (Stenfelt and Good, 2005b; Stenfelt *et al.*, 2002). This means that the vibration of the middle ear ossicles is greatest in the low-frequency pistonlike stapes motion direction when the excitation is at the ipsilateral mastoid; vibrations of the ossicles in the other directions is assumed to only have a minor influence.

The relative motion of the middle ear ossicles with BC vibration was obtained by stimulation of isolated temporal bone specimens in the direction of the low-frequency pistonlike motion of the stapes while measuring the velocity of the ossicles and surrounding bone. The relative motion is then calculated as the velocity difference between the ossicle and bone divided by the stimulation velocity (Stenfelt *et al.*, 2002). This relative motion is shown in Fig. 6(a) (OD_{BC_REL}). Using the knowledge of the promontory displacement at BC thresholds [Eq. (3)] shown in Fig. 6(b), the ossicles displacement at BC threshold stimulation can be calculated as

$$OD_{RETFL} = OD_{BC\_REL} \cdot PD_{RETFL}. \quad (4)$$

The ossicles displacement at BC threshold stimulation is shown in Fig. 6(b).

The ossicle displacement at BC thresholds shown in Fig. 6(b) is derived from measurements in isolated cadaver heads and temporal bone specimens. It may be argued that isolated temporal bones are not a good model for BC stimulation of

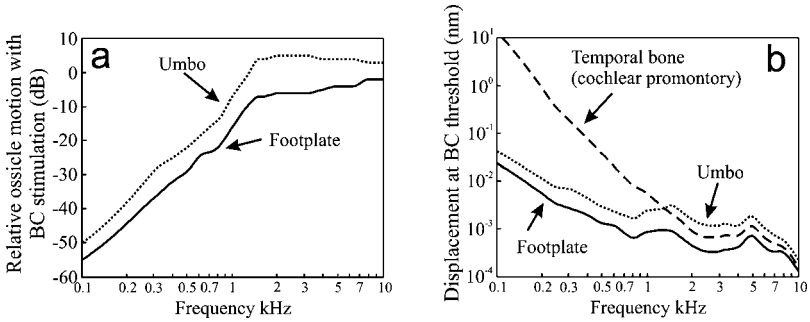


FIG. 6. (a) The relative motion between the stapes and cochlear promontory (solid line) and the relative motion between the umbo and surrounding bone (dotted line) when the stimulation is by BC in the direction of the low-frequency piston-like motion of the stapes. The data are taken from Stenfelt *et al.* (2002). (b) The dashed line shows the cochlear promontory bone displacement in the direction of the low-frequency pistonlike motion of the stapes when the stimulation is at behavioral thresholds at the ipsilateral mastoid (PD_{RETFL}). By combining the promontory displacement at threshold stimulation data (dashed line) with the relative ossicle displacements in (a) the stapes footplate (solid line) and malleus umbo (dotted line) displacements at behavioral BC threshold stimulation is estimated (OD_{RETFL}).

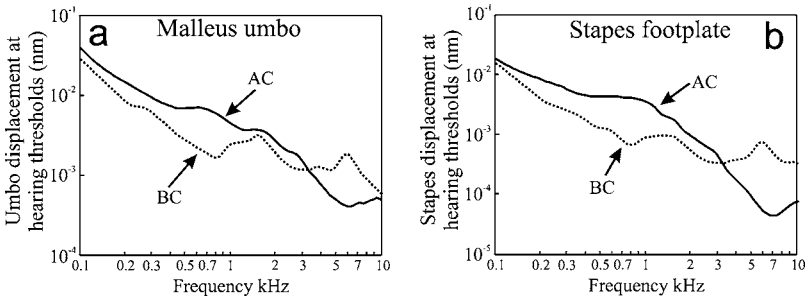


FIG. 7. The ossicle displacement at hearing threshold stimulation (OD_{MAP} and OD_{RETFL}). (a) Malleus umbo displacement and (b) stapes footplate displacement. The displacements using AC stimulation are shown with solid lines and the displacements using BC stimulation are shown with dotted lines.

compression or some other stimulation mode unaccounted for. However, stapes footplate as well as round window membrane motion has been measured in limited whole cadaver heads using BC stimulation producing vibration results similar to average results from isolated temporal bone specimens (Stenfelt *et al.*, 2002, 2004). Further, unpublished results of cochlear promontory vibration data obtained with a LDV aimed through an enlarged ear canal and open middle ear from two patients in our laboratory showed similar results as the cadaver head vibration data in Stenfelt and Good (2005b).[1] Although more experiments are required to verify these cadaver and temporal bone models for BC studies, they seem to be valid for first order effects. However, some caution is recommended interpreting the results in the coming sections.

V. RESULTS AND DISCUSSION

A. Middle ear ossicle motion

Using the analysis in the preceding sections, the displacement of the stapes and umbo can be compared for the two stimulation modalities, AC sound and BC vibration. For the comparison to be interesting, the absolute displacement of the ossicles at AC thresholds (Fig. 3, OD_{MAP}) and the relative displacement between the ossicles and the surrounding bone at BC thresholds [Fig. 6(b), OD_{RETFL}] are compared. This comparison for the stapes and umbo displacements are shown in Fig. 7. The overall displacements at threshold stimulation are similar, whereas AC or BC excitation is used. This similarity can be found in both umbo and stapes data. Although overall similar, a slight difference between the displacements with the two different stimulations is seen. This difference is similar for both the umbo and stapes data: at frequencies below 1.5 kHz the displacement of the ossicles is greater with AC stimulation than with BC stimulation, at frequencies between 1.5 and 3.1 kHz the displacements with AC and BC stimulation are close (the ossicle displacement with AC stimulation is still slightly greater), and at 3.1 kHz and above, the ossicles displacement with BC stimulation is greater than with AC stimulation.

In what follows the assumption is made that, with AC stimulation, the motion of the ossicles at threshold level is the lowest required motion of the middle ear ossicles to produce a hearing sensation. Consequently, if the motion of the ossicles at threshold stimulation is less than this, the hearing sensation cannot be caused by, or transmitted by, the middle ear ossicles. To rephrase this, if, at a specific frequency or frequency range, the motion of the ossicles at BC threshold stimulation is less than the ossicles motion at AC threshold stimulation ($OD_{MAP} > OD_{RETFL}$), the hearing sensation produced at BC thresholds is not primarily caused by the inertial motion of the middle ear ossicles or by sound in the external ear canal that is transmitted to the inner ear via the middle ear ossicles. One limitation with this approach is that it requires the BC stimulation to be generated by inertial forces. If the BC hearing is dominated by other phenomena, for example compression and expansion of the cochlear space, the stapes footplate is not necessarily related to the hearing threshold and the assumption fails. This limits the analysis primarily at higher frequencies. It should also be noted that the ossicle displacement data with BC stimulation are extracted from isolated temporal bone specimens that may not be a good model for compressional or some yet unaccounted for response with BC stimulation; this adds to the uncertainties in the analysis.

The results in Fig. 7 are rearranged to displacement at threshold with BC stimulation relative to AC stimulation (OD_{RETFL}/OD_{MAP}). Viewing the result in this way, Fig. 8(a) shows the relative stapes footplate displacement with a solid line and the relative umbo displacement as a dotted line. Displayed like this, the results found in Fig. 7 are even more evident: at frequencies below 1.5 kHz the ossicles threshold

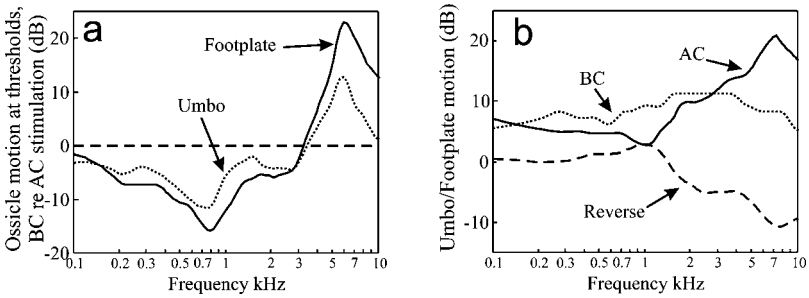


FIG. 8. (a) The ossicle motion at threshold stimulation displayed as the displacement with BC stimulation divided by the displacement with AC stimulation (OD_{RETFL}/OD_{MAP}). The relative stapes footplate motion is shown with a solid line, the relative umbo motion with a dotted line, and the 0 dB level is indicated with a dashed line. The data is taken from Fig. 7. (b) The umbo motion divided by the stapes footplate motion at hearing threshold stimulation. Results using AC stimulation are shown with a solid line and using BC stimulation with dotted line. Also included is the umbo divided with the footplate motion when the ossicles are driven from the vestibule in the cochlea (dashed line). More details about the latter are given in the text.

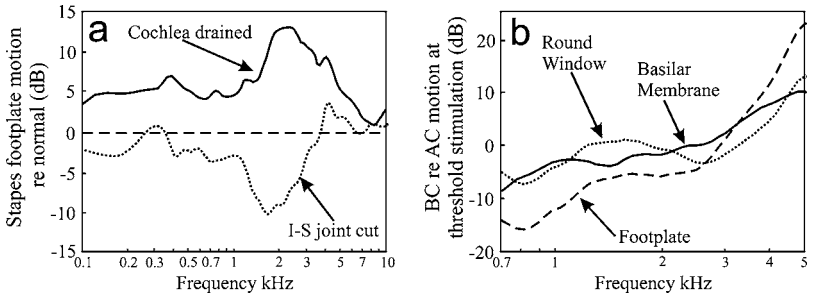


FIG. 9. (a) Alteration of the stapes footplate motion with BC stimulation after artificial lesion of the middle and inner ear. Changes of the motion after the cochlea was drained are shown with a solid line and after the incudo-stapedial joint was severed are shown with a dotted line. (b) The relative motion of the basilar membrane (solid line), round window (dotted line), and stapes footplate (dashed line) at BC threshold stimulation compared with AC threshold stimulation. The footplate data is the same as shown in Fig. 8(a); the frequency range is different.

vibration with BC stimulation is some 5 to 15 dB lower than with AC stimulation, at frequencies between 1.5 and 3.1 kHz the vibration of the ossicles is within 5 dB at AC or BC threshold stimulation, and at frequencies above 3.1 kHz, the stapes footplate vibration is up to 20 dB and the malleus umbo vibration is up to 10 dB greater with BC than AC stimulation at threshold levels.

An interpretation of the results is as follows. At frequencies below 1.5 kHz, the motion of the ossicles with BC stimulation (OD_{RETFL}) is below that of AC stimulation (OD_{MAP}) meaning that at these frequencies, the outer and middle ear is not the dominant contributor to BC sound perception. At the frequencies between approximately 1.5 and 3.1 kHz, the ossicles motion at AC and BC thresholds are within 5 dB, indicating that the middle ear ossicles can contribute to the BC perception in this frequency range. At frequencies above 3.1 kHz, the ossicles displacement at BC thresholds (OD_{RETFL}) are up to 20 dB greater (stapes footplate) or 10 dB greater (umbo) than at AC thresholds (OD_{MAP}). At these high frequencies, wave propagation at a wavelength that can cause alteration of the cochlear space is present (Stenfelt and Good, 2005b). As stated previously, with that type of BC excitation of the inner ear, the footplate vibration does not represent the input to the cochlea and the current analysis is no longer valid.

It is not obvious, given the many uncertainties in the analysis and extraction of the data, the significance of the differences found in Fig. 8(a). The limit for a change or difference to be considered a true change or difference is set to ±5 dB. The reason for the 5 dB significance level is primarily due to the results of basilar membrane motion in Sec. V B. However, as stated previously, the results from the present analysis should be interpreted with some caution.

Another way to present the data in Fig. 7 is to plot the relation between umbo and footplate motion for AC and BC stimulation. This is done in Fig. 8(b) where the AC umbo to footplate motion is shown with a solid line, and the BC umbo to footplate motion is shown with a dotted line. This plot represents the attenuation of motion between the umbo and footplate. Normally, when the ossicles are driven by a sound pressure in the ear canal (AC stimulation), the low-frequency motion of the stapes is slightly lower than the motion of the umbo. The ossicles have a resonance frequency around 1 kHz and above this frequency, the umbo to stapes motion increases at around 20 dB/decade [Fig. 8(b), solid line].

Although not exactly equal, the BC data in Fig. 8(b) (dotted line) shows a trend similar to the AC data (solid line) up to 1.7 kHz. Between this frequency and 4 kHz, the ratio between umbo and stapes motion with BC stimulation is almost constant; above 4 kHz the ratio between the umbo and stapes motion decreases. Since the BC data in Fig. 8(b) is derived from isolated temporal bone models, no sound radiated into the outer ear canal contributes to the result. This means that middle ear ossicles inertia, inner ear fluid inertia, and alteration of the cochlear space are the only components expected to cause the motion of the stapes and umbo. For comparison, the umbo-stapes ratio with the stimulation reversed is also plotted in Fig. 8(b) (dashed line). This data is taken from Stenfelt *et al.* (2003b) where the cochlea in the temporal bones were opened from the internal acusticus meatus, a piston coupled to a minitransducer was attached to the stapes footplate on the vestibular side, and the velocity of the umbo and stapes was measured with a laser Doppler vibrometer. The umbo-stapes ratio shown as the dashed line is expected when the ossicles are purely driven from the cochlea, whereas the solid line (AC) is the umbo-stapes ratio expected when the ossicles are driven purely from a sound pressure in the ear canal. Since the inertial forces of the ossicles also influence the umbo-stapes ratio, the BC curve (dotted line) is neither purely driven from the ear canal sound pressure (solid line) nor purely driven from the cochlea (dashed line). As stated previously, for frequencies up to 1.7 kHz, the trend of the umbo-stapes ratio with BC stimulation is similar to the AC stimulation whereas above this frequency, the trend of the ossicles ratio with BC stimulation is more similar to the driven in reverse curve (dashed line). Although no conclusive evidence can be found by investigating the umbo-stapes ratio using different stimulation modalities, some trends and indications are seen.

Next, the stapes footplate motion with BC stimulation is investigated after the incudo-stapedial (I-S) joint was artificially severed and after the cochlea was drained [Fig. 9(a)]. The results displayed are the stapes motion in temporal bone models with artificially lesions compared with the stapes motion in temporal bones with normal intact middle and inner ear (Stenfelt *et al.*, 2002). The dotted line in Fig. 9(a) shows the stapes motion change after the I-S joint was severed; the solid line shows the stapes motion change after the cochlea was drained.[2] These data were obtained during BC stimulation (the temporal bones were vibrated in the direction of the low-frequency in-and-out motion of the stapes). According to Fig. 9(a), the stapes motion was lowered 5 to 10 dB between the frequencies 1.2 and 2.7 kHz after the I-S joint was severed. After the cochlea was drained, an increase of stapes vibration of around 5 dB was seen up to frequencies of

1.5 kHz. Between 1.5 and 5.0 kHz, the increase is between 5 and 13 dB, and for frequencies above 5.0 kHz, the increase stays between 1 and 5 dB.

The loss in the mid-frequency range with a severed ossicular chain is attributed to the change in resonance frequency of the ossicular chain. With an intact ossicular chain the resonance frequency of the ossicles vibration with BC stimulation is close to 1.5 kHz, but with a disrupted chain (at the I-S joint) the resonance frequency of the stapes vibration is closer to 4 kHz. This resonance frequency change is primarily due to the mass reduction when the malleus and incus no longer is coupled to the stapes. Removing some of the inertial force by severing the ossicular chain results in a decrease of the stapes motion around 2 kHz. This data indicates that the inertia of the ossicles can, with an intact ossicular chain, contribute to the perception of BC sound around this frequency.

The increase of the stapes motion seen after the cochlea was drained is probably due to the removal of the cochlear fluid interaction. The total mass of the middle ear ossicles is 50 to 60 mg whereas the mass of the fluid within the cochlea and vestibule is around twice that, approximately 100 mg (Wever and Lawrence, 1954). The mass of the ossicles is centered to the head of the malleus and body of incus; the ossicles center of rotation goes through these masses. This results in a lower effective mass of the ossicular chain producing the inertial force during BC stimulation than the total sum of the three ossicles masses. However, it is scarcely likely that, during BC stimulation, the total mass of the inner ear fluid contributes to the inertial forces of the inner ear fluid. Although the effective mass of the inner ear fluid and corresponding inertial force is less than the total mass of the inner ear fluid, the inertial force of the inner ear fluid is probably greater than the inertial force of the ossicles. The increase of stapes motion with removal of the cochlear fluid seen in Fig. 9(a) is therefore believed to rather be the removal of fluid motion than the impedance reduction for the stapes motion at the stapes-vestibuli interface.

B. Basilar membrane and round window motion

A last analysis is made with measured basilar membrane and round window motion using both AC and BC stimulation (Stenfelt *et al.*, 2003a; Stenfelt *et al.*, 2004). If the assumptions and calculations here are correct, one can expect equal vibration magnitude of the basilar membrane at threshold stimulation independent if the excitation is by AC or BC. The basilar membrane vibration at BC thresholds compared with the vibration at AC thresholds is displayed in Fig. 9(b) (solid line). This comparison is computed similar as the stapes footplate and malleus umbo data previously. Since the BC data of basilar membrane motion was uncertain at low and high frequencies, only results for frequencies between 0.7 and 5 kHz are included here. For comparison, the stapes footplate relative displacement (dashed line) and the round window relative volume displacement (dotted line) at thresholds are also included in Fig. 9(b). The footplate data are the same as displayed in Fig. 8(a); the frequency range of the plots differ. According to the calculations deriving the results presented in this study, the basilar membrane vibration differ less than 5 dB between BC and AC threshold stimulation in the frequency range 0.8 to 3.5 kHz. Beyond these frequencies, the difference is greater.

The basilar membrane motion is calculated using almost the same data, assumptions, and simplifications as for deriving the stapes footplate and the malleus umbo motion. As a consequence, the results of the basilar membrane motion can be used as a verification of the results presented previously. As stated above, the basilar membrane motion should be equal at AC threshold and BC threshold stimulation. Within the frequency range 0.8 to 3.5 kHz the basilar membrane motion differ less than 5 dB and within the frequency range 0.7 kHz and 5 kHz it differ less than 10 dB between AC and BC threshold stimulation using the calculations presented here. Therefore, in the comparisons, changes and differences less than 5 dB is not considered as significant and differences in the 5 to 10 dB range should be interpreted with care in the frequency range 0.7 to 5 kHz; no validation data is applicable at frequencies outside this range. Since individual AC and BC hearing thresholds may differ more than 5 dB among normal hearing subjects, the basilar membrane data provides a strong indication of the validity of the current analysis, at least in the frequency range 0.8 to 3.5 kHz.

One interesting finding seen in Fig. 9(b) is that the relative round window volume displacement at threshold stimulation is similar to the relative basilar membrane motion. This could implicate that the round window volume displacement, correctly measured, is a good estimate of the hearing excitation, whereas the stimulation is by AC or BC. It should be noted that only one or a couple positions measured on the round window membrane does not give a good estimate of the round window volume displacement; the round window membrane motion depends on the stimulation itself (Stenfelt *et al.*, 2004). It should also be remembered that the basilar membrane data were obtained in cadaver temporal bone specimens; no active mechanism enhancing the basilar membrane motion was present. Although the active mechanism is expected to influence AC and BC stimulation of the basilar membrane similarly, no direct measurement of the basilar membrane motion in the human with AC and BC stimulation has been conducted. As explained in the introduction, the anticipated similarity of the basilar membrane motion with AC and BC stimulation is based on the ability to produce distortion product otoacoustic emissions with both stimulations as well as cancellation of pure tones and similarities of resonance frequencies on the basilar membrane with the two stimulation modalities.

VI. CLINICAL FINDINGS

It would be interesting to compare the results from this analysis to reported clinical findings. Therefore, the computed estimates are compared with findings of superior semicircular dehiscence, mass-loading of the ossicles, otosclerosis of the stapes footplate, and removal of middle ear ossicles.

A. Superior semicircular dehiscence

In rare occasions, an opening can occur between the cranial space and the superior semicircular canal enabling a fluid pathway between the cerebrospinal fluid and fluid in the vestibule. The audiological findings in these patients are, beside sound induced vertigo, worse low-frequency AC thresholds and improved low-frequency BC thresholds (Minor *et al.*, 2003; Mikulec *et al.*, 2004). The worse AC thresholds can be explained by the shunt effect of the dehiscence on the vestibular side reducing the fluid flow from the scala vestibuli to scala tympani and stimulation of the basilar membrane. The sensitivity increase at low frequencies with BC stimulation can be explained by the impedance decrease for fluid flow on the scala vestibule side (paralleling the oval window) that would increase the effect of the fluid inertia. Another explanation for the improved low-frequency BC sensitivity could be a pressure transmission from the CSF to the cochlea that is improved due to the patent fluid connection. Regardless whether the first or second explanation is correct, the improvement of low-frequency hearing thresholds with BC stimulation indicates that hearing of BC sound is not dominated by sound in the outer ear canal or inertia of the middle ear ossicles at these frequencies. However, a dehiscence does not normally alter the hearing thresholds at the frequency range where middle ear inertia is identified as a potential contributor to BC sound (approximately 1.5 to 3 kHz).

B. Mass-loading of the ossicles

Several investigators have increased the mass of the ossicles to manipulate the BC sensitivity (Bárány, 1938; Kirikae, 1959; Brinkman *et al.*, 1965). An increase of the ossicles mass lowers the resonance frequency of the mechanical system and improves the relative low-frequency (below the resonance frequency of the normal ossicular chain) vibration between the ossicles and surrounding bone during BC stimulation. Huizing (1960) used water and mercury on the eardrum to increase the mass of the middle ear ossicles and alter the effect from inertia on BC thresholds. He reported that the BC hearing thresholds below 1 kHz improved with increasing mass for masses up to approximately 250 mg. A further increase of the mass over 250 mg did not change the BC hearing thresholds. In his study, an increase of 100 mg produced a threshold shift of around 10 dB for frequencies below 1 kHz. That result can be compared with the results in Stenfelt *et al.* (2002), where 100 mg was added to the malleus and produced around 20 dB greater relative vibration levels of the stapes for frequencies below 0.7 kHz. Consequently, with BC stimulation an improvement in stapes vibration level of 20 dB corresponds to a threshold shift of 10 dB at low frequencies. This can be interpreted as follows: the vibration of the ossicles with BC excitation due to inertia at frequencies below 1 kHz causes a hearing stimulation that is around 10 dB lower than other phenomena contributing to sound perception via BC stimulation. This also agrees well with the results presented in Fig. 7(a) where the stapes vibration at BC threshold is 5 to 15 dB below the vibration level of the stapes at AC thresholds for frequencies below 1 kHz. Once again, no evidence supporting or refuting the BC middle ear ossicles inertial effect between 1.5 and 3 kHz can be found from these investigations.

C. Otosclerosis of the stapes footplate

Results of the BC thresholds in patients with otosclerosis of the stapes footplate have been reported extensively in the literature. Carhart (1962) drew attention to the typical BC threshold alteration caused by stapes otosclerosis: a loss of approximately 20 dB at 2 kHz, with lesser losses at frequencies below and above this frequency, now known as the Carhart notch. One explanation for the typical BC threshold change with otosclerosis was given by Tonndorf (1966) who suggest that this loss is the lack of excitation from the inertial component of the ossicles. Although this explanation has some support from the result in Stenfelt *et al.* (2002), the threshold elevation can originate from the impedance change at the oval window seen from the inside of the cochlea. For example, fluid flow caused either by fluid inertia or by alterations of the cochlear space can be impeded by increased impedance at the oval window.

The elevation of BC thresholds with otosclerosis is found at, and also slightly below and above, the frequency range identified as the area of possible ossicles inertia contribution to BC sound. The otosclerosis removes effectively any contribution from the ossicles and also from the outer ear with BC stimulation. However, since the otosclerosis also blocks the fluid flow at the oval window, it is difficult to separate the middle ear and inner ear parts of the contribution to the BC sound. Moreover, Everberg (1968) reported that patients without oval windows due to malformations had BC thresholds close to normal. This suggests that even though middle ear ossicle inertia contributes to the hearing of BC sound in the frequency range 1.5 to 3 kHz, other phenomena is nearly as effective producing a hearing sensation with BC stimulation.

D. Middle ear discontinuity

As stated previously, resection of the incudo-stapedial joint increases the resonance frequency for vibration of the stapes when driven by inertial effects during BC stimulation. This can be understood by the decrease of mass (malleus and incus) while maintaining the compliance attaching the stapes (annular ligament). This is not the whole story since some ligaments of the malleus and incus together with the TM influences the stiffness of the middle ear ossicles as well; these are removed from the stapes when severing the I-S joint. However, seen from the stapes, the reduced mass-effect is greater than the alteration of the stiffness. Severing the I-S joint increased the resonance frequency of the stapes with slightly more than one octave while driven by BC. For simplicity, let us assume that the increase is precise one octave. Below the resonance frequency, the relative motion between the stapes and cochlear promontory rises at 12 dB/octave [Fig. 6(a)]. This means that, by severing the I-S joint and causing a resonance frequency shift of one octave, the stimulation at the oval window is reduced by 12 dB for frequencies below the original resonance frequency (approximately

1.5 kHz), between the original and the new resonance frequency the reduction decreases (improves) from 12 dB to 0 dB as a function of frequency. Above the new resonance frequency no alteration of the stimulation is found.

It has been reported that discontinuity of the middle ear ossicles, e.g., lesion of the incudo-stapedial joint, has a small effect on the BC thresholds at frequencies below 2 kHz (Møller, 2000). This agrees well with temporal bone experiments displayed in Fig. 9(a); the change in stapes motion after I-S joint resection is within 5 dB up to 1 kHz where it decreases to below −10 dB at 1.5 kHz. Above this frequency the decrease of the stapes motion recovers and is close to zero for frequencies of 3.5 kHz and above. An interpretation of this result from the model above is as follows: at frequencies below 1 kHz, the middle ear does not contribute to perception of BC sound. At frequencies between 1.5 and 3 kHz the middle ear inertia contributes to the perception of BC sound, and above 3 kHz no conclusions can be made since no alteration is expected from the model. Hence, results with middle ear discontinuity indicate that the inertial forces of the ossicular chain do contribute to the perception of BC sound in the frequency range close to the resonance frequency of the ossicles. It should be noted that the alteration of BC thresholds depend on the position and type of middle ear discontinuity; e.g., if the discontinuity is between the malleus and umbo, less change of BC hearing thresholds is expected.

VII. HIGH-FREQUENCY RESULTS

From Fig. 8 it is clear that, according to the present analysis and at frequencies above 5 kHz, the stapes footplate and umbo vibrates 10 to 20 dB more at BC thresholds than at AC thresholds. Since, according to the previously stated hypothesis, the vibration of the stapes at AC thresholds is considered the lowest required vibration level to cause a hearing sensation, the high-frequency stapes motion at BC thresholds produces cochlear stimulation that are some 20 dB above thresholds. Consequently, the assumptions made are incorrect at these high frequencies. The reason is most probably that for BC stimulation at these high frequencies, the stapes footplate vibration does not reflect the fluid flow in the cochlea causing the basilar membrane vibration.

One explanation is that there are compliant fluid pathways in the cochlea paralleling the oval window allowing fluid flow in the cochlea without involving the oval window and attached stapes. Another plausible explanation is that the dominant excitation mode at these high frequencies is expansion and compression of the cochlear space caused by wave propagation in the skull bone. In such cases, the fluid flow at the oval window is expected to be greater than the fluid flow over the basilar membrane causing a hearing stimulation. A third explanation is that with BC stimulation at high frequencies, the stimulation of the basilar membrane is caused locally, e.g., by the inertial forces of the osseus spiral lamina (Stenfelt *et al.*, 2003a), not involving the fluid flow at the oval window and vibration of the stapes. A fourth explanation for the result is that isolated temporal bone specimen is not a good model of BC stimulation at the high frequencies; threshold data are obtained in live humans while ossicle motions are estimated from temporal bone models and the BC stimulation of the cochlea and middle ear differ between the two.

When examining Fig. 9(b) and comparing threshold motion of the stapes footplate, basilar membrane, and round window, the third explanation appears less plausible. The relative round window and basilar membrane vibration level is similar at threshold stimulation for both stimulation modalities (AC and BC), indicating that the round window fluid flow reflects the stimulation of the basilar membrane. Figure 9(b) only shows results up to frequencies of 5 kHz, and at frequencies above 4 kHz the results begins to deviate, once again indicating that the present analysis may not be valid at those high frequencies. However, according to clinical findings reported in the literature (Martin *et al.*, 2002; Groen and Hoogland, 1958), obstructing the round window affects BC hearing thresholds more than obstructing the oval window. This is another indication that a fluid flow at the round window is essential for hearing by BC and it reflects the stimulation of the basilar membrane when the stimulation is by BC.

VIII. CONCLUSIONS

The analysis presented is based on human behavioral threshold data with AC and BC stimulation together with physiological and mechanical vibration data of the middle ear ossicles, round window, and basilar membrane. With the assumption that the stapes footplate vibration at AC threshold stimulation represents the lowest cochlear excitation required to elicit a neural response and hearing perception, the contribution to BC hearing perception from the middle ear ossicles was investigated. According to that approach, the vibration of the middle ear ossicles with BC stimulation is approximately 10 dB too low to contribute to the BC hearing perception at frequencies below 1.5 kHz. For frequencies above 4 kHz, the assumptions made for this analysis is no longer valid, and no conclusive evidence for the importance of the middle ear ossicles to BC hearing can be given. In the frequency range between 1.5 and 3.1 kHz, the stapes footplate and umbo vibrations were similar at AC and BC threshold stimulation. This frequency range includes the resonance frequencies for the middle ear ossicles vibration with BC stimulation. Although, given the uncertainties of using isolated temporal bone specimens and cadaver heads for the results with BC stimulation, it is proposed that in the frequency range between 1.5 and 3.1 kHz the inertial effect of the middle ear ossicles contributes to hearing by BC stimulation. However, according to clinical findings, as long as there is an opening allowing fluid flow in the scala vestibule side of the cochlea (similar to the oval window), effects within the cochlea produces almost similar excitation of the basilar membrane as the inertial effects of the middle ear ossicles in the frequency range 1.5 to 3.1 kHz.

ACKNOWLEDGMENT

This work was supported by a grant from the Swedish Research Council (Grant No. 621-2002-5624).

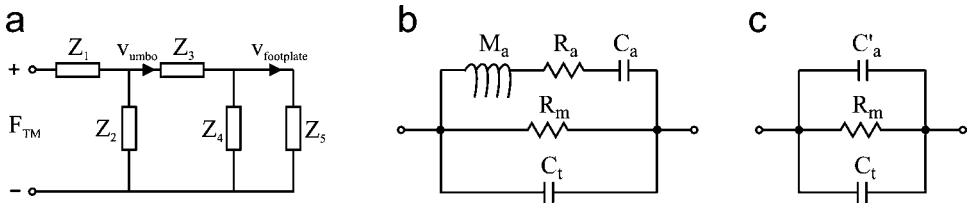


FIG. 10. (a) The impedance model of the middle ear according to Zwislocki (1962). The physical interpretation of the five impedance branches is given in the text. (b) The lumped parameters of impedance Z_1 for the intact middle ear. (c) The lumped parameters of impedance Z_1 for the artificially altered middle ear space.

APPENDIX: EFFECT OF THE MIDDLE EAR CAVITY

Zwislocki (1962) describes the sound (vibration) transmission through the middle ear with a lumped impedance network according to Fig. 10(a). According to this model the input at the left-hand side is the sound pressure in front of the eardrum multiplied with the eardrum area, depicted as F_{TM} ($F_{\mathrm{TM}}=P_{\mathrm{TM}}\cdot A_{\mathrm{TM}}$). The general impedance network and the five impedance branches in Fig. 10(a) were primarily derived considering the anatomy and using acoustic impedance measurements. All five impedances (Z_1–Z_5) can be related to physical structures. The effects of the middle ear cavity are described by Z_1 while Z_2 represents the part of the eardrum that moves differently from the malleus. The part of the eardrum together with the malleus and incus that vibrates nearly as a unit is represented by Z_3 and Z_4 describes primarily the stiffness of the incudo-stapedial joint. Last, Z_5 is the impedance of the stapes and the annular ligament attaching it to the oval window, the input impedance of the cochlea, and the RW membrane. A more detailed description of the impedance model and corresponding parameters is found in Zwislocki (1962).

The details of the impedance branch representing the middle ear cavity, Z_1, are shown for the intact middle ear in Fig. 10(b); this is taken from Zwislocki (1962). The air compliance within the tympanic cavity is modeled by C_t and R_m is the absorption in the tympanic cavity and Eustatian tube. The branch comprising M_a, R_a, and C_a is the resonance from the antrum and mastoid air cells where M_a is the mass behavior of the narrow passage between the tympanic cavity and antrum, R_a models the losses of the same passage, whereas C_a describes the compliance of the air in the antrum and mastoid air cells.

The measurements of ossicle displacement with AC stimulation were conducted in cored temporal bone specimens where part of the mastoid air cells and occasionally part of the antrum are missing (Hato *et al.*, 2003). In these specimens, a hole was drilled into the antrum and tympanic cavity enlarging the small passage between the antrum and tympanic cavity. Before the measurements are executed the isolated temporal bone specimen is enclosed with a latex cover and the opening into the middle ear cavity is sealed with a glass plate. This means that in these manipulated specimens, the resulting middle ear space is best described as a single large air volume and the corresponding impedance model is the one depicted in Fig. 10(c). In this model, R_m and C_t are the losses and compliance of the tympanic cavity as in Fig. 10(b). The difference is that the rest is modeled by C'_a which is basically the rest of the air enclosed in the altered middle ear space, i.e., C'_a+C_t represent the total air volume enclosed in the altered middle ear cavity. The normal tympanic cavity is in the order of 0.5 cc and the artificially opened space is 1.0 to 1.5 cc. This means that C'_a represents an air volume two to three times that of C_t. For the calculations estimating the difference of malleus umbo and stapes footplate displacement between the intact and artificially manipulated middle ear, the relation $C'_a=2C_t$ was used (corresponds to a total middle ear space volume of approximately 1.5 cc).

In the model in Fig. 10(a) the velocities of the umbo (v_{umbo}) and footplate ($v_{\mathrm{footplate}}$) are indicated with arrows. The displacements of the ossicles can be computed by integrating these velocities. Using the parameters given by Zwislocki (1962) and using the model in Fig. 10(b) for an intact middle ear and the model in Fig. 10(c) for an altered middle ear cavity, the change of the umbo and footplate displacements can be estimated. The estimated ratio between the ossicle displacement with an intact middle ear and the artificially altered middle ear space is shown in Fig. 2(b).

[1]The unpublished data were obtained in two patients equipped with a titanium implant anchoring bone conduction hearing aids. This enabled direct BC stimulation of the skull bone behind the pinnae. These two patients had enlarged ear canals and had their eardrums, malleus and incus removed. This allowed for the LDV beam to reach the cochlear promontory and the displacement of the cochlear promontory could be measured when the BC stimulation was applied at the mastoid, similar to the setup in Stenfelt and Good (2005b).

[2]These data with the drained cochlea were never shown in Stenfelt *et al.* (2002). The data shown in that study were with the cochlea drained and the I-S joint severed. The data presented here in Fig. 9(a) were obtained during that same study, using the same setup and measurement method as all other data presented in that study.

Bárány, E. (**1938**). "A contribution to the physiology of bone conduction," Acta Oto-Laryngol., Suppl. **26**, 1–223.

Brinkman, W., Marres, E., and Tolk, J. (**1965**). "The mechanism of bone conduction," Acta Oto-Laryngol. **59**, 109–115.

Carhart, R. (**1962**). "Effect of stapes fixation on bone conduction response," in *Otosclerosis*, edited by H. Schuknecht (Little, Brown and Company, Boston), pp. 175–197.

Everberg, G. (**1968**). "Congenital absence of the oval window," Acta Oto-Laryngol. **66**, 320–332.

Flottorp, G., and Solberg, S. (**1976**). "Mechanical impedance of human headbones (forehead and mastoid portion of the temporal bone) measured under ISO/IEC conditions," J. Acoust. Soc. Am. **59**, 899–906.

Groen, J. J., and Hoogland, G. A. (**1958**). "Bone conduction and otosclerosis of the round window," Acta Oto-Laryngol. **49**, 206–212.

Håkansson, B., Carlsson, P., and Tjellström, A. (**1986**). "The mechanical point impedance of the human head, with and without skin penetration," J. Acoust. Soc. Am. **80**, 1065–1075.

Hato, N., Stenfelt, S., and Goode, R. L. (**2003**). "Three-dimensional stapes footplate motion in human temporal bones," Audiol. Neuro-Otol. **8**, 140–152.

Huizing, E. H. (**1960**). "Bone conduction—The influence of the middle ear," Acta Oto-Laryngol., Suppl. **155**, 1–99.

ISO 389-3 (**1994**). "Acoustics—Reference zero for the calibration of audiometric equipment. Part 3 Reference equivalent thresholds force levels for pure tones and bone vibrators."

ISO 226 (**2003**). "Acoustics—Normal equal-loudness-level contours."
Khanna, S. M., Tonndorf, J., and Queller, J. (**1976**). "Mechanical parameters of hearing by bone conduction," J. Acoust. Soc. Am. **60**, 139–154.
Killion, M. C. (**1978**). "Revised estimate of minimum audible pressure: Where is the "missing 6 dB"?," J. Acoust. Soc. Am. **63**, 1501–1508.
Kirikae, I. (**1959**). "An experimental study on the fundamental mechanism of bone conduction," Acta Oto-Laryngol., Suppl. **145**, 1–111.
Martin, C., Tringali, S., Bertholon, P., Pouget, J.-F., and Prades, J.-M. (**2002**). "Isolated congenital round window absence," Ann. Otol. Rhinol. Laryngol. **111**, 799–801.
Mikulec, A. A., McKenna, M. J., Rosowski, J. J., Herrmann, B. S., Rauch, S. D., Curtin, H. D., and Merchant, S. N. (**2004**). "Superior semicircular canal dehiscence presenting as conductive hearing loss without vertigo," Otol. Neurotol. **25**, 121–129.
Minor, L. B., Carey, J. P., Cremer, P. D., Lustig, L. R., Streubel, S. O., and Ruckenstein, M. J. (**2003**). "Dehiscence of bone overlying the superior canal as a cause of apparent conductive hearing loss," Otol. Neurotol. **24**, 270–278.
Møller, A. R. (**2000**). *Hearing. Its Physiology and Pathophysiology* (Academic, San Diego), pp. 379–394.
Purcell, D., Kunov, H., Madsen, P., and Cleghorn, W. (**1998**). "Distortion product otoacoustic emissions stimulated through bone conduction," Ear Hear. **19**, 362–370.
Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.
Stenfelt, S., and Good, R. L. (**2005a**). "Bone-conducted sound: Physiological and clinical aspects," Audiol. Neurotol. **26**, 1245–1261.
Stenfelt, S., and Good, R. L. (**2005b**). "Transmission properties of bone conducted sound: Measurements in cadaver heads," J. Acoust. Soc. Am. **118**, 2373–2391.
Stenfelt, S., and Håkansson, B. (**1999**). "Sensitivity to bone-conducted sound: Excitation of the mastoid vs the teeth," Scand. Audiol. **28**, 190–198.
Stenfelt, S., Hato, N., and Goode, R. L. (**2002**). "Factors contributing to bone conduction: The middle ear," J. Acoust. Soc. Am. **111**, 947–959.
Stenfelt, S., Hato, N., and Goode, R. L. (**2004**). "Fluid volume displacement at the oval and round windows with air and bone conduction stimulation," J. Acoust. Soc. Am. **115**, 797–812.
Stenfelt, S., Puria, S., Hato, N., and Goode, R. L. (**2003a**). "Basilar membrane and osseous spiral lamina motion in human cadavers with air and bone conduction stimuli," Hear. Res. **181**, 131–143.
Stenfelt, S., Wild, T., Hato, N., and Goode, R. L. (**2003b**). "Factors contributing to bone conduction: The outer ear," J. Acoust. Soc. Am. **113**, 902–912.
Tonndorf, J. (**1966**). "Bone conduction. Studies in experimental animals," Acta Oto-Laryngol., Suppl. **213**, 1–132.
Wever, E. G., and Lawrence, M. (**1954**). *Physiological Acoustics* (Princeton University Press, Princeton, NJ), pp. 1–454.
von Békésy, G. (**1932**). "Zur theorie des hörens bei der schallaufnahme durch knochenleitung," Ann. Phys. **13**, 111–136.
Zwislocki, J. (**1962**). "Analysis of the middle-ear finction. Part 1: Input impedance," J. Acoust. Soc. Am. **34**, 1514–1523.

A geometrically nonlinear finite-element model of the cat eardrum

Hanif M. Ladak[a)]
Department of Medical Biophysics, The University of Western Ontario, London, Canada N6A 5C1;
Department of Electrical and Computer Engineering, The University of Western Ontario, London, Canada N6A 5B9 and
Department of Otolaryngology, The University of Western Ontario, London, Canada N6A 5A5;
Imaging Research Labs, Robarts Research Institute, London, Canada N6A 5K8

W. Robert J. Funnell
Departments of BioMedical Engineering and Otolaryngology, McGill University, Montréal, Canada H3A 2B4

Willem F. Decraemer and Joris J. J. Dirckx
Laboratory of BioMedical Physics, University of Antwerp, Antwerpen, Belgium B-2020

(Received 10 September 2005; revised 22 February 2006; accepted 25 February 2006)

Current finite-element (FE) models of the eardrum are limited to low pressures because of the assumption of linearity. Our objective is to investigate the effects of geometric nonlinearity in FE models of the cat eardrum with an approximately immobile malleus for pressures up to ±2.2 kPa, which are within the range of pressures used in clinical tympanometry. Displacements computed with nonlinear models increased less than in proportion to applied pressure, similar to what is seen in measured data. In both simulations and experiments, there is a shift inferiorly in the location of maximum displacement in response to increasingly negative middle-ear pressures. Displacement patterns computed for small pressures and for large positive pressures differed from measured patterns in the position of the maximum pars-tensa displacement. Increasing the thickness of the postero-superior pars tensa in the models shifted the location of the computed maximum toward the measured location. The largest computed pars-tensa strains were mostly less than 2%, implying that a linearized material model is a reasonable approximation. Geometric nonlinearity must be considered when simulating eardrum response to high pressures because purely linear models cannot take into account the effects of changing geometry. At higher pressures, material nonlinearity may become more important. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188370]

PACS number(s): 43.64.Ha [BLM] Pages: 2859–2868

I. INTRODUCTION

Existing finite-element (FE) models of the eardrum (e.g., Funnell and Decraemer, 1996; Beer *et al.*, 1999; Bornitz *et al.*, 1999; Eiber 1999; Ferris and Prendergast, 2000; Koike *et al.*, 2002; Gan *et al.*, 2004) are limited to low pressure levels because of the simplifying modeling assumption that the relationship between applied pressure and the resulting displacements is linear (i.e., that an increase in pressure results in a proportional increase in displacement). However, modeling the response to high pressures would be useful in understanding the mechanics of impedance tympanometry in which the static pressures that are used are so high that the responses become nonlinear.

As one step toward our goal of developing and evaluating an FE model of the eardrum that is valid for high pressures, we measured the shape and displacement patterns of the cat eardrum in response to static pressures (Ladak *et al.*, 2004) by using phase-shift shadow-moiré topography, a noncontacting optical technique that was originally adapted for measurements on the eardrum by Dirckx *et al.* (1988). Our experiments involved cyclically pressurizing the middle ear up to ±2.2 kPa after immobilizing the malleus. The immobile-malleus condition allows us to investigate the behavior of the eardrum in isolation from any possible nonlinearities of the cochlea and middle-ear structures. The immobile-malleus data indicate that the eardrum response itself is nonlinear at the pressures used in the study. In other words, eardrum displacements do not increase in proportion to the applied pressure; specifically, the displacements grow less than in proportion to applied pressure, like that of a stiffening structure.

In order to model the measured data, nonlinearities must be added to current models. Two basic types of nonlinearities must be considered in modeling the eardrum or any other structure: material nonlinearity and geometric nonlinearity. Material nonlinearity manifests itself as a nonlinear relationship between stress and strain. Experimental measurements on the pars tensa indicate that the relationship between stress and strain is indeed nonlinear, with the material becoming stiffer as it undergoes large stretches (Decraemer *et al.*, 1980). For small changes in strain from an equilibrium state, however, materials can often be approximated as linear.

[a)]Electronic mail: hladak@uwo.ca

Geometric nonlinearity is related to changes in the shape of the structure while the structure is being loaded. Large displacements can significantly alter the shape and therefore the stiffness of a structure as well as the line of action of applied loads. For small displacements, geometric nonlinearity can be neglected. However, for thin-shell structures such as the eardrum, geometric nonlinearity starts to become important when the displacements are as large as the thickness of the structure (Timoshenko and Woinowsky-Krieger, 1959; Fung, 1965). At the highest pressures in our experimental study (±2.2 kPa), eardrum displacements measured with an immobile malleus are many times larger than the thickness of the pars tensa (Ladak *et al.*, 2004), implying that geometric nonlinearity must be included in eardrum models when simulating the response to such high pressures.

For soft tissues, both geometric and material nonlinearities often occur together. We hypothesize that for our experiments on the eardrum with an immobile malleus and for pressures up to ±2.2 kPa, geometric nonlinearity must be taken into account because the displacements are larger than the thickness of the pars tensa. However, the relative importance of material nonlinearity is not clear. Although the measured displacements are large relative to the thickness of the pars tensa, they are small relative to the overall dimensions of the eardrum, and the resulting strains may also be small. If the strains are indeed small, then linearizing the relationship between stress and strain under these specific experimental conditions may be a reasonable approximation.

The objective of this work is to investigate the effects of geometric nonlinearity alone on the behavior of the eardrum with an immobile malleus measured for pressure levels up to ±2.2 kPa, by incorporating such nonlinearity into our FE models of the cat eardrum and by comparing computed displacements with measured ones. To investigate the eardrum in isolation from the loading effects of the middle-ear structures and cochlea, we confined our simulations to the case of an immobile malleus as in the experimental study.

II. FINITE-ELEMENT MODELS

A. Mesh geometry

The response of the eardrum depends significantly on its shape (Funnell and Laszlo, 1978; Funnell and Decraemer, 1996), which can vary from subject to subject; hence, individualized FE meshes were constructed from *resting-shape* data measured in our previous experimental study (Ladak *et al.*, 2004); the resting shape is taken to be the shape measured at the beginning of an experiment, with no pressure applied to the eardrum. Meshes were defined for three cats from the previous study (MY1216L, MY1630L, and MY0923R) in which experiments were done with an immobile malleus. A sample image of the resting shape for cat MY1612L is shown in Fig. 1. The grey levels in the image vary from black (points furthest from reader) to white (points closest to reader). A vertical profile through the pars tensa just inferior to the manubrium is shown to the left of the image, and a horizontal profile through the posterior pars tensa is shown below the image. The locations of the profiles in relation to the original image are shown by the dashed lines. The figure also shows a lateral view of the corresponding eardrum mesh along with vertical and horizontal profiles through the mesh corresponding to the profiles shown through the moiré shape image. The procedure for constructing an FE mesh from a moiré shape image has previously been described in detail (Funnell and Decraemer, 1996). The mesh shown has a *nominal resolution* of 25 elements/diameter (Funnell, 1983). Each element of the mesh represents an S3R general-purpose shell element of the ABAQUS commercial FE software package (Hibbit, Karlsson, and Sorensen Inc., Pawtucket, RI). This element is accurate for both thin and thick shells. No elements are generated for the manubrium, as it is assumed to be completely immobile along its entire length in order to model our previously reported immobile-malleus data. In some areas, the mesh only appears to overlap the bony part of the ear canal; however, here the eardrum is actually concealed by overhanging bone. In such cases, the location of the periphery of the eardrum was estimated by extrapolation as described by Funnell and Decraemer (1996).

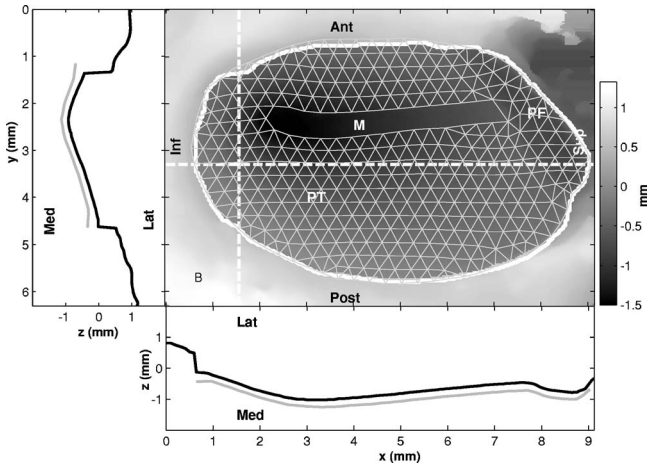


FIG. 1. Grey-level image of the resting shape of the eardrum of cat MY1216L with corresponding FE mesh superimposed. The thick white line encloses the visible portion of the eardrum. The grey levels in the image vary from black (point furthest away from reader) to white (point closest to reader), whereas the mesh is shown in a uniform light grey. The regions of the mesh corresponding to the pars tensa (PT), pars flaccida (PF), and manubrium (M) are labeled. The anterior (Ant), posterior (Post), inferior (Inf), and superior (Sup) directions are approximate because the cat eardrum is actually tilted, and therefore the pars flaccida is both posterior and superior (dorsal) and the manubrium runs in the antero-inferior to postero-superior direction. The bottom panel shows horizontal profiles through the image (black profile) and through the mesh (grey), taken just posterior to the manubrium as indicated by the horizontal dashed line in the image. The left panel shows vertical profiles through the image (black profile) and through the mesh (grey), taken through the pars tensa just inferior to the manubrium as indicated by the vertical dashed line in the image. Mesh profiles are offset so as not to overlap with image profiles.

B. Mechanical properties

As in previous simulations (e.g., Funnell and Decraemer, 1996), both the pars tensa and the pars flaccida are assumed to be elastic, that is, the eardrum is assumed to return to its resting shape upon the removal of applied pressure. Because the purpose of this study was to investigate the effects of geometric nonlinearity, the material of the eardrum

was assumed here to be linear, as in our previous simulations. That is, the relationship between stress and strain is assumed to be linear. Furthermore, again as in previous reports, the pars tensa and the pars flaccida are assumed to be isotropic and homogeneous. The pars tensa is characterized by a single Young's modulus or stiffness of 20 MPa, a thickness of 40 μm and a Poisson's ratio of 0.3; the pars flaccida is assumed to have a Young's modulus of 1 MPa, a thickness of 80 μm and a Poisson's ratio of 0.3 (Funnell and Decraemer, 1996).

The periphery of the eardrum is assumed to be fully clamped to the ear canal, so that nodes on the boundary of the model do not undergo translation or rotation. The boundary of the manubrium is also assumed to be fully clamped in order to simulate an immobile manubrium.

C. Solution procedure

For each model, static-displacement patterns were computed in response to uniform static middle-ear pressures from 0 to +2.2 kPa in steps of 0.1 kPa and from 0 to −2.2 kPa in steps of −0.1 kPa. The static pressure is applied to the medial (middle-ear) surface of the tympanic membrane, such that positive pressures cause laterally directed motions (into the external ear), while negative pressures produced medially directed motions (into the middle ear).

The FE method results in a system of equations of the form

$$\mathbf{Ku} = \mathbf{f}, \tag{1}$$

where $\mathbf{K}$ is a global stiffness matrix that is assembled from the known geometry and material properties of each element, $\mathbf{f}$ is a vector of applied nodal forces that is assembled from the known applied pressure and geometry, and $\mathbf{u}$ is an unknown vector of nodal displacements and rotations that needs to be computed. In *linear* problems, both $\mathbf{K}$ and $\mathbf{f}$ are constants and approximated as being independent of $\mathbf{u}$; thus, $\mathbf{u}$ can easily be computed as $\mathbf{u} = \mathbf{K}^{-1}\mathbf{f}$.

In geometrically nonlinear problems, however, the displacements and rotations are large, and the dependence of $\mathbf{K}$ and $\mathbf{f}$ on $\mathbf{u}$ cannot be ignored: Changes in the shape of the structure as it is being loaded can significantly alter $\mathbf{K}$ and $\mathbf{f}$. Because $\mathbf{K}$ and $\mathbf{f}$ depend on $\mathbf{u}$, simple matrix inversion cannot be used to compute the unknown vector $\mathbf{u}$. Instead, in ABAQUS, an incremental-iterative solution procedure is used to compute $\mathbf{u}$ at each pressure step. The nonlinear solution procedure used in this work is well described in the literature (e.g., Bathe, 1982). Basically, each pressure step of size ±0.1 kPa is automatically broken into smaller increments, the sum of which equals the applied step size. Applying the total pressure step in small increments allows displacements to be accurately computed as the pressure is increased to its final level. At the beginning of each increment, an initial stiffness matrix $\mathbf{K}_0$ is computed from the current geometry of the mesh and its mechanical properties. Based on this stiffness, a linear solution is computed and the geometry of the mesh is updated. Generally, force equilibrium will not be satisfied within the deformed mesh, and the residual vector defined by $\mathbf{r} = \mathbf{f} - \mathbf{K}^{-1}\mathbf{u}$ will be nonzero. ABAQUS uses Newton's method to find a deformed configuration that satisfies force equilibrium and results in an approximately zero residual vector. Note that we calculate the solution in steps of 0.1 kPa instead of the larger step sizes used in the experiments (see Sec. III A). The smaller step size used with the numerical solution procedure ensures convergence to an equilibrium solution at each step.

III. RESULTS

A. Full-field displacement and strain patterns

Displacement patterns computed at a pressure of +2.2 kPa from FE models for cats MY1216L, MY1630L, and MY0923R, are shown by iso-amplitude contours in the left-hand column of Fig. 2. In all plots, we present the component of displacement perpendicular to the plane of the tympanic ring. The difference in amplitude between adjacent contours is 25 μm. As a guide, some of the contours are labeled. Simulated eardrum displacements are maximal in the postero-superior portion of the pars tensa, close to the pars flaccida. Patterns for smaller positive pressures are similar to that for a pressure of +2.2 kPa. (The exact variations in displacement as a function of pressure are described in Sec. III B.)

For comparison, the right-hand column of Fig. 2 shows our previously published displacement patterns measured with the moiré method (Ladak *et al.*, 2004). In those experiments, three cycles of pressure were applied to the middle-ear cavity. Each cycle started from rest (i.e., zero pressure in the cavities) and involved loading the eardrum by applying positive middle-ear pressures in the order 0.1, 0.2, 0.4, 0.7, 1.1, 1.6, and 2.2 kPa, then unloading back to 0 Pa in the reverse order. The eardrum was then loaded by applying negative pressures to the cavities in the same order and unloaded in the reverse order. A moiré measurement is made 5 s after the pressure is adjusted to the new level, and requires approximately 45 s. The figure shows the displacements measured at a pressure of +2.2 kPa during the first cycle; patterns for other positive pressures, including other cycles, are qualitatively similar but with differences in size as described in Sec. III B. Some of the iso-amplitude displacement contours appear to end abruptly because the periphery of the eardrum is obscured by overhanging tissue. The measured patterns are not smooth because they contain small systematic errors associated with the moiré apparatus (Ladak *et al.*, 2000). Nevertheless, a notable difference between the two sets of patterns (simulated and measured) is that the maximal pars-tensa displacement in the simulated patterns is located more superiorly than that in the measured data. The sizes of the simulated and experimental maximal pars-tensa displacements also differ. In cat MY1216L, the maximum value based on the simulation is twice as large as the measured value, and for cat MY1630L, the simulation-based value is 1.5 times as large as the measured value. For cat MY0923R, the simulation result is smaller than the measured value. The measured displacement pattern for cat MY0923R exhibits local maxima in both the anterior and the posterior pars tensa.

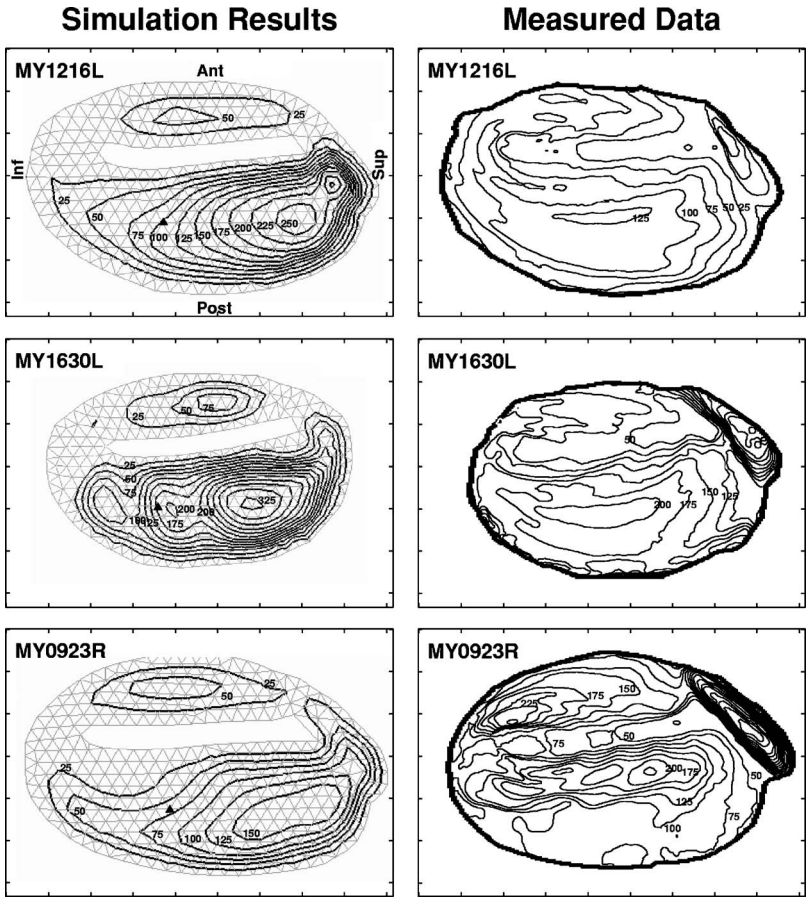


FIG. 2. Simulated and measured iso-amplitude displacement contours for a pressure of +2.2 kPa with an immobile malleus. Computed patterns are shown on the left, and corresponding measured data from Ladak *et al.* (2004) are shown on the right. The measured data are for the first cycle of pressurization. Displacements are in microns and adjacent contours are 25 μm apart. In cat MY0923R, measurements were made on the right ear instead of on the left ear as in the other two cats. In order to facilitate comparison of MY0923R with the other cats, the image of the eardrum was mirrored to look like the others. The triangles indicate points for which results are plotted in Fig. 5.

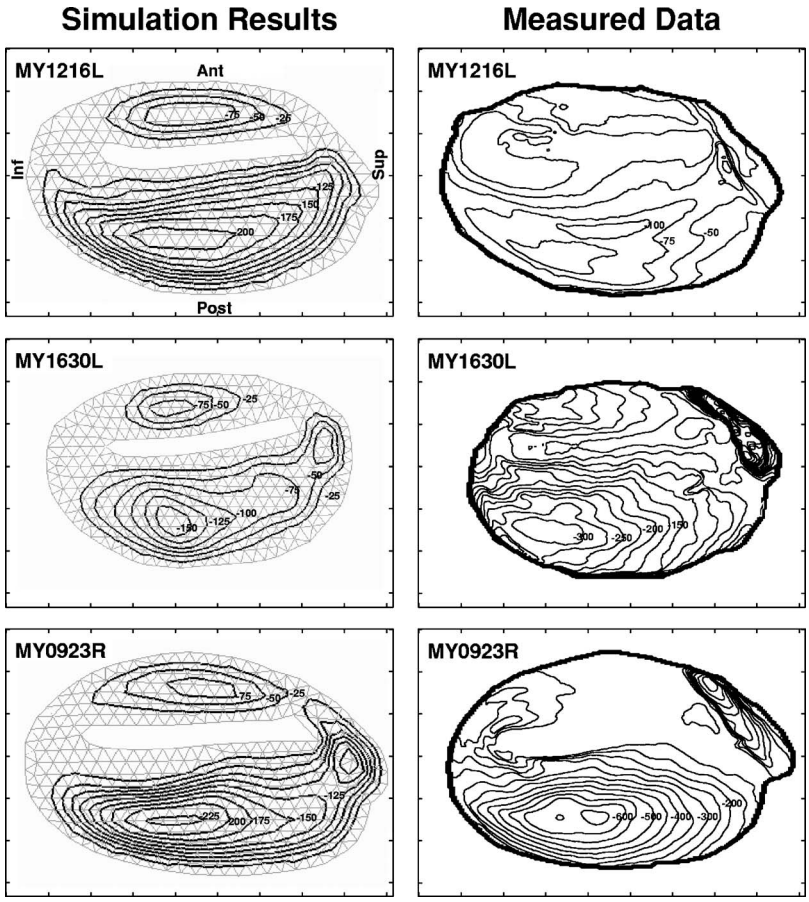


FIG. 3. Simulated and measured iso-amplitude displacement contours for a pressure of −2.2 kPa with an immobile malleus, displayed as in Fig. 2. The negative signs indicate that the displacements are directed medially.

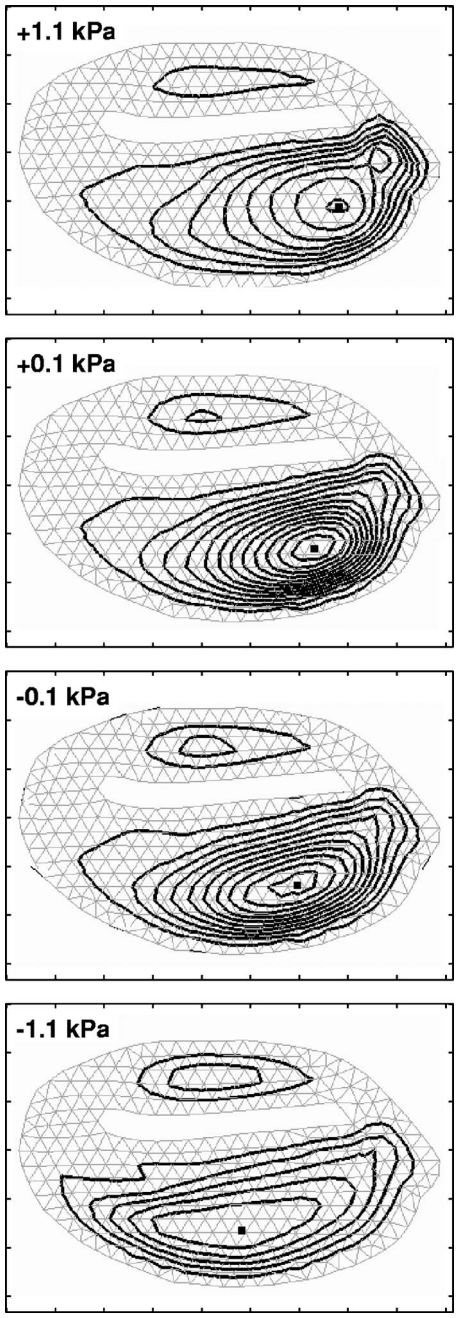


FIG. 4. Simulated iso-amplitude displacement contours for cat MY1216L for pressures of +1.1, +0.1, −0.1, and −1.1 kPa. Displacements are in microns. For pressures of +1.1 and −1.1 kPa, adjacent contours are 25 μm apart, whereas for pressures of +0.1 and −0.1 kPa, adjacent contours are 2.5 μm apart. The square indicates the location of maximal pars-tensa displacement.

Figure 3 shows displacement patterns, both simulated and measured, for a pressure of −2.2 kPa. For large negative pressures, the point of maximum pars-tensa displacement in the simulated patterns is shifted inferiorly compared with that for positive pressures. A shift in the inferior direction can also be seen in the experimental data shown in the right-hand column of the same figure; however, the degree of shift is larger in the simulations than in the measured data. In both the simulations and the experiments, the shift in the location of maximum displacement occurs gradually with pressure. This is illustrated in Fig. 4 which shows computed displacement patterns for +1.1, +0.1, −0.1, and −1.1 kPa for cat MY1216L; the patterns for pressures of +2.2 and −2.2 kPa are shown in Figs. 2 and 3, respectively. For negative pressures close to zero (e.g., −0.1 to −0.2 kPa), patterns are similar to those for positive pressures, with the location of maximum displacement occurring in the postero-superior region of the pars tensa. As the eardrum is further loaded toward −2.2 kPa, the location of the maximum moves inferiorly. Trends in the experimental data are consistent with the computed displacement patterns, but artifacts in the measurements make it difficult to discern the shift.

For negative pressures, there are also differences in maximal pars-tensa displacement magnitudes between the simulations and the experimental data. Specifically, for −2.2 kPa, the displacement magnitude in the simulation for cat MY1216L is twice as large as the measured value. For cat MY1630L the simulation result is half the size of the measured value, and for cat MY0923R the simulation gives a result that is about one third the size of the measured value.

The FE models presented here are not truly subject specific because only the shapes of the models are individualized to match each cat's eardrum; the thickness and Young's modulus are not subject specific because these data were not available. Nevertheless, subject-specific comparisons between simulations and measured data are made because individual eardrum shape significantly affects displacement patterns.

Figure 5 shows the calculated maximum principal nominal strains at the two highest pressures used in this study. The nominal strain is defined as

$$\boldsymbol{\varepsilon}^N = \mathbf{V} - \mathbf{I}, \tag{2}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{V} = \sqrt{\mathbf{F}\mathbf{F}^{\mathbf{T}}}$ and $\mathbf{F} = \partial\mathbf{x}/\partial\mathbf{X}$ is the deformation gradient tensor which is computed from knowledge of the initial location of a material particle at some three-dimensional position in space specified by the vector $\mathbf{X}$ and its new position $\mathbf{x}$ after deformation. The nominal strain is computed with respect to the initial unpressurized state. *Maximum principal* strains represent the largest possible strains in the tissue. The largest strains in the pars tensa are generally less than 2%. In cats MY1216L and MY1630L, however, strains of 3 to 5% occur in a very small strip along the periphery of the postero-superior pars tensa. Strains in the pars flaccida are generally less than 5% with a few localized areas of higher strain. Although the calculated strains are small, displacements are large because the eardrum has a relatively large diameter and is very thin.

B. Pressure-displacement curves

Figure 6 shows pressure-displacement curves for the locations indicated by triangles in the left-hand column of Fig. 2. Simulated pressure-displacement curves are shown in black, whereas all cycles of measured data are shown in grey. The measured curves are for the same locations indicated by the triangles in Fig. 2. The measured curves vary from one cycle of pressurization to the next and exhibit hysteresis. All three measured cycles are presented for cats MY1216L and MY1630L; however, only the second and third cycles are

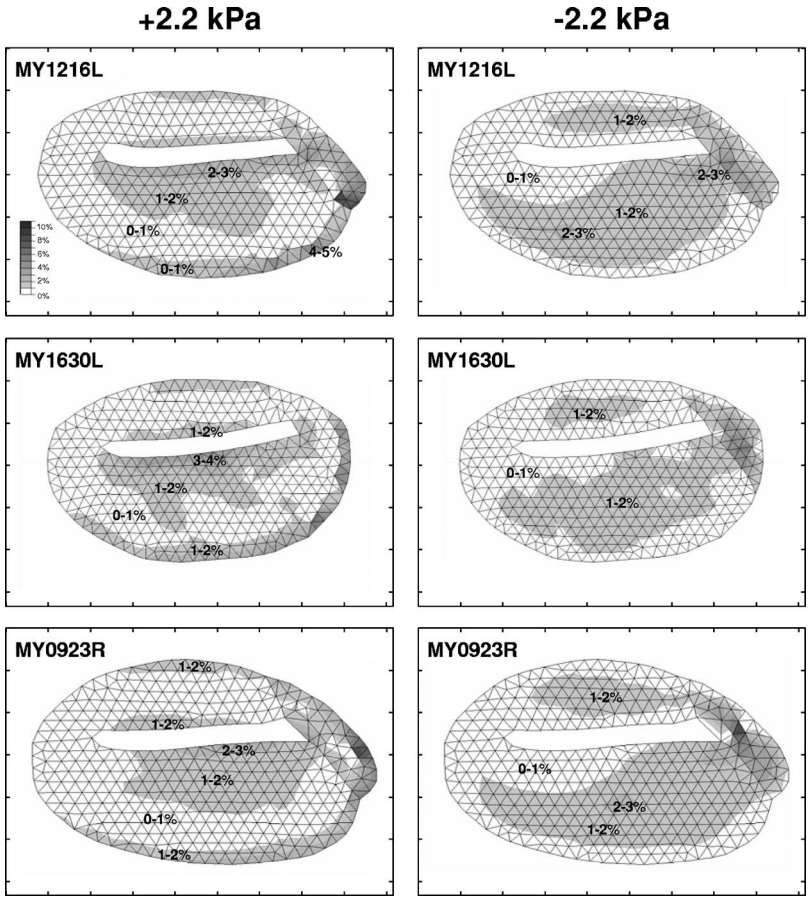


FIG. 5. Calculated maximum principal nominal strains for pressures of ±2.2 kPa with an immobile malleus.

presented for cat MY0923R. As previously noted (Ladak *et al.*, 2004), for cat MY0923R, the first cycle is considerably different from the second and third cycles. After the first cycle (not shown), the response appears to stabilize. Hysteresis is pronounced in this cat as indicated by the large nonzero displacement from the resting shape when the pressure is increased from 0 to +2.2 kPa and brought back to 0 kPa. There is only one simulated curve in the figure because the material is assumed to be elastic, with no hysteresis. For positive pressures, the simulated pressure-displacement curves agree reasonably well with the measured data. For negative pressures, displacements calculated from the nonlinear model for cat MY1216L are larger in magnitude than measured values. In cats MY1630L and MY0923R, there is substantial variation in measured displacements from one cycle to the next, especially for negative pressures. The variation for negative pressures may be due to mechanical instability (Ladak *et al.*, 2004). In any case, for both cats, simulated displacements fall within the range of measured displacements.

Both simulated and measured displacements grow less than in proportion to the applied pressure. For example, at a pressure of +1.1 kPa, the displacement calculated from the model for cat MY1216L is 73 μm; when the pressure is doubled to +2.2 kPa, however, the displacement does not double but increases by a factor of only 1.6, to 120 μm. Similarly, the measured displacement increases by a factor of 1.6, from 84 to 136 μm.

Displacements were also computed from purely linear models. The pressure-displacement curves in these cases are straight lines passing through the origin. Displacements calculated from linear and geometrically nonlinear models agree well with each other for low pressure levels. For instance, at a pressure of 0.1 kPa the displacements calculated from linear models are only 5 to 7% larger than those calculated from geometrically nonlinear models. For very high pressures, however, the linear models predict much larger displacement magnitudes than do the geometrically nonlinear models. For instance, when a linear model is used, the displacement magnitude computed for the point indicated by the triangle in Fig. 2 for cat MY1216L is twice as large as the value computed when a geometrically nonlinear model is used.

For all cats, the discrepancy between simulated (linear or nonlinear) and measured pressure-displacement curves is greatest for points in the superior portion of the posterior pars tensa. Simulation results here are larger in magnitude than measured values, which is consistent with the observation that for positive pressures the point of maximum pars-tensa displacement is located more superiorly in the simulations.

C. Effect of nonuniformity

The location of the displacement maximum can be shifted inferiorly in the simulations, to better match the measured patterns, by including nonuniformity in the models. Since the mechanical behavior of the eardrum is sensitive to both its thickness and its Young's modulus (Funnell and Laszlo, 1978), stiffening the superior region of the posterior pars

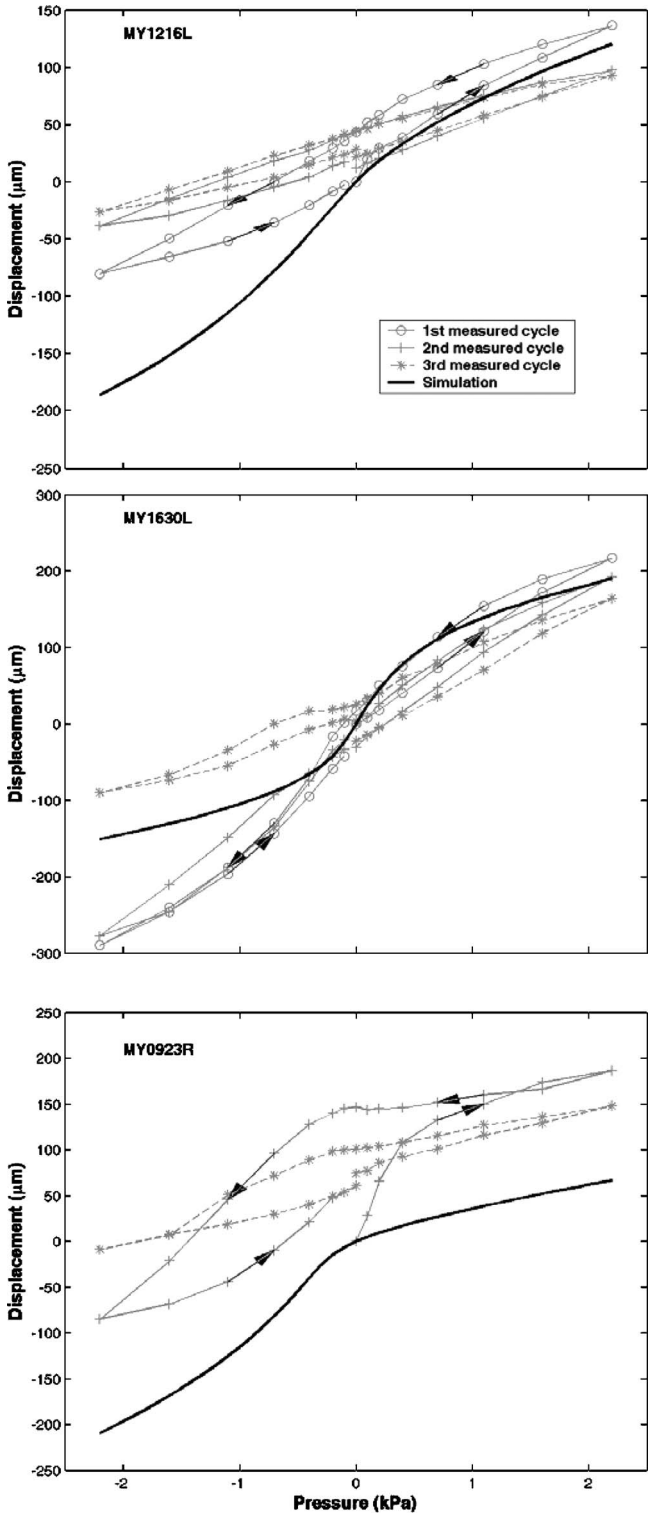


FIG. 6. Computed (black) and measured (grey) pressure-displacement curves for cats MY1216L, MY1630L, and MY0923R after mallear fixation, for locations indicated by triangles in the left-hand column of Fig. 2. The three cycles of measured data are from Ladak *et al.* (2004). The arrows on the measured curves indicate the temporal order of the measurements. Lateral displacements are positive, whereas medial displacements are negative.

tensa by increasing either the Young's modulus or the thickness can potentially shift the location of the maximum inferiorly. To demonstrate this, in the model for cat MY1216L, we arbitrarily increased the thickness of the superior third of the posterior pars tensa from 40 to 80 μm. Figure 7 shows

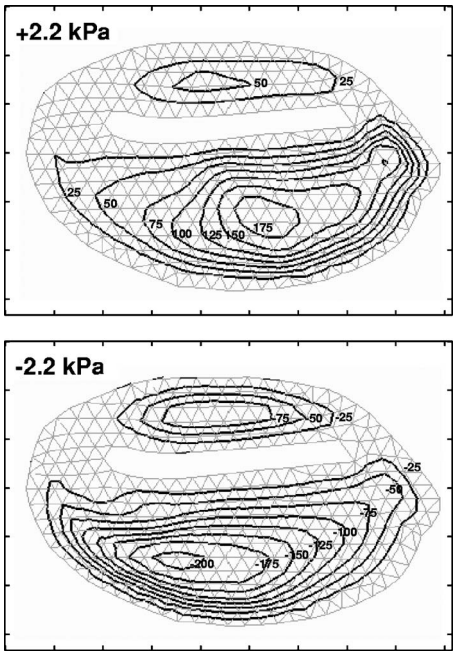


FIG. 7. Computed iso-amplitude displacement patterns for cat MY1216L for pressures of +2.2 and −2.2 kPa after increasing the thickness of the superior third of the posterior pars tensa from 40 to 80 μm. Adjacent iso-displacement contour lines (black) are 25 μm apart and are superimposed on the FE mesh (grey).

displacement patterns computed for pressures of +2.2 and −2.2 kPa. Increasing the thickness in this region has the effect of shifting the pars-tensa maximum inferiorly for positive pressures by decreasing the size of displacements in the postero-superior pars tensa; however, displacements in the inferior region are not significantly altered. Although displacements in the postero-superior region are smaller after the increase in thickness, they remain larger than measured values; no attempt was made to exactly match the simulated and measured displacements by adjusting the thickness of the pars tensa. The location of the maximum for negative pressures does not shift as much and remains in the inferior portion of the pars tensa.

IV. CONCLUSIONS AND DISCUSSION

When geometric nonlinearity is incorporated into FE models, the calculated displacements increase less than in proportion to the applied pressure for high pressures, i.e., those greater than about ±0.1 kPa. Such a lack of proportionality was also observed in the experimental data. By contrast, in *linear* simulations, computed displacements increase in proportion to the applied pressure because the effects of changing geometry on the stiffness matrix and on the direction of applied pressure are ignored.

For small pressures (both positive and negative) and for large positive pressures, the position of the maximum pars-tensa displacement is located more superiorly for the simulation results than it is for the experimental results if uniform mechanical properties are assumed for the model pars tensa. This difference between experiment and theory is also present in linear models (Funnell and Decraemer, 1996) and is not specific to the present addition of geometric nonlinear-

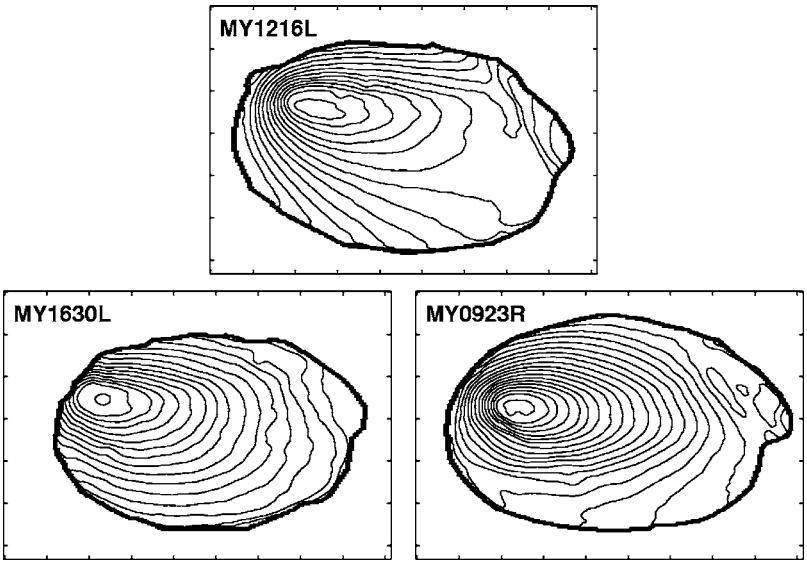


FIG. 8. Iso-depth shape contours. Adjacent contours are 0.1 mm apart. Ticks along the horizontal and vertical axes are 1 mm apart. For each cat, the thick black line encloses the visible portion of the eardrum. For cats MY1216L and MY1630L, measurements were made on the left eardrum, whereas the right eardrum was used for cat MY0923R. To facilitate comparison of cat MY0923R with the other cats, the image of the eardrum was mirrored to look like the others.

ity. By including nonuniformity in the models (e.g., by increasing the thickness of the superior portion of the posterior pars tensa), it is possible to shift the location of the maximum in the simulations toward the experimentally observed location. Such a thickness increase is consistent with recent experimental point-by-point thickness measurements (Kuypers *et al.*, 2005). Differences between simulated and measured displacement patterns and magnitudes can also be reduced by varying the Young's modulus of the pars tensa across its surface. The value used here for the pars-tensa thickness, adopted from our previous models and originally based on a value reported by Lim (1968), is considerably larger than the values found by Kuypers *et al.* (2005). Any adjustment downward of this model parameter could be counterbalanced by an increase in the value of the Young's modulus, as done by Fay *et al.* (2005). The Young's modulus used here, also adopted from our previous models, was originally based on a value reported by Békésy (1949) for a human-cadaver eardrum. The current work underscores the need for detailed quantitative measurement of eardrum properties across its surface.

The similarity in displacement patterns and magnitudes computed from the nonlinear models for *small* negative and positive pressures is expected since for small pressures, between 0 and ±0.1 kPa, the behavior of the nonlinear models is approximately linear, and a linearized model would predict the same displacements (apart from sign) for both positive and negative pressures. At higher negative pressures, the inclusion of geometric nonlinearity in the models accounts for the change in location of the maximum displacement. The same trends are observed in the measurements, where patterns for small negative pressures are found to be similar to patterns for small positive pressures, but the maximum is found to shift inferiorly for higher negative pressures.

As already noted, eardrum displacement magnitudes are significantly affected by the thickness, Young's modulus, and shape of the eardrum (Funnell and Laszlo, 1978), all of which can vary from one subject to the next. Furthermore, as discussed above, displacement patterns are affected by variations in the thickness and Young's modulus across the surface of an individual eardrum. Although we have taken into account individual eardrum shape in our FE models, we have not taken into account inter- and intra-individual differences in eardrum thickness and Young's modulus because of the unavailability of such data. The discrepancies between corresponding simulated and measured displacement patterns seen in Figs. 2 and 3 presumably arise in part because inter- and intra-individual variations in thickness and Young's modulus were not taken into account. As discussed below, the degree of malleus fixation may also affect eardrum displacement magnitudes, but to a smaller degree.

Inspecting the left-hand column of Figs. 2 and 3 shows that simulated displacement patterns vary between cats. In constructing the models, only the shapes of the models were individualized to the particular cat; the thickness and Young's modulus were assumed to be the same. Hence, any variations in simulated displacement patterns between individual models arise only from variations in the shapes. Figure 8 shows iso-depth contours for the measured resting shapes of the three cats; the difference in depth between adjacent contours is 0.1 mm. The depth of the eardrum is approximately 1.5 mm for the three cats used in the present study. However, there are subtle local variations in curvature between the eardrums as indicated by local variations in spacing between adjacent iso-depth contours. Simulated displacement magnitudes have previously been shown to be sensitive to both global (Funnell and Laszlo, 1978) and local (Funnell and Decraemer, 1996) variations in curvature. There is a noticeable difference in the lengths of the major and minor axes of the tympanic ring. The sensitivity study of Funnell *et al.* (1993) indicates that expansion or contraction of the tympanic ring does affect displacement magnitude.

In our models, the eardrum was assumed to be isotropic. This simplification is consistent with our previous models of the cat eardrum (e.g., Funnell and Laszlo, 1978; Funnell and Decraemer, 1996). Funnell and Laszlo (1978) found that the inclusion of anisotropy did not change the basic shape of the displacement patterns as compared to an isotropic model,

although the maximal pars-tensa displacement did increase. The actual nature and degree of anisotropy in the eardrum are not known.

The present study is limited to the case of an immobile-malleus condition. However, measurements indicate that the lack of proportionality between pressure and eardrum displacements is more pronounced with a normal mobile malleus (Ladak *et al.*, 2004). Under normal conditions, eardrum response would be affected by the mechanical load exerted on it by the rest of the middle ear and cochlea. One source of nonlinearity in middle-ear response may be due to an asymmetry in ossicular displacements at high static pressures caused by slippage in the joints between the malleus and the incus (Cancura, 1980; Hüttenbrink, 1988) and between the incus and the stapes (Hüttenbrink, 1988). Another source of nonlinearity in middle-ear response is thought to be due to the annular ligament connecting the stapes to the cochlea, which limits the displacement of the stapes (Price, 1974; Price and Kalb, 1986). Indeed, Price and Kalb (1991) feel that this is the main source of nonlinearity in middle-ear response at high pressures.

In the simulations, the manubrium is assumed to be completely immobile along its length. This only approximately simulates the experimental condition. In the experiments, an attempt was made to immobilize the malleus by gluing a point on the mallear head to the middle-ear wall. Although this does substantially reduce manubrial displacements relative to the case where the malleus is not fixed in this manner, it does not ensure that all points on the manubrium will be immobile (Ladak *et al.*, 2004). Indeed, the experimental data in Figs. 2 and 3 indicate that the manubrium deforms. Specifically, the ratio of umbo displacement after mallear fixation to the displacement without fixation was 21% for MY1216L, 30% for MY1630L, and 31% for MY0923R at a pressure of +2.2 kPa during the first cycle (Ladak *et al.*, 2004). The ratio of the displacement of the superior end of the manubrium after fixation to the displacement before fixation was 7% for cat MY1216L, 17% for cat MY1630L, and 22% for cat MY0923R (Ladak *et al.*, 2004). As the focus of this study was to investigate the effects of geometric nonlinearity on eardrum displacements, we have for simplicity assumed the manubrium to be immobile along its entire length. Although it would be more realistic to explicitly model the malleus with only its head constrained, any resulting small displacements of the manubrium would probably have only a very small effect on the displacements of the eardrum. For example, Funnell and Laszlo (1978) showed that doubling the ossicular-hinge stiffness in their middle-ear model decreased the maximal pars-tensa displacement by only 11%, with little change in the form of the displacement pattern. The effects in the present case would presumably be even smaller.

Material nonlinearity has been ignored in this work in order to investigate the effects of geometric nonlinearity alone. Uniaxial tension tests by Decraemer *et al.* (1980) on strips of pars tensa cut out from the eardrum suggest that the relationship between stress and strain is nonlinear for small strains and approximately linear for high strain levels. Data in their paper are presented as graphs of stress versus specimen length instead of stress versus strain since the resting length of the specimen could not be accurately determined in order to compute strain. It is difficult to determine the resting length of eardrum strips as they tend to wrinkle when no stress is applied. *A priori* it is not clear in which portion of the stress-strain curve (linear or nonlinear) the eardrum is functioning because quantitative experimental data on eardrum prestress are not available; existing data based on perforation tests are of a qualitative nature (Békésy, 1949; Kirikae, 1960) and their validity is questionable (Funnell and Laszlo, 1982). In any case, for small variations in strain with pressure, the stress-strain curve may be linearized. The largest pars-tensa strains computed from each of the models are generally less than 2%, which implies that a linearized material model is a reasonable approximation for most of the pars tensa for pressures up to ±2.2 kPa, at least under the condition of an approximately immobile malleus. For the pars flaccida, material nonlinearity may be more important since the strain variations are larger.

As our focus is to investigate the effects of including geometric nonlinearity, we have not modeled the hysteresis observed experimentally in the pressure-displacement curves during cyclical loading, which would require the inclusion of viscoelastic effects. This should then permit the quantitative modeling of eardrum behavior during impedance tympanometry, including the fact that differences in measurement protocols cause pronounced differences in tympanometric results (Osguthorpe and Lam, 1981).

ACKNOWLEDGMENTS

Funding for this work was provided through an operating grant from the Canadian Institutes of Health Research (WRJF), and through fellowships and grants from the Natural Sciences and Engineering Research Council of Canada (HML) and the Québec-Belgium Exchange Programme (HML). The authors thank J. Lauziére for editing the manuscript.

Bathe, K.-J. (**1982**). *Finite Element Procedures in Engineering Analysis* (Prentice-Hall, Englewood Cliffs), xiii+735 pp.

Békésy, G. v. (**1949**). "The structure of the middle ear and the hearing of one's own voice by bone conduction," J. Acoust. Soc. Am. **21**, 217–232.

Beer, H.-J., Bornitz, M., Hardtke, H.-J., Schmidt, R., Hofmann, G., Vogel, U., Zahnert, T., and Hüttenbrink, K.-B. (**1999**). "Modelling of components of the human middle ear and simulation of their dynamic behaviour," Audiol. Neuro-Otol. **4**, 156–162.

Bornitz, M., Zahnert, T., Hardtke, H.-J., and Hüttenbrink, K.-B. (**1999**). "Identification of parameters for the middle ear model," Audiol. Neuro-Otol. **4**, 163–169.

Cancura, W. (**1980**). "On the statics of malleus and incus and on the function of the malleus-incus joint," Acta Oto-Laryngol. **89**, 342–344.

Decraemer, W. F., Maes, M. A., and Vanhuyse, V. J. (**1980**). "An elastic stress-strain relation for soft biological tissues based on a structural model," J. Biomech. **13**, 463–468.

Dirckx, J. J. J., Decraemer, W. F., and Dielis, G. (**1988**). "Phase shift method based on object translation for full field automatic 3-D surface reconstruction from moiré topograms," Appl. Opt. **25**, 1164–1169.

Eiber, A. (**1999**). "Mechanical modeling and dynamical behaviour of the human middle ear," Audiol. Neuro-Otol. **4**, 170–177.

Fay, J. P., Puria, S., Decraemer, W. F., and Steele, C. R. (**2005**). "Three approaches for estimating the elastic modulus of the tympanic membrane," J. Biomech. **38**, 1807–1815.

Ferris, P., and Prendergast, P. J. (**2000**). "Middle-ear dynamics before and after ossicular replacement," J. Biomech. **33**, 581–590.

Fung, Y. C. (**1965**). *Foundations of Solid Mechanics* (Prentice-Hall, Englewood Cliffs, New Jersey).
Funnell, W. R. J. (**1983**). "On the undamped natural frequencies and mode shapes of a finite-element model of the cat eardrum," J. Acoust. Soc. Am. **73**, 1657–1661.
Funnell, W. R. J., and Decraemer, W. F. (**1996**). "On the incorporation of moiré shape measurements in finite-element models of the cat eardrum," J. Acoust. Soc. Am. **100**, 925–932.
Funnell, W. R. J., Decraemer, W. F., Dirckx, J. J. J., and Khanna, S. M. (**1993**). "On the effects of shape variations in a finite-element model of the cat eardrum," *16th ARO Midwinter Meeting* (The Association for Research in Otolaryngology, St. Petersburg Beach, Florida).
Funnell, W. R. J., and Laszlo, C. A. (**1978**). "Modeling of the cat eardrum as a thin shell using the finite-element method," J. Acoust. Soc. Am. **63**, 1461–1467.
Funnell, W. R. J., and Laszlo, C. A. (**1982**). "A critical review of experimental observations on ear-drum structure and function," ORL **44**, 181–205 [list of references is available at http://audilab.bmed.mcgill.ca/~funnell/AudiLab/earref.html].
Gan, R. Z., Feng, B., and Sun, Q. (**2004**). "Three-dimensional finite element modeling of human ear for sound transmission," Ann. Biomed. Eng. **32**, 847–859.
Hüttenbrink, K. N. (**1988**). "The mechanics of the middle ear at static air pressures," Acta Oto-Laryngol., Suppl. **451**, 1–35.
Kirikae, I. (**1960**). *The Structure and Function of the Middle Ear* (Univ. Tokyo Press, Tokyo).
Koike, T., Wada, H., and Kobayashi, T. (**2002**). "Modeling of the human middle ear using the finite-element method," J. Acoust. Soc. Am. **111**, 1306–1317.
Kuypers, L. C., Decraemer, W. F., Dirckx, J. J. J., and Timmermans, J.-P. (**2005**). "Thickness distribution of fresh eardrums of cat obtained with confocal microscopy," J. Assoc. Res. Otolaryngol. **6**, in press.
Ladak, H. M., Decraemer, W. F., Dirckx, J. J. J., and Funnell, W. R. J. (**2000**). "Systematic errors in small deformations measured using shadow-moiré topography," Appl. Opt. **39**, 3266–3275.
Ladak, H. M., Decraemer, W. F., Dirckx, J. J. J., and Funnell, W. R. J. (**2004**). "Response of the cat eardrum to static pressures: Mobile versus' immobile malleus," J. Acoust. Soc. Am. **116**, 3008–3021.
Lim, D. J. (**1968**). "Tympanic membrane. Electron microscopic observation. I. pars tensa," Acta Oto-Laryngol. **66**, 181–198.
Osguthorpe, J. D., and Lam, C. (**1981**). "Methodologic aspects of tympanometry in cats," Otolaryngol.-Head Neck Surg. **89**, 1037–1040.
Price, G. R. (**1974**). "Upper limit to stapes displacement: implications for hearing loss," J. Acoust. Soc. Am. **56**, 195–197.
Price, G. R., and Kalb, J. T. (**1986**). "Mathematical model of the effect of limited stapes displacement on hazard from intense sounds," J. Acoust. Soc. Am.(Suppl. 1), **80**, S122.
Price, G. R., and Kalb, J. T. (**1991**). "Insights into hazards from intense impulses from a mathematical model of the ear," J. Acoust. Soc. Am. **90**, 219–227.
Timoshenko, S., and Woinowsky-Krieger, S. (**1959**). *Theory of Plates and Shells*, 2d ed. (McGraw-Hill, New York).

Simultaneous latency estimations for distortion product otoacoustic emissions and envelope following responses[a)]

David W. Purcell,[b)] Patricia Van Roon, M. Sasha John, and Terence W. Picton
The Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, Canada, M6A 2E1

(Received 26 October 2005; revised 6 March 2006; accepted 10 March 2006)

The purpose of this research was to simultaneously estimate processing delays in the cochlea and brainstem using the same acoustic stimuli. Apparent latencies were estimated from ear canal measurements of $2f_1$-f_2 distortion product otoacoustic emissions (DPOAEs), and scalp recordings of the f_2-f_1 envelope following response (EFR). The stimuli were equal level tone pairs (65 dB SPL) with the upper tone f_2 set at either 900 or 1800 Hz to fix the initiation site of the DPOAE and EFR. The frequency of f_1 was swept continuously between frequency limits chosen to keep the EFR response between 150 and 170 Hz. The average DPOAE latencies were 9.6 and 6.2 ms for f_2 =900 and 1800 Hz, and the corresponding EFR latencies were 12.4 and 8.8 ms. In a control condition, a third (suppressor) tone was added near the DPOAE response frequency to evaluate whether the potential source at f_{dp} was contributing significantly to the measured emission. DPOAE latency is the sum of both inward and outward cochlear delays. The EFR apparent latency is the sum of inward cochlear delay and neural processing delay. Neural delay was estimated as approximately 5.3 ms for both frequencies of stimulation.
[DOI: 10.1121/1.2191616]

PACS number(s): 43.64.Jb, 43.64.Ri [BLM] Pages: 2869–2880

I. INTRODUCTION

Acoustic stimuli in the ear canal propagate through the middle ear, travel as a wave to the appropriate place in the cochlea where neural encoding begins, and ascend to higher auditory centers before being perceived as sound. The goal of this research was to estimate processing delays in the cochlea, using acoustic responses from the hair cells, and in the brainstem, using electrical responses from the auditory nervous system. A tone pair with higher frequency f_2 and lower frequency f_1 elicited both responses simultaneously. The acoustic response from the cochlea was the distortion product otoacoustic emission (DPOAE) occurring at the frequency $2f_1$-f_2. The electrical signal was the envelope following response (EFR) occurring at the frequency f_2-f_1.

Processing delays were estimated using the phase versus frequency relationship to calculate apparent latency (Regan, 1966, 1989) or group delay (Goldstein *et al.*, 1971; O Mahoney and Kemp, 1995). Both DPOAEs (Talmadge *et al.*, 1999; Shera and Guinan, 1999; Knight and Kemp, 2000) and EFRs (Dolphin, 1997; Schoonhoven *et al.*, 2001) are initiated at the f_2 characteristic place on the basilar membrane where the stimulus tones overlap most prevalently. The upper tone f_2 was therefore fixed in frequency to maintain the initiation site at a fixed characteristic place (O Mahoney and Kemp, 1995; Moulin and Kemp, 1996a). The f_1 tone was swept slowly but continuously over a small range to obtain phase measurements for the calculation of apparent latency. There is some debate about whether to hold f_1 or f_2 fixed during sweep measurements (Moulin and Kemp, 1996a, b; Bowman *et al.*, 1997). An "f_1-sweep" was chosen to limit the DPOAE and EFR initiation sites to the two f_2 frequencies employed (900 and 1800 Hz).

Latency of the DPOAE measured in the ear canal is the sum of both inward cochlear delay (50% to 82% of the total) and outward cochlear delay (Moulin and Kemp, 1996a, b; Schoonhoven *et al.*, 2001). Propagation delays in the ear canal and middle ear are very short (Gardner and Hawley, 1972; Lutman and Martin, 1979; Giguere and Woodland, 1992, 1994; Kimberley *et al.*, 1993) and will be implicitly combined with the basilar membrane travel times in the treatment of inward and outward cochlear delays. Inward propagation delay is at least 50% of the total DPOAE latency because the response frequency f_{dp} is lower than stimulus f_2. Inward and outward delays are more likely near equal when the response is at the same frequency as the stimulus, such as for the stimulus frequency or transiently evoked otoacoustic emissions. Assuming a single DPOAE source at the f_2 characteristic place, the phase accumulation of the inward f_2 traveling wave will be greater than that of the outward response f_{dp} (O Mahoney and Kemp, 1995; Moulin and Kemp, 1996a, b; Shera and Guinan, 1999). As a traveling wave approaches its characteristic place, wavelength decreases and more phase accumulates for a given distance on the basilar membrane. Since the characteristic place of f_{dp} is more apical than f_2, an outward DPOAE launched from the characteristic place f_2 would experience fewer cycles.

Interpreting DPOAE latency is complicated when a source at the f_{dp} characteristic place contributes significantly to the measured response in the ear canal. In this study, the stimulus ratio f_2/f_1 was chosen to minimize the influence of

[a)]Part of these results were presented at the 19th International Evoked Response Audiometry Study Group meeting in Havana, Cuba, June 2005.
[b)]Author to whom correspondence should be addressed. Currently affiliated with the Department of Psychology, Queen's University, Kingston, Ontario, Canada, K7L 3N6. Electronic mail: purcelld@nca.uwo.ca

any source at f_{dp}. Ratios significantly smaller than 1.1 can elicit DPOAEs with a dominant source at f_{dp} (Shera and Guinan, 1999; Knight and Kemp, 2000, 2001; Dhar *et al.*, 2005). The smallest ratios were for f_2=1800 Hz and varied with swept f_1 from 1.091 to 1.104. To evaluate whether a significant source existed at f_{dp}, a third tone was added in a control condition at frequency f_{dp}+25 Hz. The purpose of this tone was to suppress any potential source at f_{dp} (Gaskill and Brown, 1996; Heitmann *et al.*, 1998; Talmadge *et al.*, 1999; Konrad-Martin *et al.*, 2001). If a significant f_{dp} source existed, significant changes in the response would occur.

Whereas DPOAE latency is the sum of inward and outward delays, EFR apparent latency is the sum of inward stimulus propagation delay and neural processing delay. These latencies and constituent delays are as follows: w is the DPOAE latency, l is the EFR latency, t is the inward stimulus delay of f_2 to the characteristic place of f_2, d is the outward DPOAE delay and n is the neural response delay.

It would be interesting theoretically and useful clinically to derive the last three delays. However, separating latencies measured for the DPOAE (w) and the EFR (l) into inward and outward cochlear traveling wave propagation, and neural processing delays, is problematic since the two latency equations $t+d=w$ and $t+n=l$ contain three unknown components. These equations can be solved if additional constraints are applied. A useful additional constraint is that the phase accumulated in the traveling wave of a tone to its characteristic place on the basilar membrane is independent of its frequency (Shera and Guinan, 1999). This constraint, together with recordings from different f_2 frequencies and the assumption that n is independent of f_2 frequency, can effectively provide a third equation to allow the estimation of separate cochlear (t,d) and neural delays (n).

II. METHODS

A. Subjects

Twenty-four adults (13 female) varying in age from 19 to 37 years took part in the experiment. All participants had normal hearing thresholds (≤20 dB HL) measured at 500, 750, 1000, 1500, 2000, 3000, and 4000 Hz. During measurements, participants were encouraged to sleep in a comfortable reclining chair located in an Industrial Acoustics Company (IAC) sound-insulated room. The experiment, including the audiogram, was generally completed in less than 90 min.

B. Auditory stimuli

A modified version of the MASTER research system (John and Picton, 2000b, see www.mastersystem.ca) was used to generate stimuli and extract the DPOAE and EFR responses. The stimuli were two different tone pairs. The higher frequency tone of each pair (f_2) was held constant at 900 or 1800 Hz. The lower tone (f_1) was swept slowly over a frequency range of approximately 20 Hz. This stimulus "sweep" lasted 15.36 s. Frequency f_1 was linearly ramped up 20 Hz for the first half of each sweep and ramped down for the second half. The stimuli were designed such that sweeps could be presented sequentially without any discontinuity and 60 sweeps were recorded in each block.

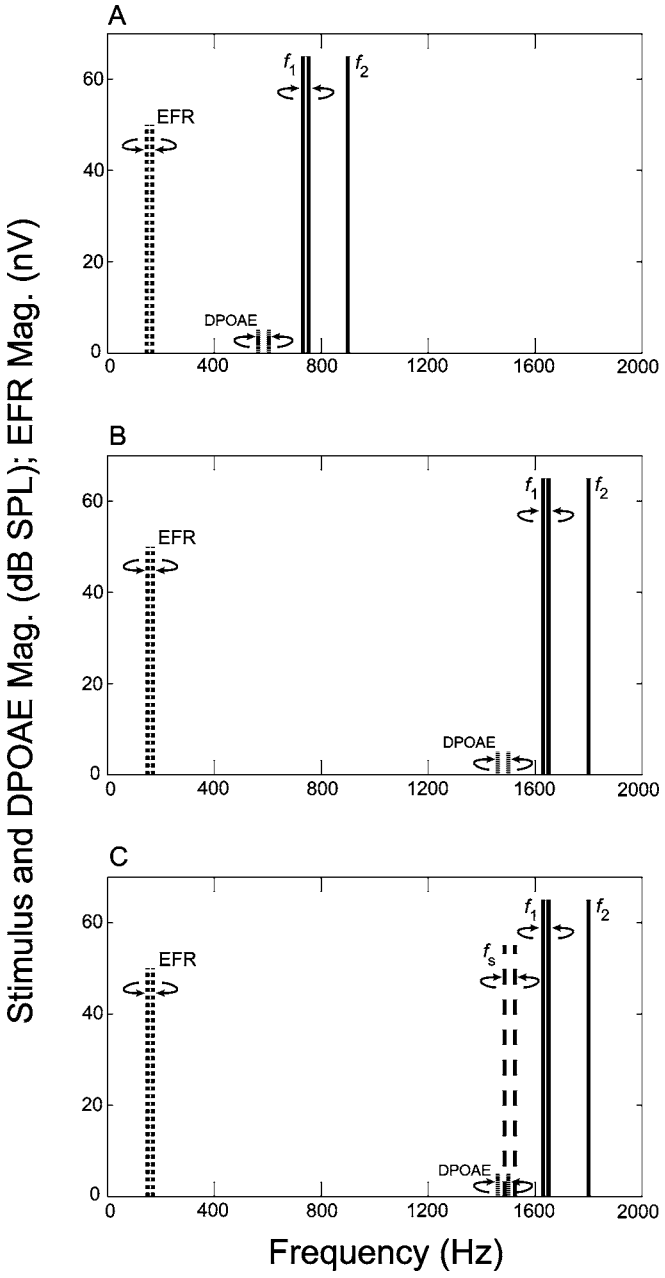


FIG. 1. The three stimulus conditions used in the study are represented in the frequency domain. The horizontal axes show frequency (Hz), and the vertical axes show the acoustic signal levels (dB SPL) and EFR response level (nV). Panel (a) displays the 65 dB SPL stimulus tone pair (solid lines) with f_2=900 Hz, and the bounds of f_1 that were swept from 732 to 752 Hz. An example robust 5 dB SPL DPOAE response is shown at 564 and 604 Hz (short dash lines) for the extremes of f_1. Similarly, a robust 50 nV EFR response is shown (dot lines) at 148 and 168 Hz for the extremes of f_1. Panel (b) displays the 65 dB SPL stimulus tone pair (solid lines) with f_2 =1800 Hz, and the bounds of f_1 swept from 1630 to 1650 Hz. A DPOAE response is displayed (short dash lines) at 1460 and 1500 Hz for the extremes of f_1. Similarly, the EFR is shown (dot lines) at 150 and 170 Hz, in response to the extremes of f_1. Panel (c) is identical to (b), but 55 dB SPL suppressor tones have been added (dash lines). These are 25 Hz above the instantaneous DPOAE frequency and are shown at 1485 and 1525 Hz.

For f_2=900 Hz, stimulus tone f_1 was swept from 732 to 752 Hz [Fig. 1(a)]. The EFR response frequency varied with f_2-f_1 between 148 and 168 Hz, and the DPOAE response frequency varied with $2f_1$-f_2 between 564 and 604 Hz. For f_2=1800 Hz, stimulus tone f_1 was swept from 1630 to

1650 Hz [Fig. 1(b)]. The EFR response frequency varied between 150 and 170 Hz, whereas the DPOAE response frequency varied between 1460 and 1500 Hz. As f_1 increased, the DPOAE increased in frequency, whereas the EFR decreased. Stimulus presentation order was randomized across subjects, and each f_2 condition was presented for two recording blocks in a balanced design for a total of 120 sweeps. The individual tones were presented at 65 dB SPL.

In addition to the two f_2 conditions described above, responses were also recorded for $f_2=1800$ Hz with an additional swept tone 25 Hz higher than the instantaneous DPOAE frequency, f_{dp} (Heitmann *et al.*, 1998). The purpose of this third tone was to act as a suppressor of any potential OAE source at the f_{dp} characteristic place. This suppressor tone (f_s) was presented on the f_1 channel at a level 10 dB below the level of the main stimulus tones f_2 and f_1 [Fig. 1(c)]. Only a single block of 60 sweeps was collected for this condition, and it was always recorded as the third data block (e.g., ABCBA). A single block was sufficient because it is easier to obtain an acceptable signal-to-noise ratio (SNR) with less averaging for the DPOAE. Since the purpose of this condition was to evaluate DPOAE sources, the EFR was not analyzed.

A National Instruments 6052E input/output board was used to perform digital-to-analog conversion at 32 kHz with 16-bit precision. Electrical stimulus amplitude was controlled using a Grason-Stadler Model 16 audiometer, and a pair of Etymotic ER-2 transducers generated the acoustic stimulus. These have a flat frequency response at the eardrum up to 10 kHz and were coupled to an Etymotic ER -10B+OAE probe with a rubber tip selected for fitting each subject's ear canal. To avoid equipment-generated distortion products, the f_2 and f_1 tones were produced on separate channels. Stimulus tones were calibrated using a Knowles DB-100 Zwislocki coupler and Brüel & Kjaer sound level meter (Siegel, 1994; Whitehead *et al.*, 1995).

C. Recordings

The ear receiving the stimulus was randomized across subjects. Microphones in the acoustic probe were used to measure the ear canal sound pressure. This signal was amplified using the +40 dB setting of the Etymotic ER-10B+ battery-powered amplifier and was band-pass filtered 40–4000 Hz with a Krohn-Hite Model 3322 filter with 24 dB/oct slope.

The EEG was measured using gold-plated Grass electrodes located at the vertex (Cz) and just below the hairline at the posterior midline of the neck (reference), with a ground on the collarbone. Electrode impedances were measured at 10 Hz as less than 8 kOhm using a FEZM-5D Grass electrode impedance meter. A Grass P50 battery-powered amplifier applied a gain of 10 000 and band-pass filtered the signal 1–1000 Hz. A further gain of 5 was applied by the acquisition board that sampled the acoustic and EEG signals at a rate of 16 kHz with 16-bit resolution. The EEG signal was down-sampled after acquisition to 2 kHz for analysis.

D. DPOAE and EFR analysis

Although the experimenter was provided with online response feedback, data analysis was performed offline so that noise-reduction techniques could be employed (Purcell *et al.*, 2004). For a given stimulus condition, acoustic and EEG data recorded in response to each stimulus sweep were synchronously averaged in the time domain. Each sweep was divided into 15 epochs of 1.024 s. The noise rejection method excluded epochs of data that exceeded a threshold noise level estimated from a broad frequency band containing the response frequency. This threshold noise level was obtained by first determining the mean and standard deviation (SD) of the noise level in all epochs of the acoustic and EEG data from a given measurement block. The threshold was then set as the mean noise level plus 3 SD for the DPOAE and plus 1.5 SD for the EFR. These values have been previously found by the authors to be a good compromise between producing good quality data without rejecting too many epochs. This rejection method led to a different number of epochs included in the average for each epoch of the 15-epoch sweep, but in practice fewer than ten (out of 160) were typically rejected.

A software implemented Fourier analyzer (FA) extracted the DPOAE from the average acoustic microphone record and the EFR from the average background EEG signal (Regan, 1989; Purcell *et al.*, 2004). Orthogonal reference sinusoids were matched in frequency to the desired signal to be extracted from the averaged response sweeps. For the EFR, the response frequency was f_2-f_1. For the DPOAE, the response frequency was $2f_1$-f_2. The level and phase of the stimulus tones in the ear canal were also obtained using "response" references of f_2 (fixed) and f_1. Two 1.024-s rectangular window filters were applied in series to each of the complex outputs of the FA. Since the stimuli were designed with symmetrical increasing and decreasing response frequencies in each half of the sweep, the second half of the FA sweep output was folded and vector averaged with the first. Physiological delays (up to a maximum of about 30 ms) were neglected during this fold and average because of the broad filter windows and the slow change in frequency over the sweep.

An F ratio (John and Picton, 2000b) determined whether the instantaneous response amplitude at a given frequency was statistically different from the background noise (acoustic for the DPOAE, and EEG for the EFR). A discrete Fourier transform (DFT) estimated the background noise of the average sweeps folded in the time domain. Specifically, the mean noise levels in $-/+20$ DFT frequency bins ($-/+1.3$ Hz) surrounding the instantaneous response frequencies were calculated and subsequently multiplied by a scaling factor of 2.38 (determined using simulated noise) to reconcile the narrower effective bandwidth of the DFT with that of the FA. The signal was judged to be significantly different from noise at $p<0.05$ when the SNR reached approximately 5 dB ($F=3$; df=2,80).

Apparent latency was estimated from the response phase (measured as cosine onset phase) versus frequency data. Stimulus delays from the transducers to the ear canal were

TABLE I. Arithmetic average response magnitudes and apparent latencies across participants (with SD shown in brackets) where $L_1=L_2=65$ dB SPL. Only individuals with valid apparent latencies are included in these summary statistics. Values are not reported for the EFR when the suppressor tone was added because only 60 sweeps were collected instead of the normal 120. For this final row, only individuals who had valid apparent latencies both with and without the suppressor are reported.

	DPOAE			EFR		
f_2(Hz)	Magnitude (dB SPL)	Apparent latency (ms)	N subjects in mean (out of 24)	Magnitude (nV)	Apparent latency (ms)	N subjects in mean (out of 24)
900	2.0 (7.2)	9.6 (2.0)	22	49.8 (15.2)	12.4 (5.1)	15
1800	0.3 (8.2)	6.2 (3.3)	21	45.8 (17.3)	8.8 (2.0)	21
1800 with suppressor	2.9 (12.5)	6.0 (5.1)	20	⋯	⋯	⋯

removed by correcting the response phases with the stimulus phases (denoted ϕ_2 and ϕ_1 for tones f_2 and f_1, respectively). For the EFR, ϕ_2-ϕ_1 was subtracted from the response phase, whereas for the DPOAE, $2\phi_1$-ϕ_2 was subtracted (O Mahoney and Kemp, 1995). Measurement system phase shifts were evaluated, but were not subtracted because they were very small. The estimation of apparent latency assumes that a single source with frequency-independent delay dominates the measured phase response and that phase therefore changes linearly with frequency. It was calculated as the slope of phase (measured in degrees) versus frequency:

$$L = -\delta\phi/(360\delta f). \tag{1}$$

Since noise can corrupt the estimated EFR and DPOAE phase responses, a weighted least squares linear regression was employed to estimate the phase versus frequency slope ($\delta\phi/\delta f$). The weighting varied with the SNR at each measurement frequency. For SNR $\leqslant$3 dB (i.e., when the response power was less than twice that of the noise power) a weight of 0 was assigned, whereas for SNR $\geqslant$15 dB the weight was 1. In between, the weight was scaled linearly between 0 and 1 with SNR (as expressed in dB). The points at the start and end of the phase versus frequency data were not employed due to end-effects associated with the filters of the FA.

Even with the weighted least squares approach, some subjects did not have phase versus frequency responses that could be adequately described by a linear fit, or that led to rational latency estimates. This could have been due to noise (if the noise estimates were inaccurate) or to the superposition of several response sources with different phases and similar magnitudes. Subjects were removed from the apparent latency analysis for a given response type (DPOAE or EFR) if (*i*) more than 75% of their response sweep had zero weighting, (*ii*) the weighted phase versus frequency slope was positive (indicating a negative latency), or (*iii*) the correlation between the phase data and the residual phase error (the *difference* between the data and model) had a Pearson coefficient that was greater than 0.80 (an indication that the linear regression was not adequately modeling the data). These procedures eliminated about 10% of the DPOAE measurements and about a third of the EFR measurements (see "*N*" columns in Table I).

E. Separation of delay constituents

In order to separate the apparent latencies into their constituent parts, the assumption was made that the EFR neural processing delay was independent of stimulus frequency f_2 (900 and 1800 Hz). The neural conduction delays from the 900- and 1800-Hz characteristic places might be slightly different due to their different positions along the modiolus of the cochlea, but the rest of the pathway to the response generator in the brainstem (Herdman *et al.*, 2002) is basically the same and the neural differences within the cochlea are probably not significant. Assuming that neural delay is independent of frequency, the difference in EFR apparent latency between the two values of f_2 is due only to the change of inward f_2 stimulus delay,

$$l_{900} - l_{1800} = t_{900} - t_{1800}, \tag{2}$$

where l is the latency of the EFR and t is the inward stimulus delay.

In DPOAE measurements, the stimulus delay is limited by the time for the f_2 tone to build up at its characteristic place (O Mahoney and Kemp, 1995; Moulin and Kemp, 1996a, b). The time required for the f_2 tone to reach its characteristic place is

$$t = \theta/(2\pi f_2), \tag{3}$$

where θ is the phase accumulated by the f_2 traveling wave to the characteristic place of f_2. Both delay t and θ are unknown for a given measurement.

To solve this equation it is necessary to employ the theory that the phase accumulated in the traveling wave from the oval window to a particular characteristic place is almost frequency independent (Shera and Guinan, 1999). Substituting Eq. (3) in Eq. (2),

$$l_{900} - l_{1800} = \theta(1/(2\pi 900) - 1/(2\pi 1800)). \tag{4}$$

After solving for θ, the inward stimulus delays for both values of f_2 can be calculated with Eq. (3). Outward DPOAE delays can then be evaluated with

$$d = w - t, \tag{5}$$

where d is the outward DPOAE delay and w is the measured DPOAE apparent latency. Finally, the neural portion of the EFR apparent latency can be determined as

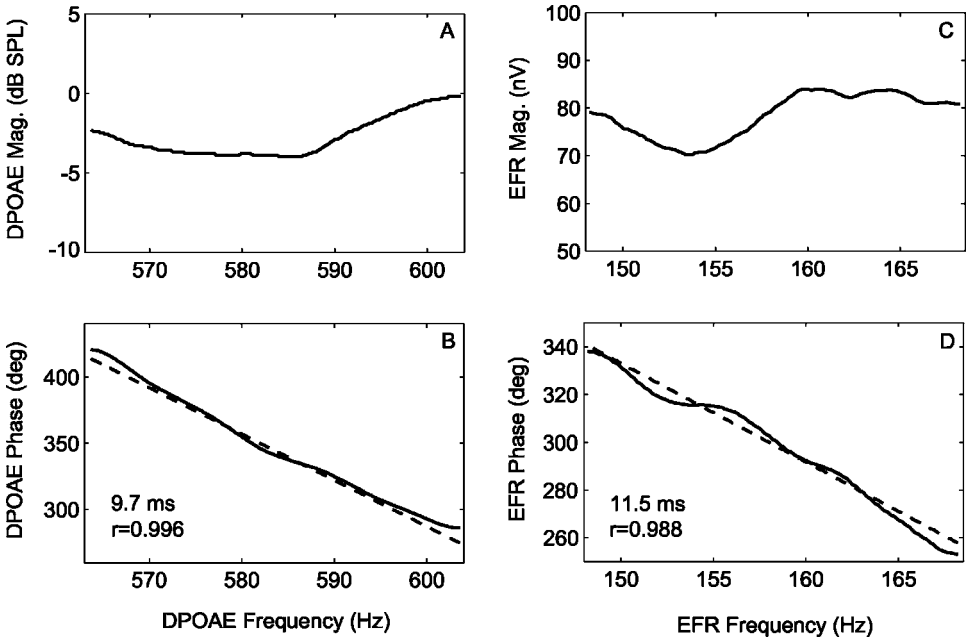


FIG. 2. Responses from a single individual for 120 averaged sweeps are presented for f_2=900 Hz at 65 dB SPL. Panel (a) shows the DPOAE response in dB SPL across frequency as f_1 was swept in frequency. Panel (c) shows the EFR magnitude response that was obtained simultaneously in nV. Phase responses for the DPOAE and EFR are displayed in panels (b) and (d), respectively (solid lines). Weighted linear regressions are shown with dash lines. Apparent latencies and correlation coefficients are reported in the lower left corner of the phase versus frequency plots.

$$n = l - t, \tag{6}$$

where n is the estimated neural processing delay.

The variability of the latency estimates was such that we could not make rational estimates of the constituent delays from an individual subject's data. We therefore combined data across subjects using the arithmetic mean value of the measured apparent latencies.

III. RESULTS

A. Magnitude and apparent latency

The magnitudes and apparent latencies, arithmetically averaged across participants, are given in Table I. A two-way analysis of variance (ANOVA) was performed using Greenhouse-Geisser corrections for apparent latency with factors of response type (DPOAE or EFR) and f_2 frequency (900 or 1800 Hz). The only significant effect was f_2 frequency (F=12.8; df=1,10; $p<0.01$) and there was no interaction. DPOAE and EFR latencies were significantly longer for 900 Hz than for 1800 Hz. Two-tailed separate variance t tests also showed that there was no significant effect of f_2 on the magnitude of the elicited DPOAE or EFR (DPOAE t =0.72; df=40; p=0.48;EFR t=0.74; df=33;p=0.46). The DPOAE measures of magnitude and apparent latency were not significantly affected by the suppressor tone, as assessed using two-tailed separate variance t tests (magnitude t= −0.80; df=33;p=0.43; latency t=0.24; df=35;p=0.81; similar results for paired t tests).

Figure 2 shows example data from a single participant where f_2=900 Hz. For this individual both the DPOAE and EFR had a high SNR (not shown) across the entire measurement band. The DPOAE magnitude varied by about 4 dB across frequency, and the EFR varied by about 14 nV. The DPOAE phase plot shows an apparent latency of 9.7 ms, and linear regression provided a good model of the response (r =0.996). The correlation between the residual phase and the data was only r=0.08. From the EFR phase plot an apparent latency of 11.5 ms was calculated. The phase versus frequency curve shows more deviation from a straight line, but the corresponding linear regression was adequate (r=0.988) albeit less ideal than for the case of the DPOAE. For the EFR, the correlation between the residual phase and the data was r=0.16.

Figure 3 shows f_2=1800 Hz data from a second individual who had a less ideal response. This participant's responses were significant for most of the frequency ranges, but the SNR tended to decrease from low to high frequency. The range of DPOAE magnitudes was about 6 dB, and the weight of points in the phase versus frequency plot decreased from 1.0 to about 0.2 at higher frequencies (not shown). The DPOAE apparent latency was 6.0 ms with a less ideal correlation between modeled and measured data (r=0.946). The correlation between residual phase and the DPOAE data was r=0.35. The EFR decreased about 40 nV from low to high frequency. The corresponding SNR led to weights from about 0.8 to 0.0 as frequency increased. The calculated EFR apparent latency was 9.4 ms with r=0.936. The correlation between residual phase and EFR data was r=0.61.

B. Separation of delay constituents

Apparent latencies separated into estimates of inward stimulus delay and response delays are given in Table II. Using the measured EFR apparent latencies, parameter θ was first calculated using Eq. (4) in the fourth row of Table II. This phase accumulated in the traveling wave was then used to determine the inward stimulus delays for both f_2 frequencies with Eq. (3) in rows 5 and 6. With the estimates of inward stimulus delays and the measured EFR apparent latencies, the neural response delay was evaluated using Eq. (6) in row 7. This delay is considered independent of stimulus frequency f_2. Again with the inward stimulus delays and the measured DPOAE apparent latencies, the outward DPOAE response delays were estimated using Eq. (5) in

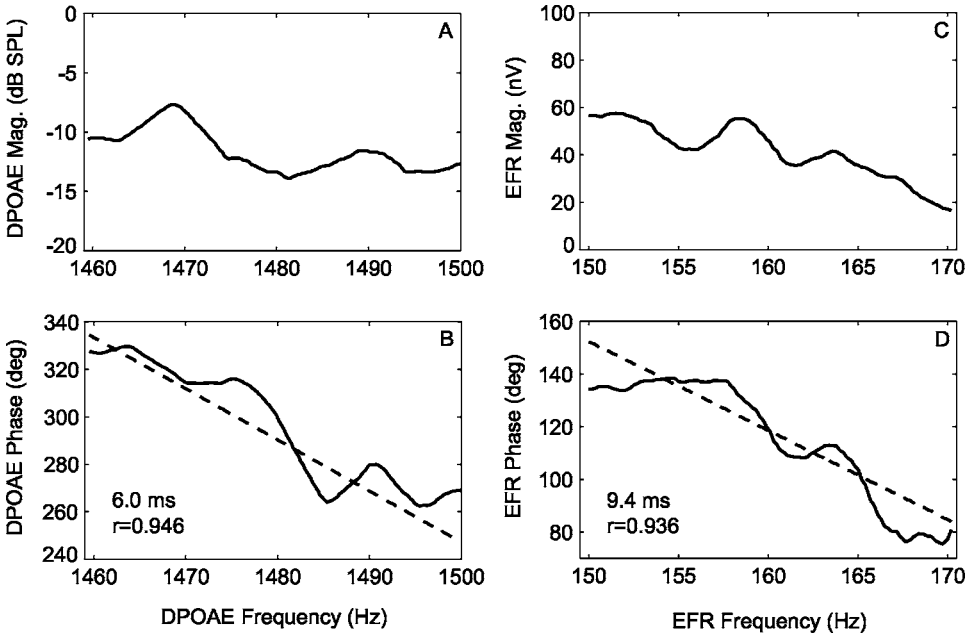


FIG. 3. Responses from a second single individual for 120 averaged sweeps with f_2=1800 Hz at 65 dB SPL. Panels display magnitude and phase data for the DPOAE and EFR as in Fig. 2.

rows 10 and 11. The inward stimulus delays were also represented as percentages of the total DPOAE latencies in rows 12 and 13.

IV. DISCUSSION

The experiment reported here measured DPOAE and EFR responses simultaneously using the same stimuli. While concurrent measurements have been reported previously for transient (e.g., Hoth and Lochmann, 1999) and steady-state responses (Purcell *et al.*, 2003; Rosner *et al.*, 2005), the stimulus here has been optimized for the estimation of apparent latency by employing a continuous sweep of the response frequency. DPOAE apparent latency or group delay has generally been reported using tone pairs of fixed frequency, where f_1 or f_2 is incremented in small steps between measurements (e.g., O Mahoney and Kemp, 1995; Bowman *et al.*, 1996). When f_2 is changed, it is termed an f_2 sweep and, correspondingly, when f_1 is changed, it is referred to as an f_1 sweep. Usually there are relatively few step measurements available for the estimation of latency, whereas continuous sweeps allow for many phase/frequency points (albeit not independent due to the analyzer filters). Continuous stimulus sweeps have been used extensively in EEG studies (e.g., Regan, 1989). An f_1 sweep was chosen here to fix the initiation site of the DPOAE at the characteristic place of f_2 during the estimation of apparent latency.

TABLE II. Estimates of component delays derived from measured DPOAE and EFR apparent latencies. Average measured apparent latencies are from Table I. Values are presented with extra precision to avoid rounding error when making illustrative calculations with values in this table. In the text, values are presented to one decimal place.

Description	Parameter	Equation	Value
Measured EFR latency for f_2=900 Hz	l_{900} (ms)	measured	12.4
Measured EFR latency for f_2=1800 Hz	l_{1800} (ms)	measured	8.8
Traveling wave accumulated phase	θ (rad)	(4)	39.85
Stimulus f_2=900 Hz inward delay	t_{900} (ms)	(3)	7.05
Stimulus f_2=1800 Hz inward delay	t_{1800} (ms)	(3)	3.52
Neural response delay of EFR	n (ms)	(6)	5.32
Measured DPOAE latency for f_2=900 Hz	w_{900} (ms)	measured	9.6
Measured DPOAE latency for f_2=1800 Hz	w_{1800} (ms)	measured	6.2
Outward DPOAE delay for f_2=900 Hz	d_{900} (ms)	(5)	2.54
Outward DPOAE delay for f_2=1800 Hz	d_{1800} (ms)	(5)	2.64
Inward delay as % of total for f_2=900 Hz	$\%t_{900}$(%)	100 t_{900}/w_{900}	73.5
Inward delay as % of total for f_2=1800 Hz	$\%t_{1800}$(%)	100 t_{1800}/w_{1800}	57.2

A. Variability of apparent latency measurements

Apparent latency varies considerably across subjects as reflected by the SD values in Table I. This is in stark contrast to ABR latencies, which are very tightly distributed in the normal-hearing population [e.g., reviewed on p. 265 in Regan (1989)]. The small variance in ABR latencies suggests that the forward stimulus latencies applicable to responses elicited with transient stimuli are not so variable between individuals (Hoth and Weber, 2001). The apparent latency observed from a single delay has a linear relationship between phase and frequency in the sufficiently broad frequency range of interest. Accordingly, any significant filtering introduced by the measurement equipment must be removed from the phase response prior to the calculation. Unfortunately, the auditory system itself may apply significant filtering that cannot be characterized, both in the generation and transmission of DPOAEs, or in neural processing (Bijl and Veringa, 1985; Regan, 1989). In some measure-

ments, a physiological filter can be inferred from the magnitude versus frequency plot, and the phase response can be corrected accordingly (Regan, 1989). Indeed, the EFR magnitude response of Fig. 3 looks low-pass filtered, which would increase the (unweighted) apparent latency estimate. Without weighting, the EFR latency estimate for this participant was 10.6 ms. If the effects of a hypothetical low-pass filter were included in the analysis, the latency estimate would be decreased. These corrections were not, however, attempted in the analysis here because subjects showed different patterns of amplitude fluctuations.

In addition, the physiological signals that are measured tend to have poor SNR, and thus the phase response is distorted by noise. This can be attenuated but never completely removed through signal processing and averaging. Efforts were made to prevent extraneous noise from influencing the estimation of apparent latency, but the intersubject SDs were still from 2 to 5 ms. There are large individual differences in DPOAEs (Probst *et al.*, 1991), and apparent latency can be affected by spontaneous emissions (O Mahoney and Kemp, 1995). Substantial intersubject variability has been reported previously for DPOAE latency, possibly due to individual acoustic and mechanical differences between middle ears and cochleas (Moulin and Kemp, 1996a). The EFR is a small signal relative to the background noise, and across subjects the mean SNR was about 9 dB for $f_2=900$ Hz. Noise in the signal could have played a role in the variance of the EFR apparent latency. This might be improved by averaging more sweeps and/or using a higher stimulus level.

Finally, the presence of multiple generators with different delays leads to a calculated apparent latency of a "virtual" source that can be quite different from the contributors. If the sources happen to sum constructively due to a fortuitous relative delay between them, then the estimated apparent latency can be roughly approximated by the magnitude weighted mean of the two contributors. If they sum destructively, however, the estimate of apparent latency can vary widely and become irrational. The measured DPOAE and EFR can both be generated by multiple sources with distinctly different delays if the stimulus parameters are chosen unfavorably. As discussed in the next section, the second long-latency DPOAE source at f_{dp} did not contribute significantly to the measured emission of individuals who passed the linear regression criteria outlined in the Methods (end of Sec. II D). The scalp-recorded 150–170-Hz EFR may be generated at several brainstem sites, and in different proportions across subjects, but the relatively high response frequency should have prevented the contribution of cortical sources (Herdman *et al.*, 2002; Purcell *et al.*, 2004).

Although the use of apparent latency is fraught with caveats, it is worth exploring as an estimate of true delay since there are few alternatives for use with steady-state responses in humans (Regan, 1989). Despite the constraints imposed by simultaneous recording of DPOAE and EFR responses (Purcell *et al.*, 2003), and despite the variability and uncertainty associated with apparent latency, the delays estimated here are similar to those reported by others studying one response type at a time with various methods.

B. DPOAE latencies

Measured DPOAEs in the ear canal can come from two main sources within the cochlea (Brown *et al.*, 1996; Gaskill and Brown, 1996; Heitmann *et al.*, 1998; Talmadge *et al.*, 1999). The initial initiation site of the DPOAE is at the f_2 characteristic place where the two stimulus tones interact most prevalently. For stimulus frequency ratios of about $f_2/f_1 \leqslant 1.1$, a second source at the characteristic place of the DPOAE f_{dp} can dominate the measured response (Shera and Guinan, 1999; Knight and Kemp, 2000, 2001; Dhar *et al.*, 2005). The apparent latency of a DPOAE dominated by an f_{dp} source is longer than that from an f_2 source because of additional traveling wave times within the cochlea (Knight and Kemp, 2000).

Since interpreting apparent latencies from multiple sources is difficult, the present experiment included a condition with a suppressor tone 25 Hz above f_{dp}. This was to evaluate whether a source at f_{dp} might be contributing significantly to the measurement when the stimulus frequency ratio was close to the transition between source types (i.e., for $f_2=1800$ Hz). The magnitude and apparent latency of individuals who passed the linear regression criteria (which implicitly required a single source) were not statistically affected by the suppressor tone. Thus the dominant DPOAE source was probably at the f_2 characteristic place for the included measurements. The DPOAE apparent latency was therefore simply divided into two components: inward stimulus and outward response delays (O Mahoney and Kemp, 1995; Moulin and Kemp, 1996a).

Stimulus frequency f_2 had a significant effect on the apparent latency of the DPOAEs. The longer latencies for $f_2=900$ Hz were expected since the inward stimulus traveling wave delay would be longer at lower frequencies. DPOAE latencies have been found to decrease monotonically with increasing f_2 (e.g., Bowman *et al.*, 1997; Schoonhoven *et al.*, 2001). There was, however, no effect of frequency f_2 on the DPOAE magnitude. There might have been an effect if the stimulus frequency ratio f_2/f_1 had changed enough between $f_2=900$ and 1800 Hz. DPOAE magnitude varies nonmonotonically with stimulus frequency ratio (Abdala, 1996). For $f_2=1800$ Hz, the ratio was about 1.10, whereas it was about 1.21 for $f_2=900$ Hz. On average these ratios elicited DPOAEs with similar response magnitude.

Table III summarizes some latency values from the literature that can be compared to those obtained here. Using concurrent measurement of DPOAEs and EFRs with multiple tone pairs of fixed frequency, Purcell *et al.* (2003) previously reported DPOAE apparent latencies that were a few milliseconds longer than those found here. The table gives latency values for $f_2=900$ Hz with $f_2/f_1=1.26$ (closest ratio to the 1.21 used here) and for $f_2=1800$ Hz with $f_2/f_1=1.12$ (compared to 1.10 used here). The longer latencies may in part be due to the 25 dB lower stimulus levels. For lower stimulus levels, the f_2 traveling wave should experience sharper tuning at its characteristic place, and corresponding

TABLE III. Summary of delay estimates from OAE and evoked potential (EP) studies. DPOAE studies used f_1 sweeps. Some entries are derived from functions or average values in the specified report, and the delay equations described here. Stimulus levels are in dB SPL unless otherwise noted. Click level is given for derived-band ABRs rather than masking noise level. Tabulated responses were those elicited with stimulus frequencies closest to the f_2 values employed here; actual stimulus frequencies are given in the fourth column.

Study	Technique	Stimulus level (dB SPL)		Actual stimulus frequency (Hz)	Emission latency (ms)		Evoked potential latency (ms)		Inward stimulus delay (ms)		Neural delay (ms)		Outward DPOAE delay (ms)	
		L1	L2	f_2 used here →	900	1800	900	1800	900	1800	900	1800	900	1800
Present	DPOAE, tone pair EFR	65	65	900, 1800	9.6	6.2	12.4	8.8	7.0	3.5	5.3	5.3	2.5	2.6
Purcell *et al.* (2003)	DPOAE, tone pair EFR	40	40	900, 1800	11.5	9.0	17.7	11.3	9.1	⋯	8.6	⋯	2.4	⋯
Whitehead *et al.* (1996)	DPOAE	75	75	1000–8000	<5.0	<5.0	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
Kimberley *et al.* (1993)	DPOAE	60	45	900, 1800	7.1	4.7	⋯	⋯	3.6	2.4	⋯	⋯	3.6	2.4
Bowman *et al.* (1997)	DPOAE	70	55	1587	⋯	4.4	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
Moulin and Kemp (1996a, b)	DPOAE	65	60	900,1800	7.8	5.7	⋯	⋯	5.4 – 6.4	3.9 – 4.7	⋯	⋯	1.4–2.4	1.0–1.8
John and Picton (2000a)	SAM tone EFR	60		750, 1500	⋯	⋯	15.1	10.9	⋯	⋯	⋯	⋯	⋯	⋯
Purcell *et al.* (2004)	SAM noise EFR	60		broadband	⋯	⋯	⋯	8.4	⋯	⋯	⋯	⋯	⋯	
Schoonhoven *et al.* (2001)	DPOAE, derived band ABR	65	55	900,1800	7.0	4.2	8.2	7.0	3.4	2.1	4.85[b]	4.85[b]	3.5	2.1
		70 dB nHL												
Neely *et al.* (1988)	TEOAE, tone-burst ABR	65		900, 1800	10.4	8.1	9.7	8.6	4.7	3.6	5.0	5.0	5.2	4.1
Don and Eggermont (1978)	derived band ABR	60 dB SL		900,1800	⋯	⋯	7.5	6.0	2.9[a]	1.4[a]	4.6[b]	4.6[b]	⋯	⋯
Eggermont and Don (1980)	derived band ABR	50 dB SL		1000, 2000	⋯	⋯	8.0	6.8	3.3[a]	2.1[a]	4.7[b]	4.7[b]	⋯	⋯
Donaldson and Ruth (1993)	derived band ABR	65 dB HL		787, 1503	⋯	⋯	10.0	8.4	⋯	⋯	⋯	⋯	⋯	⋯
Eggermont (1979)	narrow band AP	70 dB SL		900, 1800	⋯	⋯	4.5	3.0	3.7	2.2	⋯	⋯	⋯	⋯

[a]Calculated (EP latency - neural delay).
[b]0.8 ms added to reported value for synaptic delay.

larger phase accumulation (larger θ) and delay (Ruggero, 1992; Giguere and Woodland, 1992, 1994; Bowman *et al.*, 1997).

Transient evoked emissions reported by Neely *et al.* (1988) had longer latencies than the DPOAEs here, but a direct comparison is difficult because they did not use a phase gradient technique. Other DPOAE latencies are reported in the literature using f_1-sweep paradigms. Whitehead *et al.* (1996) have found shorter DPOAE latencies using higher-level stimuli ($L_1=L_2$), although one ear had a latency of about 8.5 ms for a stimulus near 1000 Hz. For unequal level stimuli, Kimberley *et al.* (1993) and Schoonhoven *et al.* (2001) reported shorter latencies. Bowman *et al.* (1997) report a shorter latency for unequal level stimuli spanning the 65 dB SPL value used here for an f_2 between 900 and 1800 Hz. Moulin and Kemp (1996a, b) also found shorter latencies using an f_1 sweep and similar, albeit unequal, stimulus levels. The majority of these studies used unequal level stimuli to optimize DPOAE response magnitude, whereas stimulus levels here were identical to maximize the modulation depth of the EFR stimulus and hence the response. This may have affected the apparent latencies, but the reported values are generally similar (albeit shorter) to those found here, given the intersubject variability.

C. EFR latencies

Stimulus frequency f_2 had a significant effect on EFR latency, which was shorter for $f_2=1800$ Hz. As for the DPOAE, this was anticipated due to the longer inward stimulus traveling wave delay at lower frequencies. There was no significant effect of f_2 on the magnitude of the elicited EFR. This was not unexpected since f_1 was chosen such that the range of response frequencies was essentially the same for both values of f_2. Envelope frequency f_2-f_1 determines that the EFR source was in the brainstem, since high EFR response frequencies are entirely brainstem whereas lower frequencies are both brainstem and cortex (Herdman *et al.*, 2002). The brainstem location of the generator should therefore not change significantly with f_2.

Again referring to Table III, Purcell *et al.* (2003) report longer EFR latencies than those found here for lower level stimuli with f_2-f_1 in the range of 180 Hz. John and Picton (2000a) report EFR latencies for sinusoidal amplitude modu-

lated (SAM) tones with carrier frequencies of 750 and 1500 Hz and modulation frequencies between 150 and 190 Hz. Their latencies are longer than found here, probably due to the lower carrier frequencies. Purcell *et al.* (2004) used SAM noise to elicit the EFR from 100 to 600 Hz. The average apparent latency was shorter than found here and probably reflects responses from areas of the cochlea with characteristic frequencies >1800 Hz, which would therefore experience shorter cochlear travel times than reported here.

Latencies from the peak of ABR wave V measured in frequency-specific paradigms are generally shorter than those reported here for frequencies similar to f_2=900 and 1800 Hz. For these frequencies, shorter delays are reported for derived band ABRs by Schoonhoven *et al.* (2001, their Fig. 3D plus corrections). Using tone-burst stimuli, Neely *et al.* (1988) provide a function where peak wave V latencies can be calculated. The 900-Hz value is shorter than here, but the 1800-Hz latency is very similar. In response to clicks plus noise, Don and Eggermont (1978) and Eggermont and Don (1980) report derived-band wave V latencies that are shorter than the EFR latencies here, but Donaldson and Ruth (1993) found a derived ABR latency quite similar for stimuli near 1800 Hz. Eggermont (1979) also found shorter latencies for narrow-band action potentials (APs) at higher stimulus levels, but these measurements would mostly represent cochlear delays. Shorter wave V latencies may occur because (*i*) the steady-state response is produced by a different generator, (*ii*) neural band-pass filtering causes additional delay in apparent latency calculations (Langner and Schreiner, 1988; Regan, 1989), (*iii*) the transient ABR stimuli with noise have not been adequately matched to the level of pure-tones, and/or (*iv*) responses to transient stimuli occur before the peak of the basilar membrane filter build-up required for steady-state responses (John and Picton, 2000a; Eggermont, 1979; Eggermont *et al.*, 1996; Purcell *et al.*, 2003).

D. Separation of delay constituents

Two assumptions were necessary to separate DPOAE and EFR apparent latencies into inward stimulus and response delays. The first assumption is that neural response delays are independent of the place on the basilar membrane that the EFR is initiated. The neural component of the tone burst auditory brainstem response (ABR) was assumed 5 ms for wave V independently of frequency and level in a study by Neely *et al.* (1988). Shirane and Harrison (1991) found that inferior colliculus neurons in the chinchilla responded with a latency of approximately 8 ms and that this latency was essentially independent of the characteristic frequency of cochlear fibers stimulated using supra-threshold current pulses. Derived responses obtained by Don and Eggermont (1978; Eggermont and Don, 1980) have shown that the latencies of ABR waves I–V are constant across intensity and frequency. Schoonhoven *et al.* (2001) used these findings to estimate cochlear delays from ABR peak V latencies.

The second assumption is that the phase accumulated in the f_2 traveling wave is independent of stimulus frequency f_2. This assumption (Shera and Guinan, 1999) should be reasonable for f_2 values away from the extremes of the cochlear frequency to place map (Greenwood, 1990). This assumption implicitly requires that the stimulus levels are similar for different f_2 so that the sharpness of the cochlear filter does not change (Ruggero, 1992). The stimuli in the present experiment were calibrated for an average ear with the Zwislocki coupler, but it is possible that the relative stimulus levels for f_2=900 and 1800 Hz could be different in the cochlea due to characteristics of the individual ear, especially the transfer function of the middle ear. However, the stimulus calibration was likely adequate to ensure similar signal levels in the cochlea for both values of f_2. The stimuli and DPOAEs could also experience different phase shifts across the ear canal and middle ear for each f_2. Although the relationship between inward stimulus delays for different values of f_2 was determined using θ, other cochlear models could be applied to estimate the relationship.

E. Inward stimulus delay

It was estimated that the inward f_2 stimulus would accumulate a frequency-independent 39.85 rad (6.34 cycles) to the characteristic place of f_2 where both response types were initiated. The higher frequency delay is half that of the lower because the two values of f_2 were an octave apart (which is not a requirement of the technique). The f_2=900 Hz delay here is shorter than that obtained previously using values from a cochlear model at a lower stimulus level (Purcell *et al.*, 2003). As above, shorter cochlear delays are expected at higher stimulus levels.

Cochlear travel times determined by Eggermont (1979) with click-evoked narrow-band action potentials were shorter than found here. The same is true for derived band ABR studies by Don and Eggermont (1978) and Eggermont and Don (1980). The stimulus levels here were different (note dB SL versus SPL in Table III), and responses to transient stimuli occur before the full build-up of the basilar membrane filter (Eggermont, 1979; Eggermont *et al.*, 1996). For the slow sweep of f_1 used here, the EFR responses are essentially steady state and therefore probably experience the full build-up of the basilar membrane filter and associated longer delay (John and Picton, 2000a).

Forward or inward stimulus latency can be estimated from an empirical function provided by Neely *et al.* (1988) from measurements of transient evoked otoacoustic emissions and tone burst ABRs. Their value was shorter for f_2 =900 Hz, but very similar at 1800 Hz. Neely *et al.* (1988) point out that the ABR is generated by the onset of the stimulus. Kimberley *et al.* (1993) halved their DPOAE latencies to obtain shorter inward travel times than those found here. They noted that this was a considerable simplification.

Schoonhoven *et al.* (2001) found that f_1-sweep latency was about twice the corrected delay from electrophysiological measurements and concluded that inward stimulus delay was about 50% of the total DPOAE latency. Their inward delay values from ABR measurements were shorter than those found here. They pointed out, however, that their results may not allow predictions at stimulus levels other than $L1$=65 dB SPL, $L2$=55 dB SPL, and clicks of 70 dB nHL. Indeed, different stimuli (pure tones and clicks with noise)

were used to elicit the various responses in the study, unlike the identical stimuli and simultaneous measurements made here. As discussed above, transient electrophysiological measurements may underestimate cochlear traveling wave delays for steady-state responses, and this would result in the underestimation of the inward stimulus delay as a percentage of total latency when transient measurements are used. Other researchers (O Mahoney and Kemp, 1995; Moulin and Kemp, 1996a, b) have reported larger percentages for inward stimulus delay. This was thought reasonable (e.g., Whitehead *et al.*, 1996) because the low-frequency outward emission from a single source at the f_2 place would be expected to accumulate fewer cycles in the reverse traveling wave than the stimulus (since the DPOAE characteristic place would be more apical). Previously reported values from f_1 sweeps have ranged from 69% evaluated using older electrophysiological data (Moulin and Kemp, 1996a) to 82% found by comparing the latencies of $2f_1$-f_2 and $3f_1$-$2f_2$ DPOAEs generated at the same f_2 place (Moulin and Kemp, 1996b). These percentages were used to estimate the inward stimulus delay ranges for Moulin and Kemp listed in Table III. The values are shorter than reported here at f_2=900 Hz and longer than for 1800 Hz. Using a cochlear model, Purcell *et al.* (2003) reported an inward delay of 79% for a single source DPOAE with f_2=900 Hz at 40 dB SPL. The percentages found here (Table II) of 74% (f_2=900 Hz) and 57% (f_2=1800 Hz) are within the (rather wide) range reported in the literature.

F. Outward DPOAE response delay

For otoacoustic emissions evoked with transient stimuli, the inward stimulus and outward response delays are very similar since they occur at the same frequency (although the reverse transfer function is not identical to the forward one). By halving the transient evoked emission latencies given by a function in Neely *et al.* (1988), the obtained results are several milliseconds longer than those estimated here for the outward DPOAE shown in Table III. However, their given function only accounts for 69% of the variance in their data. Schoonhoven *et al.* (2001) suggest outward DPOAE delay is half the total latency, leading to values that are slightly longer than here for f_2=900 Hz and slightly shorter when f_2=1800 Hz. An estimate of single source DPOAE outward delay at f_2=900 Hz and 40 dB SPL can be derived from Table I in Purcell *et al.* (2003). The outward delay is estimated by taking 21% of the experimental DPOAE latency of 11.5 ms. The result is very similar to the 900-Hz value reported here. Using the percentage of outward DPOAE delay with respect to total latency values of 18% (Moulin and Kemp, 1996b) and 31% (Moulin and Kemp, 1996a) and their total DPOAE latencies, outward delays can be estimated. The upper limit of outward delay estimates derived from their data and relations for f_2=900 Hz is similar to the value reported here, whereas the 1800-Hz value is shorter.

The outward DPOAE delay estimate should have been about 1 to 1.5 ms longer for f_2=900 Hz compared to 1800 Hz if the latencies were derived from a cochlear model (Giguere and Woodland, 1992, 1994). Instead, the calculated values were similar (Table II). The delay of the higher frequency DPOAE is likely increased (in part) by the fact that f_2/f_1 was smaller for f_2=1800 Hz. This pushes the f_{dp} characteristic place closer to the f_2 place, and thus increases the cycles accumulated in the outward f_{dp} signal from the f_2 place. The similarity of the two outward delay estimates is responsible for the change in the inward delay as a percentage of total latency from 74% to 57% for f_2=900 and 1800 Hz, respectively.

G. Neural processing delay

Neural processing delay above the cochlea was assumed to be independent of stimulus frequency f_2 and independent of small changes in envelope frequency f_2-f_1, such as between 150 and 170 Hz used here. Neural delay would be expected to change if f_2-f_1 were sufficiently altered to elicit responses from two different generator regions (e.g., cortical 40 Hz versus brainstem 100 Hz). The neural delay here was estimated as 5.3 ms, which is similar to the neural 5-ms delay (including synaptic delay) employed for tone burst wave V ABRs by Neely *et al.* (1988). Wave V–I delay estimates of 3.8 ms (Don and Eggermont, 1978) and 3.9 ms (Eggermont and Don, 1980) can be adjusted to the full neural delay estimated here and in Neely *et al.* (1988) by adding about 0.8 ms for the synaptic delay between the inner hair cells and auditory nerve (Eggermont, 1979), bringing the total neural delay to about 4.7 ms. Similarly, the wave V-I delays of Schoonhoven *et al.* (2001) can be compared to total neural delay reported here by adding 0.8 ms for the synaptic delay to give a total average neural delay of 4.85 ms. Despite reasonable agreement with the 5.3 ms reported here, it is possible that the multi-synaptic envelope-following circuit between the cochlea and the EFR generators may be to some extent distinct from the circuit subserving the transient ABR (Purcell *et al.*, 2003).

V. CONCLUDING REMARKS

The purpose of this study was to simultaneously measure acoustic latencies from the cochlea and electrical latencies from the auditory brainstem. A technique has been proposed to separate these latencies into stimulus and response delays. The analysis is complicated by the fact that the DPOAE response frequency is different from the pure tones that evoke it. In this regard it might be simpler to use stimulus frequency or transient evoked otoacoustic emissions where the stimulus and response share the same frequencies. Both stimulus frequency emissions and EFRs might be elicited by a SAM tone, for example (Rosner *et al.*, 2005). However, stimulus-frequency emissions are susceptible to electroacoustic artifacts whereas the DPOAE is robust and easy to measure. When DPOAEs and EFRs are simultaneously elicited by the same stimulus, the two responses have the same inward stimulus delay. Concurrent measurements may be helpful to understand the effects of stimulus parameters (overall intensity, frequency ratios, intensity ratios, etc.) on the different delays. Improved methods will be needed to reduce the variability of the measurements before the present technique can be used for individual subjects.

ACKNOWLEDGMENTS

Research supported by the Hearing Foundation of Canada, the Canadian Institutes of Health Research, and the National Institutes of Health Research.

Abdala, C. (**1996**). "Distortion product otoacoustic emission (2f1-f2) amplitude as a function of f2/f1 frequency ratio and primary tone level separation in human adults and neonates," J. Acoust. Soc. Am. **100**, 3726–3740.

Bijl, G. K., and Veringa, F. (**1985**). "Neural conduction time and steady-state evoked potentials," Electroencephalogr. Clin. Neurophysiol. **62**, 465–467.

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.

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.

Dhar, S., Long, G. R., Talmadge, C. L., and Tubis, A. (**2005**). "The effect of stimulus-frequency ratio on distortion product otoacoustic emission components," J. Acoust. Soc. Am. **117**, 3766–3776.

Dolphin, W. F. (**1997**). "The envelope following response to multiple tone pair stimuli," Hear. Res. **110**, 1–14.

Don, M., and Eggermont, J. J. (**1978**). "Analysis of the click-evoked brainstem potentials in man unsing high-pass noise masking," J. Acoust. Soc. Am. **63**, 1084–1092.

Donaldson, G. S., and Ruth, R. A. (**1993**). "Derived band auditory brainstem response estimates of traveling wave velocity in humans. I. Normal-hearing subjects," J. Acoust. Soc. Am. **93**, 940–951.

Eggermont, J. J. (**1979**). "Narrow-band AP latencies in normal and recruiting human ears," J. Acoust. Soc. Am. **65**, 463–470.

Eggermont, J. J., and Don, M. (**1980**). "Analysis of the click-evoked brainstem potentials in humans using high-pass noise masking. II. Effect of click intensity," J. Acoust. Soc. Am. **68**, 1671–1675.

Eggermont, J. J., Brown, D. K., Ponton, C. W., and Kimberley, B. P. (**1996**). "Comparison of distortion product otoacoustic emission (DPOAE) and auditory brain stem response (ABR) traveling wave delay measurements suggests frequency-specific synapse maturation," Ear Hear. **17**, 386–394.

Gardner, M. B., and Hawley, M. S. (**1972**). "Network representation of the external ear," J. Acoust. Soc. Am. **52**, 1620–1628.

Gaskill, S. A., and Brown, A. M. (**1996**). "Suppression of human acoustic distortion product: dual origin of 2f1-f2," J. Acoust. Soc. Am. **100**, 3268–3274.

Giguere, C., and Woodland, P. C. (**1992**). "Network representation of the middle and inner ear in a composite model of the auditory periphery," Proc. Inst. Acoust. **14**, 305–312.

Giguere, C., and Woodland, P. C. (**1994**). "A computational model of the auditory periphery for speech and hearing research. I. Ascending path," J. Acoust. Soc. Am. **95**, 331–342.

Goldstein, J. L., Baer, T., and Kiang, N. Y. S. (**1971**). "A theoretical treatment of latency, group delay, and tuning characteristics for auditory-nerve responses to clicks and tones," in *Physiology of the Auditory System; Based on the proceedings of a workshop*, edited by M. B. Sachs (National Educational Consultants, Baltimore, MD), pp. 133–141.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Heitmann, J., Waldmann, B., Schnitzler, H.-U., Plinkert, P. K., and Zenner, H.-P. (**1998**). "Suppression of distortion product otoacoustic emissions (DPOAE) near 2f1-f2 removes DP-gram fine structure—Evidence for a secondary generator," J. Acoust. Soc. Am. **103**, 1527–1531.

Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., and Picton, T. W. (**2002**). "Intracerebral sources of human auditory steady-state responses," Brain Topogr. **15**, 69–86.

Hoth, S., and Lochmann, H. (**1999**). "Dual response audiometry: a time-saving technique for enhanced objective auditory assessment," Audiology **38**, 235–240.

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**, 173–183.

John, M. S., and Picton, T. W. (**2000a**). "Human auditory steady-state responses to amplitude-modulated tones: phase and latency measurements," Hear. Res. **141**, 57–79.

John, M. S., and Picton, T. W. (**2000b**). "MASTER: a Windows program for recording multiple auditory steady-state responses," Comput. Methods Programs Biomed. **61**, 125–150.

Kimberley, 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.

Knight, R. D., and Kemp, D. T. (**2001**). "Wave and place fixed DPOAE maps of the human ear," J. Acoust. Soc. Am. **109**, 1513–1525.

Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). "Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears," J. Acoust. Soc. Am. **109**, 2862–2879.

Langner, G., and Schreiner, C. E. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," J. Neurophysiol. **60**, 1799–1822.

Lutman, M. E., and Martin, A. M. (**1979**). "Development of an electroacoustic analogue model of the middle ear and acoustic reflex," J. Sound Vib. **64**, 133–157.

Moulin, A., and Kemp, D. T. (**1996a**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. I. General characteristics," J. Acoust. Soc. Am. **100**, 1617–1639.

Moulin, A., and Kemp, D. T. (**1996b**). "Multicomponent acoustic distortion product otoacoustic emission phase in humans. II. Implications for distortion product otoacoustic emissions generation," J. Acoust. Soc. Am. **100**, 1640–1662.

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.

O Mahoney, C. F., and Kemp, D. T. (**1995**). "Distortion product otoacoustic emission delay measurement in human ears," J. Acoust. Soc. Am. **97**, 3721–3735.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.

Purcell, D. W., John, M. S., and Picton, T. W. (**2003**). "Concurrent measurement of distortion product otoacoustic emissions and auditory steady state evoked potentials," Hear. Res. **176**, 128–141.

Purcell, D. W., John, S. M., Schneider, B. A., and Picton, T. W. (**2004**). "Human temporal auditory acuity as assessed by envelope following responses," J. Acoust. Soc. Am. **116**, 3581–3593.

Regan, D. (**1966**). "Some characteristics of average steady-state and transient responses evoked by modulated light," Electroencephalogr. Clin. Neurophysiol. **20**, 238–248.

Regan, D. (**1989**). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (Elsevier Science, New York).

Rosner, T., Oswald, J., and Janssen, T. (2005). "Simultaneous measurement of auditory steady state responses and distortion product otacoustic emissions using amplitude modulated primary tones," in *XIX Int. Evoked Response Audiometry Study Group*, Havana, Cuba, p. 54.

Ruggero, M. A. (**1992**). "Responses to sound of the basilar membrane of the mammalian cochlea," Curr. Opin. Neurobiol. **2**, 449–456.

Schoonhoven, R., Prijs, V. F., and Schneider, S. (**2001**). "DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears," J. Acoust. Soc. Am. **109**, 1503–1512.

Shera, C. A., and Guinan, J. J., Jr. (**1999**). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: a taxonomy for mammalian OAEs," J. Acoust. Soc. Am. **105**, 782–798.

Shirane, M., and Harrison, R. V. (**1991**). "The effects of long and short term profound deafness on the responses of inferior colliculus to electrical stimulation of the cochlea," Acta Oto-Laryngol., Suppl. **489**, 32–40.

Siegel, J. H. (**1994**). "Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes," J. Acoust. Soc. Am. **95**, 2589–2597.

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.

Whitehead, M. L., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). "Effects of ear-canal standing waves on measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. **98**, 3200–3214.

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.

Signal to noise ratio analysis of maximum length sequence deconvolution of overlapping evoked potentials

Jorge Bohórquez
Department of Biomedical Engineering, College of Engineering, University of Miami, Coral Gables, Florida 33124

Özcan Özdamar
Department Otolaryngology and Pediatrics, Miller School of Medicine, University of Miami, Miami, Florida 33101

(Received 5 December 2005; revised 7 March 2006; accepted 9 March 2006)

In this study a general formula for the signal to noise ratio (SNR) of the maximum length sequence (MLS) deconvolution averaging is developed using the frequency domain framework of the generalized continuous loop averaging deconvolution procedure [Özdamar and Bohórquez, J. Acoust. Soc. Am. **119**, 429–438 (2006)]. This formulation takes advantage of the well known equivalency of energies in the time and frequency domains (Parseval's theorem) to show that in MLS deconvolution, SNR increases with the square root of half of the number of stimuli in the sweep. This increase is less than that of conventional averaging which is the square root of the number of sweeps averaged. Unlike arbitrary stimulus sequences that can attenuate or amplify phase unlocked noise depending on the frequency characteristics, the MLS deconvolution attenuates noise in all frequencies consistently. Furthermore, MLS and its zero-padded variations present optimal attenuation of noise at all frequencies yet they present a highly jittered stimulus sequence. In real recordings of evoked potentials, the time advantage gained by noise attenuation could be lost by the signal amplitude attenuation due to neural adaptation at high stimulus rates. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2191609]

PACS number(s): 43.64.Qh, 43.64.Ri [BLM] Pages: 2881–2888

I. INTRODUCTION

Transient evoked potentials (EP) are generally obtained with repetitive stimuli using synchronous signal averaging. The rate of stimuli is adjusted such that the response to each stimulus is complete prior to the presentation of the next stimulus. If this condition is not satisfied, overlapping of responses occurs and steady-state responses (SSR) are generated. SSRs generally display a complex set of time features which are generally analyzed using spectral analysis. Under these conditions, the transient EP, thought to be generating these responses cannot be recovered. In 1982, Eysholdt and Schreiner introduced a novel method to recover individual transient responses from overlapping responses obtained at high stimulus rates. This method uses a specially designed stimulus sequence called maximum length sequences (MLS) and since then has found many applications in basic and clinical research on evoked potential (e.g., Bell *et al.*, 2001, 2002; Burkard *et al.*, 1990, 1996a, 1996b; Chan *et al.*, 1992; Jirsa, 2001; Jiang *et al.*, 2003; Lasky, 1992; Leung *et al.*, 1998) and otoacoustic emission (Thornton, 1993a, b) studies.

MLS is part of a bigger family of sequences called pseudorandom binary sequences (PBS) which are cyclic sequences with flat frequency spectra and impulse circular autocorrelation functions. Because of their special properties, they are used in many diverse fields such as communications, cryptanalysis, systems theory, acoustics, and biomedicine. Other lesser known classes in the same family are Legendre sequences (LGS) and quadratic residue codes (QRC). Maximum length sequences are comprised of two elements, 1 and −1, and generated using well-known algorithms (Davies, 1966; De Graauw and Veltman, 1970). They are characterized by their generator base (b), number of ones (N), and sequence length (L). These parameters are interrelated by the simple rules $N=2^b$ and $L=2^*N-1$. MLS can be generated for all base integers of which the first five are listed in Table I.

For evoked response studies, a binary form (1, 0) of MLS is used with −1 substituted for 0 (Eysholdt and Schreiner, 1982). In this formulation, a "1" represents a stimulus and a "0" represents the lack of a stimulus. As a consequence, the autocorrelation does not produce a perfect impulse function. This difficulty, however, is remedied by using a recovery function with negative and positive elements. As a consequence the autocorrelation function of the binary form of the MLS with length L (MLS_L) results in

$$\mathrm{MLS}_L(t) \copyright \mathrm{MLS}_L(t) = \frac{N}{2}[\delta(t) + 1], \tag{1}$$

where t is the time index, $\delta(t)$ is the impulse function, and © designates auto or cross correlation function. The recovery sequence MLSR_L for the binary MLS_L incorporates the factor N so the following autocorrelation equation can be written as described by Eysholdt and Schreiner (1982) and Burkard *et al.* (1990).

$$\mathrm{MLS}_L(\mathrm{t}) \copyright \mathrm{MLSR}_L(\mathrm{t}) = \delta(t). \tag{2}$$

The left column of Fig. 1 shows the time diagram of the MLS_{31} function with $L=31$ and its recovery function

TABLE I. The listing of the first five maximum length sequences. Other sequences exist for lengths 7, 15, and 63.

Maximum length sequences	
Code	Sequence
MLS_3	1 1 0
MLS_7	1 1 1 0 0 1 0 1
MLS_{15}	1 1 1 1 0 1 0 1 1 0 0 1 0 0 0
MLS_{31}	1 1 1 0 1 1 0 0 0 0 1 1 1 1 1 0 0 1 1 0 1 0 0 1 0 0 0 0 1 0 1 0
MLS_{63}	1 1 1 1 1 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 1 0 1 0 0 1 1 1 1 0 1 0 1 0 0 0 0 1 1 1 0 0

$MLSR_{31}$. As seen, the recovery function $MLSR_{31}$ has essentially the same form as the MLS function with amplitudes attenuated by the number of ones (or stimuli) which is 16 in this case and zeros replaced by negative pulses with the same magnitude as the positive pulses. Corresponding frequency diagrams are shown on the right column. The flat spectrum of the MLS_{31} stimulus function is displayed on the right top with the dc value showing 16, the number of stimuli. The dc value of the deconvolution filter function is 0.0625 (inverse of 16) which is shown on the right bottom figure.

In auditory applications the interstimulus interval (ISI) or rate has a strong influence on the adaptation of the evoked response components. In conventional evoked potential (EP) studies, stimuli are presented at a fixed rate (constant ISI). For MLS responses, however, a mean rate or minimum ISI is used since the ISI is not constant and shows a wide distribution as a function of L (see Fig. 2). The ratio of the maximum to the minimum ISI can be defined as the stimulus jitter ratio and this ratio increases as a function of *L* as listed in Table II.

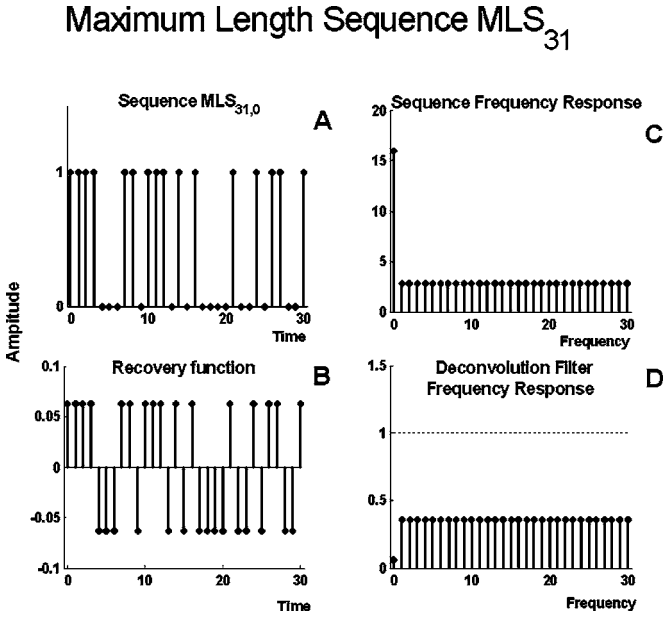


FIG. 1. (a) Time diagram of the MLS_{31} function with $L=31$. (b) Recovery function $MLSR_{31}$. (c) Stimulus frequency response function $S(f)$. (d) Deconvolution filter frequency response function $G(f)$ of the same stimulus sequence. The dotted line in (d) separates the noise amplification (above one) and attenuation (below one) regions.

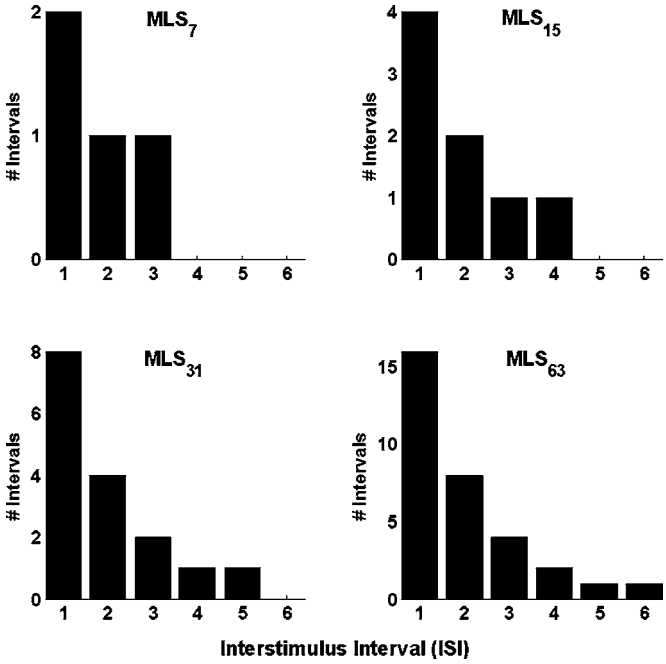


FIG. 2. Interstimulus interval (ISI) distribution histograms for the four MLS functions with lengths 7, 15, 31, and 63 listed in Table I.

Maximum length sequences have a very specific interstimulus interval (ISI) distribution, having a total of *N* components (Fig. 2). The highest component has $2^{(N-2)}$ intervals and each component's frequency decreases by a factor of 2 until reaching the *N*th component, which will have one occurrence (Picton *et al.*, 1992). Even for $N=7$, for which two distinct MLS were derived, the interpulse periods had identical patterns as shown in Fig. 2.

In order to expand the MLS deconvolution of overlapping responses to arbitrary stimulus sequences, Van Veen and Lasky (1994) presented a general matrix-based framework working in the time domain. This approach identified an infinite number of recovery sequences for any given stimulus sequences. A similar matrix-based approach was employed by Gutschalk *et al.* (1999) to deconvolve a specially prepared stimulus pattern compiled by a concatenation of click trains for the investigation of human magnetic steady state fields. Recently a similar time domain generalized procedure called continuous loop averaging deconvolution (CLAD) was presented to deconvolve overlapping responses elicited by arbitrary sequences (Özdamar *et al.*, 2003; Delgado and Özdamar, 2004).

A similar deconvolution procedure in the frequency domain was later presented by Jewett *et al.* (2004). Very re-

TABLE II. The characteristic of the first five maximum length sequences.

Code	Base *b*	Number stimuli *N*	Length *L*	Jitter ratio
MLS_3	1	2	3	2:1
MLS_7	2	4	7	3:1
MLS_{15}	3	8	15	4:1
MLS_{31}	4	16	31	5:1
MLS_{63}	5	32	63	6:1

cently we expanded the CLAD procedure to the frequency domain for fast implementation and better understanding of its noise attenuation/amplification and SNR characteristics (Özdamar and Bohórquez, 2006). This latest study presented a general formulation for the signal-to-noise ratio (SNR) of the deconvolution procedure using the frequency domain analysis. Such analyses showed that unlike conventional averaging, deconvolution averaging can attenuate or amplify phase unlocked noise depending on the frequency characteristics of the arbitrary stimulus sequence (Özdamar and Bohórquez, 2006).

In this study frequency analysis of the CLAD procedure will be used to analyze the SNR characteristics of the MLS deconvolution. Quantitative evaluation of the noise attenuation with MLS deconvolution has been difficult and controversial due to the lack of a basic theory and the interpretation of the real recordings due to the involvement of adaptation and muscle artifacts (Picton *et al.*, 1992; Marsh, 1992; Bell *et al.*, 2001, 2002). A simple formula for SNR for MLS deconvolution is derived and evaluated using sample recordings. MLS and its zero-padded variations are shown to present the optimal attenuation for noise yet it is not as effective as the conventional averaging.

II. THEORY AND MATHEMATICAL ANALYSIS

Using CLAD notation (Özdamar and Bohórquez, 2006), let $\{t_n\}$ be a L length sequence of N stimulus positions in a sweep length T acquisition buffer ($T=L$).

$$0 \leqslant t_n \leqslant T-1, \quad n=1,\ldots,N. \tag{3}$$

Let MLS_L (t) be a MLS sequence of N stimuli with length L as related

$$L = 2^{*}N - 1. \tag{4}$$

The sequence itself is the impulse response of a system triggered with the stimuli as expressed below (Özdamar and Bohórquez, 2006)

$$s(t) = \sum_{n=1}^{N} \delta_{t,t_n}. \tag{5}$$

The recorded signal $v(t)$ in response to the stimulus sequence $s(t)$ can be defined using the convolution formula as follows:

$$v(t) = a(t)^{*}s(t), \tag{6}$$

where $a(t)$ is the transient evoked response to a single stimulus and $v(t)$ is the convolved signal.

The response to a single stimulus can be computed in the frequency domain as

$$A(f) = V(f)/S(f) = V(f)G(f), \tag{7}$$

where $G(f)$ is the deconvolution filter frequency response (Özdamar and Bohórquez, 2006) and is the inverse of $S(f)$. The right portion of Fig. 1 shows the typical frequency response in the special case of MLS_{31} sequence ($L=31$). As observed the dc component has a specific value while all the other frequencies show a constant value with a flat response. As described in Özdamar and Bohórquez (2006) values below one in $G(f)$ shows attenuation while values above one shows amplification of noise at the designated frequency.

A. SNR computation of MLS sequences

MLS sequences can be treated as special cases of CLAD sequences. Spectral characteristics of MLSs are well known. The Fourier transform of a MLS sequence impulse response is flat for all the frequencies except the dc component. The value of the frequency response in the general case can be computed with the aid of the Parseval's energy conservation theorem as follows:

The frequency response $S(f)$ of the stimulus function expressed in Eq. (5) is computed using the Fourier transform definition

$$S(f) = \sum_{t=0}^{L-1} s(t)e^{-2\pi jtf/L}. \tag{8}$$

The dc component $S(0)$ can be computed as

$$S(0) = \sum_{t=0}^{L-1} s(t)e^{-2\pi jt0/L} = N. \tag{9}$$

Since the signal energies are the same in the time and frequency domains, the following equation can be written using the well-known Parseval's theorem:

$$\sum_{t=0}^{L-1} s^2(t) = \frac{1}{L}\sum_{f=0}^{L-1} |S(f)|^2 = \frac{1}{L}[S^2(0) + (L-1)S_{\mathrm{ac}}^2(f)]. \tag{10}$$

The term $S_{\mathrm{ac}}(f)$ is the value of the transfer function for all the nonzero (alternating) frequencies (ac). Since the number of ones in the sequence is known, Eq. (10) can be written as

$$\sum_{t=0}^{L-1} s^2(t) = N = \frac{1}{(2N-1)}[N^2 + (2N-2)S_{\mathrm{ac}}^2]. \tag{11}$$

Equation (11) can be reduced to

$$S_{\mathrm{ac}}^2 = \frac{N^2 - N}{2(N-1)}.$$

This can be solved to obtain

$$S_{\mathrm{ac}} = \sqrt{\frac{N}{2}} = \frac{1}{2}\sqrt{L+1}. \tag{12}$$

For the MLS deconvolution filter without conventional averaging the noise is then attenuated by the inverse of Eq. (12). Using the SNR analysis of Özdamar and Bohórquez (2006), the resulting SNR for the MLS with a given N can be expressed as follows:

$$\mathrm{SNR}_{\mathrm{MLS}} = \frac{\sigma_s}{c_{\mathrm{dec}}\sigma_n} = \sqrt{\frac{N}{2}}\frac{\sigma_s}{\sigma_n} = 2\sqrt{L+1}\frac{\sigma_s}{\sigma_n}. \tag{13}$$

The average noise gain factor c_{dec} of the MLS deconvolution filter is

$$c_{\mathrm{dec\ MLS}} = \sqrt{\frac{2}{N}} = 2\sqrt{\frac{1}{L+1}}. \tag{14}$$

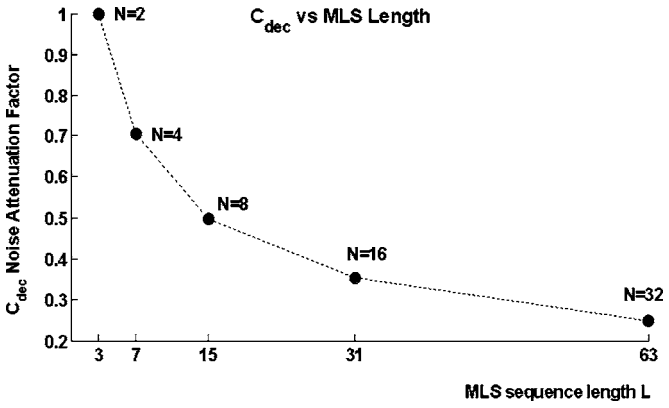


FIG. 3. Change of the noise attenuation factor c_{dec} as a function of the MLS sequence length.

Figure 3 shows the noise attenuation factor of MLS sequences as functions of the sequence length L or the number of stimuli N.

B. Noise attenuation optimality of MLS sequences

The above analysis shows that MLS stimuli have c_{dec} values below one in all frequencies making sure that noise is always attenuated and never amplified. In this section we show that there are no other sequences with the same length and stimulus number that has lower c_{dec} values; that is any arbitrary CLAD sequence with $L=2^{*}N-1$ has a deconvolution average SNR lower than or equal to a MLS sequence.

Let us start with an arbitrary sequence having the same length and stimulus number ($L=2^{*}N-1$). The c_{dec} factor for such a sequence follows the equation (Özdamar and Bohórquez, 2006):

$$c_{\mathrm{dec}}^2 = \frac{1}{L-1}\sum_{f=1}^{L-1}\frac{1}{|S(f)|^2}. \tag{15}$$

We compute the difference of c_{dec}^2 factors between the arbitrary sequence and the MLS sequence (14) to obtain

$$c_{\mathrm{dec}}^2 - c_{\mathrm{dec\ MLS}}^2 = \frac{1}{L-1}\sum_{f=1}^{L-1}\frac{1}{|S(f)|^2} - \frac{2}{N}. \tag{16}$$

Using Eqs. (10) and (11) it can be shown that

$$\frac{1}{L-1}\sum_{f=1}^{L-1}|S(f)|^2 = \frac{N}{2}. \tag{17}$$

If we substitute Eq. (17) in Eq. (16) we obtain

$$c_{\mathrm{dec}}^2 - c_{\mathrm{dec\ MLS}}^2 = \frac{1}{L-1}\sum_{f=1}^{L-1}\frac{1}{|S(f)|^2} - \frac{L-1}{\sum_{f=1}^{L-1}|S(f)|^2}. \tag{18}$$

The right side of Eq. (18) can be interpreted as the difference between the harmonic mean and the inverse of the arithmetic mean of $|S(f)|^2$. Using the arithmetic-harmonic mean inequality we can conclude that

$$c_{\mathrm{dec}}^2 - c_{\mathrm{dec\ MLS}}^2 \geqslant 0 \Rightarrow c_{\mathrm{dec}} \geqslant c_{\mathrm{dec\ MLS}}. \tag{19}$$

Equation (19) means that any other sequence with ($L=2^{*}N-1$) will have a larger or equal c_{dec} than the MLS sequence.

C. Zero padding of MLS sequences

In order to increase the time resolution of the deconvolution, zeros can be padded to the MLS sequences as shown for the MLS_7 example in Table III below. To the sequence $\mathrm{MLS}_{7,0}$, ($N=4$, $L=7$, and $T=7$), when one zero is padded (sequence $\mathrm{MLS}_{7,1}$), N and L stay the same but T increases to 14. The corresponding recovery sequences are shown which have similar patterns.

The effects of zero padding can be ascertained from Fig. 4 where convolution and deconvolution frequency responses are plotted for the MLS sequences listed in Table II. As easily observed, adding a zero adds another frequency cycle to the transfer function. Based on this property, the general formula for the noise attenuation factor of the MLS sequence with length L, N stimulus, and P padded zeros c_{dec} can be expressed as follows:

$$c_{\mathrm{dec}}(\mathrm{MLS}_{N,P}) = \frac{1}{N}\sqrt{\frac{4N^2(P+1)-4N(P+1)+P}{2N(P+1)-(P+2)}}. \tag{20}$$

When $P=0$ this equation reduces to Eq. (14) as expected.

In Fig. 5 noise attenuation factor as a function of P and N are plotted. As seen, the number of padded zeros is important only for small N and P. As the P gets larger ($P \gg 3$) it can be ignored since it does not influence the noise attenuation factor appreciably. For $N=2$ or $N=4$ noise attenuation value falls off significantly when one or two zeros are pad-

TABLE III. MLS_7 and their recovery sequences with increasing zero padding.

Code $L=7$	$N=4$	T	Stimulus sequence
$\mathrm{MLS}_{7,0}$	No 0 padded	7	1100101
$\mathrm{MLS}_{7,1}$	One 0 padded	14	10100000010010
$\mathrm{MLS}_{7,2}$	Two 0 padded	21	100100000000100000100
$\mathrm{MLS}_{7,3}$	Three 0 padded	28	1000100000000000100000001000
MLS Recovery			Recovery sequence
$\mathrm{MLSR}_{7,0}$	No 0 padded	7	1/4 1/4 −1/4 −1/4 1/4 −1/4 1/4
$\mathrm{MLSR}_{7,1}$	One 0 padded	14	1/4 0 1/4 0 −1/4 0 −1/4 0 1/4 0 −1/4 0 1/4 0
$\mathrm{MLSR}_{7,2}$	Two 0 padded	21	1/4 0 0 1/4 0 0 −1/4 0 0 −1/4 0 0 1/4 0 0 −1/4 0 0 1/4 0 0
$\mathrm{MLSR}_{7,3}$	Three 0 padded	28	1/4 0 0 0 1/4 0 0 0 −1/4 0 0 0 −1/4 0 0 0 1/4 0 0 0 −1/4 0 0 0 1/4 0 0 0

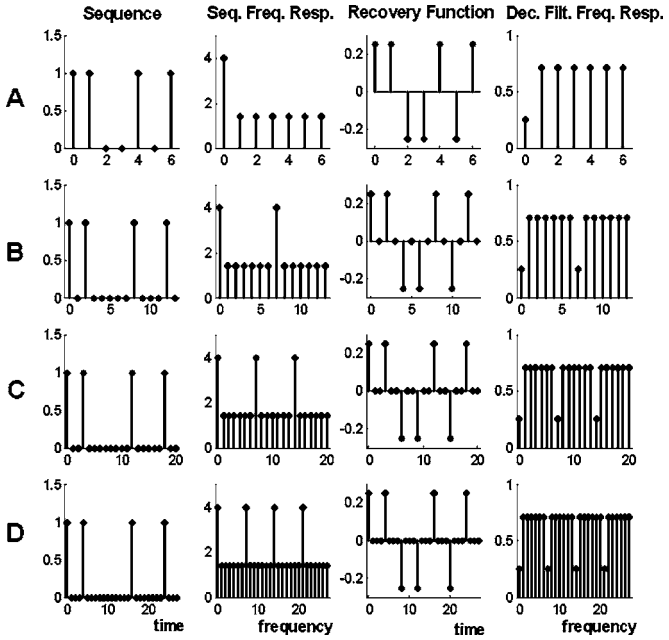


FIG. 4. Time and frequency domain plots of the MLS_7 function with increasing zero padding (a) $P=0$; (b) $P=1$; (c) $P=2$; (d) $P=3$. For each zero padding, first column shows time plots of the stimulus sequence, second column shows the stimulus frequency response, third column shows the recovery sequence, and the fourth column shows the deconvolution filter frequency response.

ded. This, however, become negligible for higher numbers of zero padding. Figure 5 also shows that the number of stimuli is a more important factor in determining the final value of noise attenuation than the number of zeros padded.

The dc value of the deconvolution filter frequency response is $(1/N)$; it is always smaller than c_{dec}. When (P) zeros are padded in time domain, this dc value will appear as notch frequencies in the deconvolution filter frequency response (P) times in equally separated frequencies (Fig. 4). It can be shown that the first notch frequency is

$$f_{nP} = \frac{f_s}{P+1}, \tag{21}$$

where f_s is the sampling frequency. The remaining notches occur at the harmonics of this frequency. This multiple notch filter creation, valid for any MLS or CLAD sequence, can be used to achieve higher noise attenuation factors at specific frequencies.

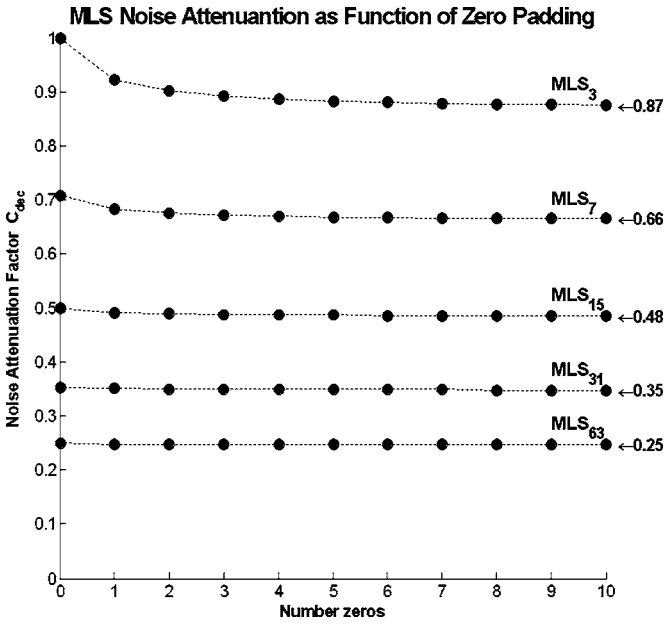


FIG. 5. Noise attenuation factor c_{dec} as a function of the number of padded zeros for the first five MLS sequences listed in Table I. Asymptotic values for infinite zeros padded are marked by arrows on the right.

III. EXPERIMENTAL METHODS

In this study human recordings were obtained to verify the validity of the theory and the equations developed in the previous section. The goal was to demonstrate the characteristics of the noise as a function of the sweeps averaged and deconvoluted. This required the recording of the data in a single sweep mode and performing averaging and deconvolution operations off line. For comparisons with the theoretical results, rms values were computed for estimating noise and signal amplitudes.

Recording parameters were chosen such that auditory brainstem responses (ABRs), middle latency responses (MLRs), and early portions of the late latency responses (LLRs) were simultaneously recorded using monaural stimulation. Scalp continuous electroencephalogram (EEG) recordings were obtained by saving individual sweeps in a digital computer. Recordings were done with 100 000 amplification at 1–1000 Hz (12 dB/oct), sampling frequency was at 20 kHz, and 1500 sweeps were recorded. Scalp differential recordings (noninverting: forehead; inverting: ipsilateral mastoid; common: contralateral mastoid) were stored digitally for later processing and analysis. Four young normal adults with normal hearing and no history of neurological or otological problems served as subjects.

MLS_{15} sequence was used in all the recordings. This sequence had a length of 15 and was padded with 255 or 267 zeros to increase the overall sweep length to $L=3840$ or $L=4020$, respectively. With 50 μs sampling time, sweep times of 192 or 201 ms were obtained. Since each sweep had eight clicks this acquisition format resulted in about 40 Hz average stimulus rate.

Acoustic stimulation was achieved by using a laptop computer which generated the sounds delivered to the subject's ear. The trigger points were recorded in one of the acquisition channels of the recording system. Tone pip stimuli (1 kHz tone pips Blackman, 5 ms duration) were exclusively used as stimuli. They were delivered monaurally to the subject's ear at 92 dB sound pressure level (SPL) using ER-3A insert phones (Etymotic Research, Elk Grove Village, IL). Acoustic calibration was done in a 2cc acoustic coupler using a precision sound level meter (Quest Model 1800) to measure peak dB SPL levels.

IV. RESULTS

An example of single sweep recordings made from a subject using the MLS_{15} sequence is shown in Fig. 6. The time pattern of the pip stimuli and a sample of the raw EEG

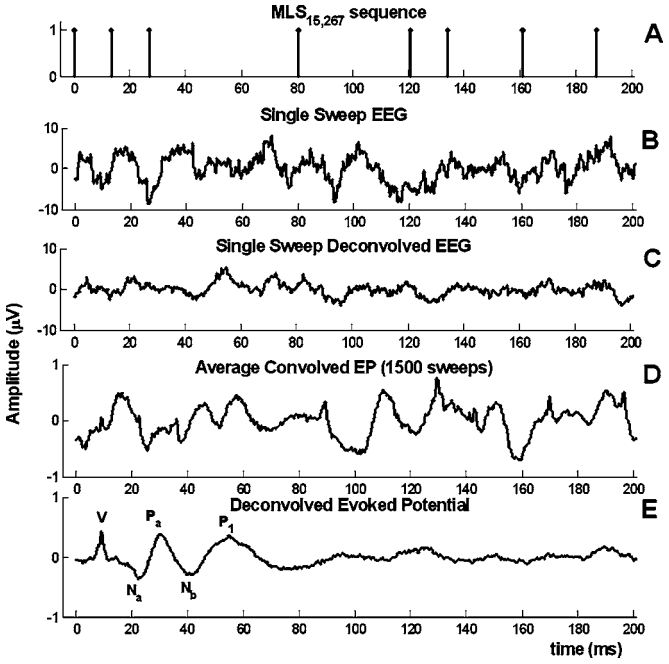


FIG. 6. Results from a recording made in an adult subject using MLS_{15} sequence with 267 zeros (see text for details). (a) Time diagram of the MLS_{15} stimulus sequence, (b) a single sweep EEG sample, (c) single sweep deconvolved EEG of the above trace, (d) averaged convolved EP, and (e) corresponding deconvolved EP. Major peaks (Wave V of ABR, N_a, P_a, N_b of MLR, and P_1 of LLR) are marked as shown.

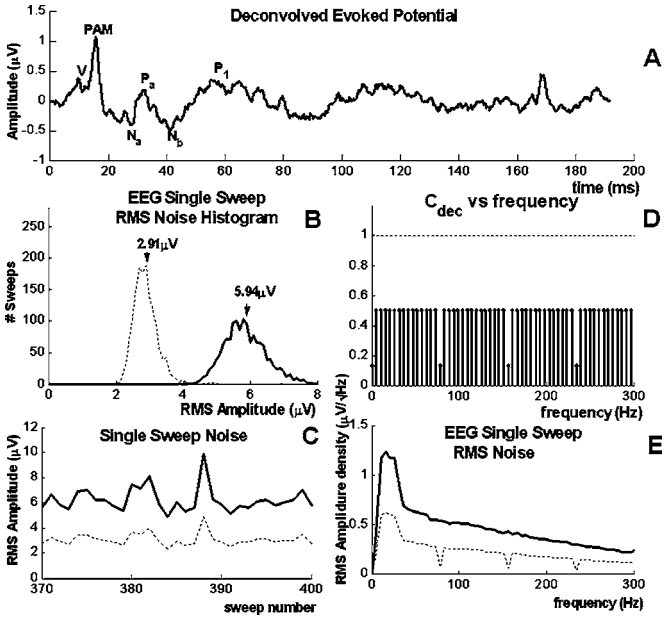


FIG. 7. Statistical and frequency results from another recording from another subject made with another MLS_{15} sequence with 255 zeros (see text for details). (a) Deconvolved EP showing major peaks [Wave V of ABR, post auricular muscle (PAM) response, N_a, P_a, N_b of MLR and P_1 of LLR], (b) RMS single sweep EEG amplitude distribution histograms for raw (solid) and deconvolved EEG (dotted), (c) RMS amplitude of single sweep recordings (370–400) raw EEG (solid) and deconvolved EEG (dotted), (d) deconvolution filter frequency plot for the MLS sequence for the dc-300 Hz region, and (e) RMS densities for each frequency bin of the corresponding frequency region.

are shown in the first two rows [Figs. 6(a) and 6(b)]. The third row shows the corresponding deconvolved EEG according to the stimulus pattern shown at the top. Reduced amplitude of the EEG [Fig. 6(c)] compared to the raw EEG [Fig. 6(b)] can readily be observed in this trace.

In the last two rows average convolved and deconvolved recordings are shown. As expected none of the (AEP) components can be readily recognized in the convolved recording [Fig. 6(d)]. Close scrutiny, however, reveals sharp positive waves corresponding to wave V of the ABR at around 6–7 ms after clicks. The remaining MLR and LLR components cannot be recognized because of heavy overlapping due to short time intervals between clicks. The last trace [Fig. 6(e)] shows a very clean easily recognized AEP recording with all components, wave V of ABR, N_a, P_a, N_b of MLR, and P_1 of LLR. A very small N_1 around 80 ms can also be detected. However, none of the other remaining LLRs (P_2 or N_2) can be observed possibly due to high stimulus rate.

In Fig. 7(a) another sample deconvolved AEP recording from another subject is shown. In this particular recording post auricular muscle (PAM) activity was present as observed in the high amplitude sharp component around 13 ms. As expected, this EP is noisier than the previous recording due to the PAM activity. A rms amplitude histogram of the single sweep convolved and deconvolved EEGs are shown in Fig. 7(b). As expected both distributions have a Gaussian shape with the means (convolved 5.94 μV and deconvolved 2.91 μV) separated by about 3.03 μV. A sample of single sweep rms EEG amplitudes are shown in Fig. 7(c) with a similar reduction between convolved and deconvolved EEG amplitudes.

In Fig. 7(d) the deconvolution filter response function of the MLS_{15} for the low frequency region is plotted. The average noise reduction is about 0.48 and constant for all frequencies other than the repetitive dips with $c_{\mathrm{dec}}=0.125$ at dc, 78.125 Hz and above harmonics due to the padded zeros. RMS EEG amplitudes corresponding to frequency bins are plotted in Fig. 7(e). As expected, EEG amplitudes are higher in low frequencies and monotonically decrease at about 35 Hz. Again the deconvolved EEGs show lower amplitudes than the raw EEG (convolved). The noise attenuation dips seen in the deconvolution filter of Fig. 7(d) are clearly reflected in the deconvolved EEG noise trace.

The recordings from all subjects showed the same results and are depicted in Figs. 6 and 7. EEG noise attenuation was the same as predicted by the theory in averaged as well as single sweep recordings. The results were the same whether the recording contained little or no EMG noise as in Fig. 6 or high PAM or EMG noise as in Fig. 7.

V. DISCUSSIONS

In conventional averaging, SNR is improved with the square root of the number of stimuli averaged. When the MLS method was developed, it was hoped that a similar improvement would be achieved since higher stimulation rates can be achieved. It was suggested that it might be possible to reduce the recording time of EPs such as ABRs by using MLS averaging at high rates (Thornton and Slaven, 1993). The experimental recordings, however, did not show much improvement and noise reduction was less than what is

achieved by conventional averaging (Marsh, 1992). The problem was further complicated by the adaptation effects that come into play with high stimulation rates. In general the SNR improving property of the MLS method is opposed by the signal attenuating property of the physiological adaptation effect.

An early formula to the SNR improvement due to MLS averaging was provided by Thornton (1993a,b) for the acquisition of otoacoustic emissions (OAEs). OAE recording with MLS has the advantage of little adaptation and no peak delay at higher rates; thus, produces more dramatic results. Thornton's proposed formula based on signal and noise improvement expectations is remarkable since it is derived without the use of basic signal and noise equations. If no adaptation is assumed ($k=1$) and percentage of stimuli per sweep ($c=N/L$) is approximated by half, this approach predicts a SNR improvement $1/\sqrt{2}=0.707$ as much as shown in this study. Since little theoretical reasoning for the proposed formula or few quantitative measurements at single sweep level are presented, no assessment of this difference can be offered. These equations are later used to show the feasibility of reducing OAE recording time as tested experimentally under controlled conditions (Thornton, 1993a,b; Hine *et al.*, 2001). In other evoked potential studies recording time of MLR was shown to improve fourfold at a MLS rate of 89/s (Bell *et al.*, 2001).

Another approach to SNR improvement for MLS averaging was provided by Van Veen and Lasky (1994) who presented a general framework for the general arbitrary sequence stimulation. The general framework was then used to develop analytical expressions for relative efficiencies of conventional and MLS averaging. It was concluded that an infinite number of recovery sequences can be identified for a given sequence. The minimization of noise with the recovery functions was analyzed and optimized. The authors' formulation of the problem precluded the use of the frequency domain and interpreting the process as an equivalent filter. Thus, arriving to a simple general solution in the time domain proved difficult to achieve. Such a time domain generalized approach will undoubtedly provide great insight to the analysis of nonlinear effects (Lasky *et al.*, 1998). Linear effects, however, are better analyzed and interpreted in the frequency domain as shown in the present study.

In this study we derived a general formula for the noise attenuation effect of the MLS method based on our frequency domain formulation of the general deconvolution method (Özdamar and Bohórquez, 2006). This analytical approach shows that with the MLS deconvolution the SNR improves with the square root of half of the stimuli [Eq. (13)]. This improvement is less than what is achieved with the conventional averaging but definitely an advantage that can be achieved in addition to the SNR increase due to the sweep averaging. SNR improvement is also influenced by the number of zeros padded when the number of stimuli is small ($N<8$). This improvement, however, becomes negligible as the number of stimuli increase as shown in Fig. 5.

Noise attenuation factors computed by Eq. (20) are consistent with the SNR percentage measured by Marsh (1992) in controlled experimental recordings. For large number of zero paddings, attenuation factors of 0.87, 0.66, and 0.48 are obtained for 2, 4, and 8 stimuli, respectively (see Fig. 5). In experimental recordings Marsh (1992) obtained 0.87, 0.69, and 0.48 for the same number of stimuli which compares very favorably with the theoretical predictions.

Our experimental recordings also confirm this prediction. In Fig. 7 the noise attenuation percentage is 0.49 as computed by dividing the rms value of the raw EEG with that of the deconvolved EEG. This percentage compares very favorably with the calculated value of 0.48 for MLS_{15} sequence with several zeros as shown in Fig. 5.

As shown in our earlier paper, the deconvolution of overlapping responses can be achieved by arbitrarily selected stimulus sequences (Özdamar and Bohórquez, 2006). Depending on their characteristics, deconvolution by such sequences can result in attenuation or amplification of noise in different frequencies. Several sequences can be found that attenuate noise. As proven in Sec. II B, the MLS sequences, however, are unique in their noise elimination since they provide the maximum attenuation possible for a given number of stimuli.

The noise attenuation formulas of the MLS deconvolution show that this method is not as efficient as the conventional averaging method. With large number of zeros padded, only about 70% noise reduction can be achieved with the same number of stimuli used as in conventional averaging. This performance, however, can be degraded with adaptation which depends highly on the rate characteristics of the EP component (Picton *et al.*, 1992; Bell *et al.*, 2001).

MLS deconvolution averaging of AEPs may offer several advantages of decreased data acquisition time and increased test sensitivity to temporal stress on the auditory system at high rates. An additional advantage of the MLS deconvolution is the ability to analyze the nonlinear behavior of the human auditory system and other nonlinear systems, as shown by Shi and Hecox (1991). In prior studies, effective design of stimuli and acquisition paradigms using MLS were hampered by the lack of frequency domain SNR formulation. With the present formulation such studies can be designed quantitatively for a better understanding of the physiological properties of the auditory system considering the differential adaptation effects of different ISIs. In addition to shortening testing time, stress testing and separation of components (especially of event related potentials) at high rates will be important applications in the future.

ACKNOWLEDGMENTS

The assistance of Dr. Marie Cheour, Dr. Tao Wang, and Dr. Erdem Yavuz and the editorial help of Roberta Turner are kindly acknowledged.

Bell, S. L., Allen, R., and Lutman, M. E. (**2001**). "The feasibility of maximum length sequences to reduce acquisition time of the middle latency response," J. Acoust. Soc. Am., **109**, 1073–1081.

Bell, S. L., Allen, R., and Lutman, M. E. (**2002**). "Optimising the acquisition time of the middle latency response using maximum length sequences and chirps," J. Acoust. Soc. Am., **112**, 2065–2073.

Burkard, R., McGee, J., and Walsh, E. J. (**1996a**). "Effects of stimulus rate on the feline brain-stem auditory evoked response during development. I. Peak latencies," J. Acoust. Soc. Am., **100**, 978–990.

Burkard, R., McGee, J., and Walsh, E. J. (**1996b**). "Effects of stimulus rate on the feline brain-stem auditory evoked response during development. II. Peak amplitudes," J. Acoust. Soc. Am., **100**, 991–1002.
Burkard, R., Shi, Y., and Hecox, K. E. (**1990**). "A comparison of maximum length sequences and Legendre sequences for the derivation of brainstem auditory evoked responses at rapid rates of stimulation," J. Acoust. Soc. Am., **87**, 1656–1664.
Chan, F. H. Y., Lam, F. K., Poon, P. W. F., and Du, M. H. (**1992**). "Measurement of human BAERs by the maximum length sequence technique," Med. Biol. Eng. Comput., **30**, 32–40.
Davies, W. D. T. (**1966**). "Generation and properties of maximum-length sequences, parts (1–3)," Control, **10(96)**, 302–304; **10(97)**, 364–365; **10(98)**, 431–433.
Delgado, R., and Özdamar, Ö. (**2004**). "Deconvolution of evoked responses obtained at high stimulus rates," J. Acoust. Soc. Am., **115**, 1242–1251.
De Graauw, T., and Veltman, B. (**1970**). "Pseudo-random binary sequences for multiplex codes," Appl. Opt., **9**, 2658–2660.
Eysholdt, U., and Schreiner, C. (**1982**). "Maximum length sequences—A fast method for measuring brainstem evoked responses," Audiology, **21**, 242–250.
Gutschalk, A., Mase, R., Roth, R., Ille, N., Rupp, A., Hähnel, S., Picton, T. W., and Scherg, M. (**1999**). "Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex," Clin. Neurophysiol., **110**, 856–868.
Hine, J. E., Ho, C., Slaven, A., and Thorton, A. R. D. (**2001**). "Comparison of transient evoked otoacoustic emission thresholds recorded conventionally and using maximum length sequences," Hear. Res., **156**, 104–114.
Jewett, D. L., Caplovitz, G., Baird, B., Trumpis, M., Olson, M. P., and Larson-Prior, L. J. (**2004**). "The use of QSD (q-sequence deconvolution) to recover superposed, transient evoked-responses," Clin. Neurophysiol., **115**, 2754–2775.
Jiang, Z. D., Brossi, D. M., Wang, J., Xu, X., Chen, G. Q., Shao, X. M., and Wilkonson, A. R. (**2003**). "Time course of brainstem pathophysiology during first month in term infants after perinatal asphyxia, revealed by MLS BAER latencies and intervals," Pediatr. Res., **54**, 680–687.
Jirsa, R. (**2001**). "Maximum length sequences-auditory brainstem responses from children with auditory processing disorders," J. Am. Acad. Audiol, **12**, 155–164.
Lasky, R. (**1992**). "Maximum length sequence auditory evoked brainstem responses in human newborns and adults," J. Am. Acad. Audiol, **3**, 383–389.
Lasky, R. E., Van Veen, B. D., and Maier, M. M. (**1998**). "Nonlinear functional modelling of scalp recorded auditory evoked responses to maximum length sequences," Hear. Res., **129**, 133–142.
Leung, S.-M., Slaven, A., Thornton, R. D., and Brickley, G. J. (**1998**). "The use of high stimulus rate auditory brainstem responses in estimation of hearing threshold," Hear. Res., **123**, 201–205.
Marsh, R. (**1992**). "Signal to noise constraints on maximum length sequence auditory brain stem responses," Ear Hear., **13**, 396–400.
Özdamar, Ö., and Bohórquez, J. (**2006**). "Signal to noise ratio and frequency analysis of continuous loop averaging deconvolution (CLAD) of overlapping evoked potentials," J. Acoust. Soc. Am., **119**, 429–438.
Özdamar, Ö., Delgado, R. E., Yavuz, E., Thombre, K. V., and Acikgoz, N. (**2003**). "Deconvolution of auditory evoked potentials obtained at high stimulus rates," *Proc.* 1st *Int. IEEE EMBS Conf. Neural Engineering*, Capri, Italy (IEEE Press, New Jersey), pp. 285–288.
Picton, T. W., Champagne, S. C., and Kellet, A. J. C. (**1992**). "Human auditory evoked potentials using maximum length sequences," Electroencephalogr. Clin. Neurophysiol., **84**, 90–100.
Shi, Y., and Hecox, K. E. (**1991**). "Nonlinear system identification by m-pulse sequences: Application to brainstem auditory evoked responses," IEEE Trans. Biomed. Eng., **BME-38**, 834–845.
Thornton, A. R. D. (**1993a**). "High rate otoacoustic emissions," J. Acoust. Soc. Am., **94**, 132–136.
Thornton, A. R. D. (**1993b**). "Click-evoked otoacoustic emissions: new techniques and applications," Br. J. Audiol., **27**, 109–115.
Thornton, A. R. D., and Slaven, A. (**1993**). "Auditory brainstem responses recorded at fast stimulation rates using maximum length sequences," Br. J. Audiol., **27**, 205–210.
Van Veen, B. D., and Lasky, R. E. (**1994**). "A framework for assessing the relative efficiency of stimulus sequences in evoked response measurements," J. Acoust. Soc. Am., **96**, 2235–2243.

Time-course of the human medial olivocochlear reflex

Bradford C. Backus[a)]
Eaton Peabody Laboratory of Auditory Physiology, Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114
and Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

John J. Guinan, Jr.
Eaton Peabody Laboratory of Auditory Physiology, Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114,
Speech and Hearing Bioscience and Technology, Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139
and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115

(Received 1 September 2005; revised 20 December 2005; accepted 7 January 2006)

The time-course of the human medial olivocochlear reflex (MOCR) was measured via its suppression of stimulus-frequency otoacoustic emissions (SFOAEs) in nine ears. MOCR effects were elicited by contralateral, ipsilateral or bilateral wideband acoustic stimulation. As a first approximation, MOCR effects increased like a saturating exponential with a time constant of 277 ± 62 ms, and decayed exponentially with a time constant of 159 ± 54 ms. However, in ears with the highest signal-to-noise ratios (4/9), onset time constants could be separated into "fast," $\tau = \sim 70$ ms, "medium," $\tau = \sim 330$ ms, and "slow," $\tau = \sim 25$ s components, and there was an overshoot in the decay like an under-damped sinusoid. Both the buildup and decay could be modeled as a second order differential equation and the differences between the buildup and decay could be accounted for by decreasing one coefficient by a factor of 2. The reflex onset and offset delays were both ~25 ms. Although changing elicitor level over a 20 dB SPL range produced a consistent systematic change in response amplitude, the time course did not show a consistent dependence on elictor level, nor did the time-courses of ipsilaterally, contralaterally, and bilaterally activated MOCR responses differ significantly. Given the MOCR's time-course, it is best suited to operate on acoustic changes that persist for 100's of milliseconds.
 [DOI: 10.1121/1.2169918]

PACS number(s): 43.64.Ri, 43.64.Jb, 43.64.Tk [BLM] Pages: 2889–2904

I. INTRODUCTION

The medial olivocochlear reflex (MOCR) is one of several sound-activated feedback circuits that control the input to the mammalian hearing system (for a review see Guinan, 1996). Activating medial olivocochlear (MOC) fibers reduces the amplitude of basilar membrane motion in response to low-level sound by 10–20 dB at the characteristic frequency (CF) (Murugasu and Russell, 1996, Dolan *et al.*, 1997, Cooper and Guinan, 2003). Despite general knowledge about how the MOCR changes cochlear and neural responses to sound, a fundamental question remains, "What use is the MOCR?" Understanding how fast the MOCR produces its effects can suggest auditory processing that is affected by the MOCR; and quantification of the time constants that describe the MOCR time-course can help determine which cellular processes are involved by focusing attention on those processes with similar time constants.

The discovery that otoacoustic emissions (OAEs) (Kemp, 1978) are modulated by MOC activation (Mountain, 1980, Siegel and Kim, 1982) has led to the investigation of the MOCR in humans via its effects on OAEs. Several investigations have used OAE-based measurements to pursue the MOCR time-course (Kim *et al.*, 2001, Maison *et al.*, 2001a, Bassim *et al.*, 2003). However, interpreting the very different results from these transient and distortion-product otoacoustic emission (TEOAE and DPOAE) studies is complicated by intra-cochlear interactions such as two-tone suppression and by efferent activation by the test stimuli.

One way to avoid these complications is to use stimulus-frequency otoacoustic emissions (SFOAEs). SFOAE-based methods require only one low-level probe-tone thereby avoiding multiple frequency interactions, providing the most frequency-specific measure, and reducing the likelihood that the probe itself will elicit efferent activity (Guinan *et al.*, 2003).

We used an SFOAE-based method to quantitatively answer five questions: (1) How rapidly does the MOCR produce its effects in the cochlea? (2) How rapidly do those effects decay? (3) Is there a relationship between the buildup and decay? (4) Is there any difference in the time-course when more intense elicitor stimuli are used? (5) Is there any difference in the time-course between the ipsilaterally, contralaterally, and bilaterally elicited reflexes?

[a)]Current affiliation: UCL Ear Institute, 332-336 Gray's Inn Road, London WC1X 8EE, England.
Electronic mail: bradford.c.backus.95@alum.dartmouth.org

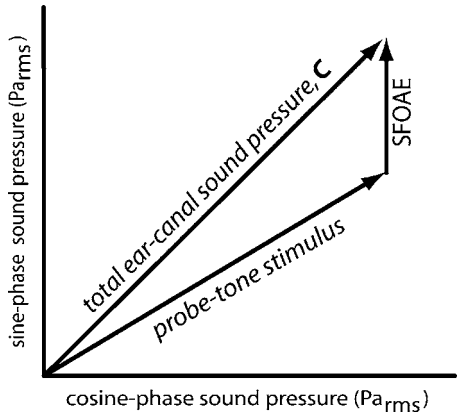


FIG. 1. Vector diagram representing the summation in the ear canal (EC) of two tonal components that differ in magnitude (length of arrows) and phase (direction of arrows). A tone stimulus (probe-tone) and the SFOAE it generates add to produce a compound-tone, **C**, that is measured as the EC sound pressure at the probe-tone frequency. If the probe-tone component and middle-ear transmission are invariant, any change in the measured EC sound pressure reflects a change in the SFOAE magnitude and/or phase.

II. METHODS

A. Overview

A "probe-tone" was played into the ear canal in order to generate an SFOAE from within the cochlea. This SFOAE combined with the probe-tone in the ear canal to produce a complex-valued compound-tone, $\mathbf{C}(t)$, that is the ear canal sound pressure at the probe-tone frequency (Fig. 1). If the acoustic probe-tone and middle-ear transmissions are invariant, e.g., if there are no middle-ear-muscle (MEM) contractions, changes in the compound-tone can be wholly attributed to changes in the SFOAE. Since activation of medial-olivocochlear (MOC) fibers reduces the sound-induced motion of the basilar membrane (BM) (Murugasu and Russell, 1996, Dolan *et al.*, 1997, Cooper and Guinan, 2003) it is likely that part or all of the MOCR inhibition of SFOAEs is caused by the reduction of BM motion.

We used a continuous (40 dB SPL) bilateral probe-tone[1] near 1 kHz (within 12%) to produce an SFOAE and to thereby create a compound-tone. A heterodyne technique (Kim *et al.*, 2001, Guinan *et al.*, 2003) was used to extract the magnitude and phase of the compound-tone across time, $\mathbf{C}(t)$, while a second acoustic stimulus, the elicitor, was introduced to activate the MOCR and thereby alter the SFOAE and the measured $\mathbf{C}(t)$. Changes in the compound tone, $\Delta\mathbf{C}(t)$, were calculated by vector subtracting the average pre-elicitor $\mathbf{C}(t)$—the average in a time window from 450 to 50 ms prior to elicitor onset—from each time point in $\mathbf{C}(t)$. The result, $\Delta\mathbf{C}(t)$, quantifies changes from the no-elicitor condition (for details see Guinan *et al.*, 2003).

We strove to keep the probe-tone stimulus invariant by using elicitor levels below middle-ear-muscle reflex (MEMR) thresholds and by asking subjects to be as still as possible during measurements. To test whether our elicitor levels ($\leqslant$60 dB SPL) were below MEMR thresholds, we performed "phase-slope MEMR tests" in which we calculated the phase-slope of $\Delta\mathbf{C}$'s produced by a 60 dB SPL wide-band noise elicitor (for details see Guinan *et al.*, 2003). No ears had measurable MEMR effects at elicitor levels of 60 dB SPL.

Since SFOAE amplitudes vary widely across frequencies and ears, our measurements of MOCR effects are reported in units relative to an estimate of the SFOAE amplitude. In other words, our MOCR measurements are *normalized* and are reported as "Δ**SFOAE**(t)" with normalized units of $\%|\mathbf{SFOAE}|$ (for detail see Appendix A). Without normalization one cannot compare the magnitude of MOCR effects on SFOAEs across ears and/or frequencies as the widely varying SFOAE amplitudes will be reflected in the responses.

B. MOCR elicitor stimuli

Two and a half second 60 dB SPL wide-band (100 Hz–10 kHz) noise bursts repeated every 5 s were used to activate the MOCR. Because wide-band noise induces two-tone suppression effects when presented into the measurement ear, we only used contralateral wide-band noise elicitors in the first set of experiments. When ipsilateral and bilateral activation of the MOCR was required (for looking at the ipsilateral and bilateral MOCR time-course) we used notched-noise elicitors. For these, a 2.1 octave spectral notch centered at the probe-tone frequency was used to remove energy from the wide-band noise elicitor at frequencies that produced measurable two-tone suppression effects (details in Appendix D).

C. Subjects and screening

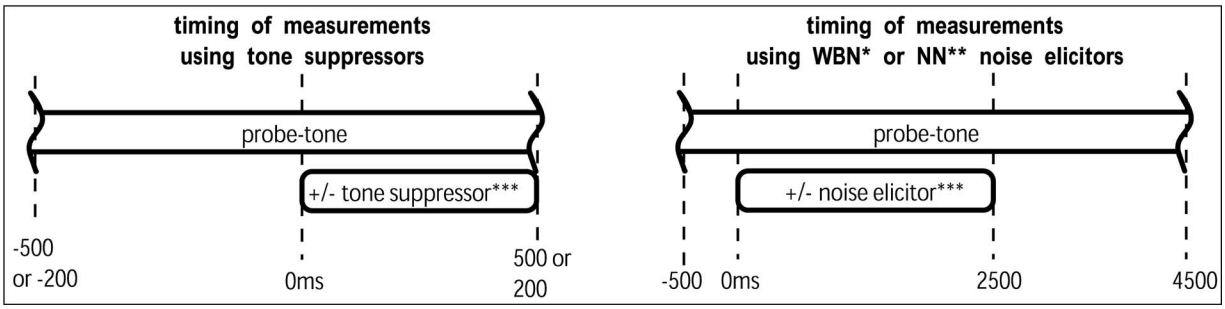
Of 28 subjects screened for the study, nine young adult subjects, 19–30 years of age, were chosen to participate (8 female, 1 male). Participation was contingent on passing four tests: (1) A two-interval forced choice audiogram using 1/3 octave band noise bursts centered at 250 Hz, 500 Hz, 1 kHz, 2 kHz, and 4 kHz was used to insure subjects had normal hearing in both ears (within 20 dB re: ANSI pure tone thresholds); (2) measurements of spontaneous emissions in quiet were used to insure that the probe-tone was at least 50 Hz from any spontaneous emission[2] greater than −5 dB SPL; (3) a MEMR phase-slope test was used to insure that MOCR effects dominated the measured changes in the compound-tone; and (4) a signal to noise ratio (SNR) test[3] was applied to the results and was used to select those subjects for whom turning on the elicitor noise produced a clear, strong and consistent $\Delta\mathbf{C}(t)$ signal.

D. Choosing a probe-tone frequency

Selecting ear/probe-tone-frequency combinations that produced large SFOAEs and thereby high SNR measurements was important for looking at details of the MOCR time-course. For this reason, each subject was initially measured at 11 or 12 probe-frequencies (20 Hz steps from 900 to 1100 Hz or from 900 to 1120 Hz); and then the ear/probe-tone-frequency combination that yielded the largest $\Delta\mathbf{C}(t)$ response was selected for further measurements. Initial measurements used 40 dB SPL continuous bilateral probe-tones and 60 dB SPL bilateral wideband noise elicitors

TABLE I. Stimulus types (top panels) and stimulus presentation timings (bottom panel) for all measurements. To form a measurement, an even number (44-816 depending on SNR and subject availablity) of presentations were made and the acoustic responses were averaged.

measurement purpose	elicitor or suppressor stimuli: Type	Level (dBSPL)	Freq.(s) (Hz)	Dur. (ms)	probe-tone (all are bilateral and continuous at 40 dB SPL)
supporting measurements					
1. SFOAE estimate	ipsi tone supressor	60	-110Hz re: probe freq.	500 or 200	within 12 % of 1 kHz for all tests
2. suppression bandwidth	ipsi tone supressor	60	-1.5 oct. +1.5 oct. re: probe freq.	200	
3. middle ear muscle test	contra or bilateral WBN* elicitor	60	100Hz to 10kHz	2500	5 or more freqs. in 20Hz spacing
MOCR measurements					
5. 'best frequency' search (bilateral efferent measure)	bilateral WBN* elicitor	60	100Hz to 10kHz	2500	11 or more freqs. in 20Hz spacing around 1 kHz
6. contralateral efferent time-course	contra WBN* elicitor	60	100Hz to 10kHz	2500	
7. Ipsilateral, contralateral and bilateral time courses	ipsi, contra, and bilateral NN** elicitor	60	100Hz to 10kHz w/2.1 oct. notch	2500	within 12 % of 1kHz for all tests
8. contralateral efferent time-course vs. level	contra WBN* elicitor	45,50,55,60	100Hz to 10kHz	2500	



* Wide band noise, (0.1—10 kHz and acoustically flat).
** Notched Noise (a WBN with a 2.1 octave square spectral notch centered at the probe frequency).
*** With raised cosine ramps (5 ms rise/fall) and alternating polarity on successive trails.

(2500 ms in duration). The response was taken to be the average $|\Delta \mathbf{C}(t)|$ in a post elicitor window, 50 to 150 ms after elicitor cessation.

E. Measurement procedure

Subjects were comfortably seated in a sound-dampening chamber during 30 min measuring sessions that were interleaved with 15 min breaks. Some subjects took several hours to measure and were measured over multiple days.

In every session, two earphones (Etymotic Research ER10c) each containing two sound sources and one microphone were fitted in the subject's ear canals, one in each ear. To eliminate potential distortion, one source from each ER10c was allocated exclusively to the continuous probe-tone stimulus while the second source was used for the elicitor (or suppressor) stimulus. Calibrations were done at the beginning, during, and at the end of every measuring session to insure proper sound pressure levels for all stimuli. Stimuli were created digitally (sampling rate of 20 kHz) and stimulus levels were set by attenuators (0.1 dB accuracy). A raised cosine ramp (5 ms rise and fall) reduced spectral splatter for the pulsed stimuli. Microphone signals were digitized using the same 20 kHz sample timer used for stimulus generation.

Elicitor and suppressor stimuli were adjusted to suit each measurement type (see Table I). For instance, because MOCR effects develop slowly compared to two-tone suppression effects, the stimulus durations for noise elicitors used in MOCR measurements (and also in the MEMR test) were 2500 ms while the stimulus durations for suppressor-tone stimuli used for SFOAE estimation were either 500 ms or 200 ms. MOCR elicitors were 60 dB SPL wideband (100 Hz–10 kHz) noise except during (a) level-dependence experiments where 40, 45, 50, 55, and 60 dB SPL elicitors were used and (b) laterality-dependence experiments where the elicitor was a 60 dB SPL wideband noise with a 2.1 octave spectral notch centered at the probe-tone frequency. Suppressor-tone stimuli were 60 dB SPL tones placed −110 Hz below the probe-tone frequency (except during suppression bandwidth measurements for which the suppressor-tone frequency was varied).

Every measurement consisted of an average of an even number of presentations. Elicitor (or suppressor) polarities

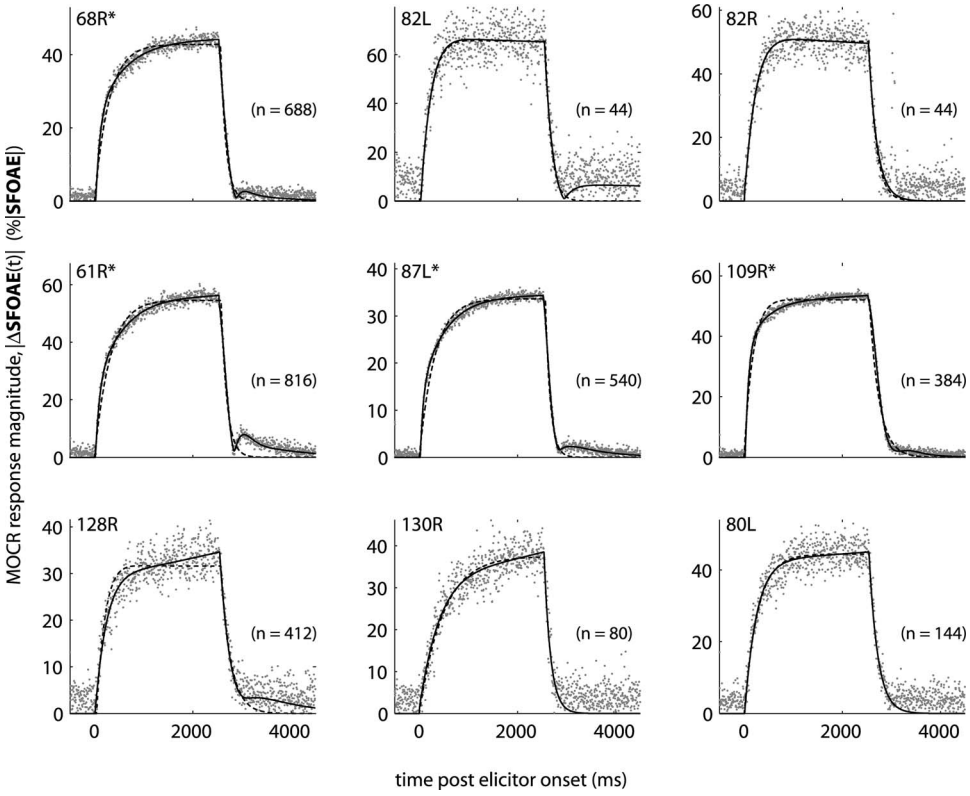


FIG. 2. MOCR time-courses for 9 ears (no special order). Responses were evoked by contralateral wideband noise (60 dB SPL). Repetition interval was 5 s and n is the number of responses that were averaged for each measurement. Points are data samples (5 ms resolution), dashed curves are single exponential characterizations, and solid curves are sum of exponentials characterizations of the data. The four subjects with *'s are those with the four highest SNRs.

were alternated on successive presentations to cancel their acoustic contributions to the averaged EC sound pressure. This manipulation did not appear to have any effect on the emission.

F. Mathematical characterizations of the responses

Each MOCR response waveform, $\Delta\mathbf{SFOAE}(t)$, was characterized mathematically to quantify its time-course. We sought to characterize each waveform's rising curvature, decaying curvature, onset delay, and offset delay (for equations and details see Appendix B).

1. "Single complex exponential" characterization

Our first characterization used a single complex saturating exponential term to represent the rising curvature, a single complex decaying exponential term to represent the decaying curvature, and included onset and offset delay parameters (the amount of time between elicitor onset or offset and a change in the response). This "single complex exponential" characterization provided reasonable approximations to the curvatures and was useful for the purpose of assigning an *overall onset time constant* and an *overall decay time constant* to the rising and decaying portions of the waveforms, respectively.

2. "Sum of complex exponentials" characterization

Although the single complex exponential characterizations captured the basic curvatures of the waveforms, subjects with high SNR measurements produced curvatures that differed visibly from their single complex exponential characterization (e.g., a bounce in the decay of subject 61R in Fig. 2). To describe these waveforms better, a more complicated mathematical characterization, using a sum of complex exponential terms, was made (for details see Appendix B). These "sum of exponentials" characterizations better approximate the waveforms and they were used to quantify the onset and offset delays; and when the SNR was large enough, they were used to extract and quantify multiple time constants.

III. RESULTS

A. The contralateral MOCR time-course

Both the amplitude and phase of SFOAEs were affected by MOCR activation; however, since most important aspects of the time-course can by shown by the amplitude alone, this paper focuses on amplitude changes. The MOCR manifested its effects on the amplitude of SFOAEs over 100's of milliseconds and produced onset rises and offset decays that looked roughly like saturating and decaying exponentials (Fig. 2).

1. MOCR overall onset and decay time constants

We assigned a single overall time constant to the rising and falling portion of each MOCR response (associated curves have dashed lines in Fig. 2). Estimates of overall onset and decay time constants varied from ear to ear, spanning 178–400 ms for the onset and 104–259 ms for the decay. For 7 of 9 ears the decay was more rapid than its buildup (Fig. 3). Only ear 109R showed the opposite trend significantly with an overall onset time constant of 178±3.2 ms, 35 ms faster than its overall decay time constant (213±3.0 ms). There was an apparent but not statistically

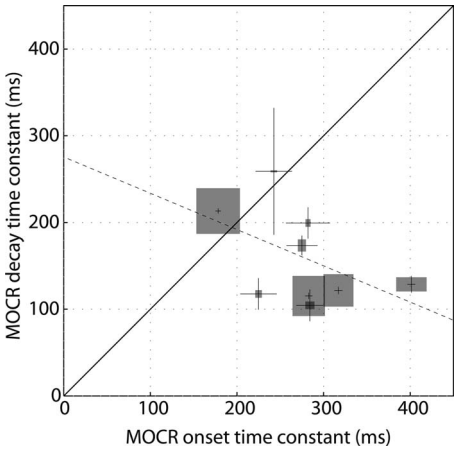


FIG. 3. MOCR overall onset vs. decay time constants for 9 ears show that MOCR effects generally buildup more slowly than they decay. Error bars =1 SD (for details see Appendix B3). To emphasize more accurate data points, i.e., those with small error bars, rectangles, whose areas represent relative accuracy, are centered on each data point (bigger area means more accurate).There was an apparent inverse relationship between the speed of the buildup of MOCR effects and the speed of the decay of those effects (dashed line, $R=-0.48$, $P=0.15$).

significant inverse relationship between the speed of the MOCR's onset and decay (as shown by the dashed line in Fig. 3, $R=-0.48$, $P=0.15$).

2. MOCR effects can decay with an overshoot

While the rising phase of each response was monotonic and appeared to approach a steady state like a saturating exponential, the decaying portion in 6 of 9 ears exhibited a "bounce" that defied exponential decay (see Fig. 2). This bounce led us to pursue a more complicated sum of complex exponentials characterization.

With a 3 term sum of complex exponentials we were able to mathematically characterize the bounce. The decay curvatures of the 4 highest SNR subjects (61R, 87L, 68R, and 109R) were well approximated by the combination of an under-damped sinusoid and a "long"[4] decay time constant (see Appendix C) indicating that the observed bounces were in fact an overshoot. In other words, as the SFOAE re-equilibrates to its pre-MOCR activation level (upon elicitor cessation), it can overshoot that level. Phase data (not shown) corroborates the overshoot.

3. The MOCR overall onset time constant is primarily a combination of "fast" and "medium" underlying time constants

The sum of exponentials characterization was also a better approximation of the MOCR's onset response, as evidenced by higher goodness-of-fit values, R^2. For the four highest SNR ears, the MOCR onset time constants (Table II) were found to be composed of a mixture of fast, ~70 ms, medium, ~330 ms, and slow, ~25 s, time constants with the bulk (88%–93%) of the curvatures determined by the fast and medium time constants. As shown in Table II, ears 61R and 109R differed widely in their overall onset time constants (317±4.59 ms versus 179±3.2 ms), yet their fast and medium underlying onset time constants were similar (69±31 ms versus 65±10 ms and 331±106 ms versus 294±68 ms). For these subjects, it was a different mix of fast and medium underlying onset time constants that produced their different overall onset time constants. Subjects 61R, 68R, 87L, and 109R all had sufficient SNR to allow the extraction of fast, medium, and slow onset time constants. All generated qualitatively consistent slow onset time constants (10s of seconds) and quantitatively similar fast (~70 ms) and medium (~330 ms) onset time constants (Table II).

Slow onset time constants could be extracted even though the measurements recorded only 2.5 s of onset and thus were not geared to detecting slow onset time constants. Values for the slow onset time constant ranged from 11 to 39 s. It was not possible to put an upper bound on all these estimates because the relatively brief MOCR elicitor,

TABLE II. Onset and decay time constants and their % contributions to contralateral MOCR time courses for 9 ears. Subjects 61R, 68R, 87L, and 109R had sufficiently large SNR to extract multiple time constants for the rising curvature. Numbers on the left are time constants in ms and the numbers on the right in parenthesis are their relative % contribution to the time course. Low SNR subjects (80L, 82L, 82R, 128R, and 130R) could not distinguish "fast" time constants from "medium" ones. MOCR decay time constants could not be well approximated by a mixture of underlying time constants, but did include a "long" decay time constant. "±" are the estimated standard deviations (see Appendix B).

Ear	Fast onset time constant (ms) and (%)	Medium onset time constant (ms) and (%)	Slow onset time constant (s) and (%)	Overall onset time constant (ms)	Overall decay time constant (ms)	Long decay time constant (ms) and (%)
61R*	69±31 (27%±12%)	331±106 (61%±11%)	~18 (12%±4%)	317±5	122±5	918±97 (14%±1%)
68R*	79±30 (43%±16%)	316±175 (49%±15%)	~23 (8%±3%)	283±5	115±4	594±254 (29%±14%)
87L*	67±28 (31%±11%)	348±99 (61%±9%)	~27 (8%±3%)	284±16	104±18	615±150 (21%±6%)
109R*	65±10 (58%±7%)	294±68 (35%±7%)	~39 (7%±2%)	179±4	213±3	449±52 (27%±6%)
	undifferentiable					
80L	235±24 (88%±5%)		~37 (12%±5%)	275±18	173±12	unmeasureable
82L	197±23 (97%±3%)		unmeasureable	225±21	118±12	unmeasureable
82R	229±14 (98%±1%)		unmeasureable	243±21	259±73	unmeasureable
128R	223±18 (84%±2%)		~14 (16%±2%)	282±26	199±18	unmeasureable
130R	404±50 (79%±5%)		~11 (21%±5%)	401±5	129±10	unmeasureable

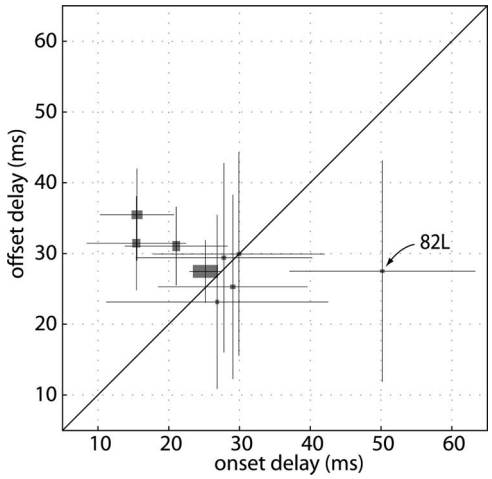


FIG. 4. MOCR onset vs offset delays for 9 ears show that both are near 25 ms. Error bars=1 SD. To emphasize more accurate data points, i.e., those with small error bars, rectangles, whose areas represent relative accuracy, are centered on each data point (bigger area means more accurate). Subject 82L's large onset delay value is probably a result of this subject's response not completely decaying to zero between presentation intervals, see Fig. 2.

2.5 s, did not allow us to sample enough of the curvatures. It was possible, however, to put a lower bound on the estimates.[5] Ear 130R produced a slow time constant >5.9 s (with 95% confidence) and all other ears produced ones >10.0 s. The average slow onset time constant estimate across ears was 24 s.

4. MOCR onset and offset delays are both near 25 ms

Onset and offset delays were calculated from the sum of complex exponentials characterizations (see Appendix B) and were generally between 15 and 40 ms (Fig. 4). Ear 82L produced a 50 ms onset delay but this subject had a very low SNR and was the only subject for which the response did not entirely decay to zero, which complicated detecting the response onset. In summary, the MOCR began to respond to an acoustic stimulus within 10's of milliseconds, and began to shut off just as quickly. Both the onset and offset delay were near 25 ms.

5. A second-order model can describe the MOCR time-course

By comparing the parameters from the sum of complex exponentials characterization with similar parameters from a simple second order system, it was found that a simple second-order model could largely account for the rising and decaying time-courses of the responses. Furthermore, the difference between the rising and decaying time-courses could be re-produced by changing the "damping factor" (i.e., the coefficient of the first derivative term in the model) by a factor of 2 while keeping all other model elements constant (see Appendix C).

B. MOCR time-course, elicitor level dependence

1. No consistent systematic dependence of MOCR time-course on elicitor level was observed

To investigate whether the MOCR time-course is affected by elicitor sound level, a second experiment was carried out on 4 of the original 9 ears using contralateral wideband noise elicitors at sound levels of 40, 45, 50, 55, and 60 dB SPL. Ears, 80L and 68R did not initially produce a measurable response at 40 dB SPL and further attempts to measure this level on those ears were abandoned.

Waveform curvatures were similar for the elicitor levels tested; and there was no obvious consistent and systematic effect of elictor level on the time course across subjects (Fig. 5). Ear 61R's overshoot was larger for larger elicitor levels both in absolute value and as a percentage of the maximum magnitude. In contrast, subject 68R had a constant percentage overshoot (8.6%). Neither the delays nor the MOCR time constants showed statistically significant systematic dependence on elicitor level (Fig. 6).

2. MOCR growth function

MOCR response amplitude increased in proportion to elicitor level almost linearly at $\sim 2\%$ per dB SPL (Fig. 7), over the elicitor range we sampled. Correlation coefficients (R values in Fig. 7) from linear least squares regression for 3 of the 4 ears strongly suggested a linear relationship. Ear 61R had a slightly lower correlation coefficient. At the highest level, 61R produced the largest response of the group, one that departed from the linear trend. This could be an early sign of reflex saturation; perhaps for this ear the 60 dB SPL elicitor was beginning to maximally stimulate the reflex or perhaps at 60 dB SPL that particular SFOAE was becoming suppressed as much as could be measured. Using higher level elicitors to determine the saturation point was not possible because of confounding middle-ear-muscle activation.

C. Time-courses of the ipsilaterally, contralaterally, and bilaterally activated MOCR

To investigate the differences between ipsilaterally, contralaterally, and bilaterally activated MOCR time-courses, a third experiment was carried out on 3 of the original 9 ears. For this experiment, elicitors were presented ipsilaterally (re: the measurement ear), contralaterally, or bilaterally. Since activating the MOCR ipsilaterally or bilaterally requires that the elicitor be presented together with the probe-tone, we developed a stimulus that would minimize interactions (e.g., two-tone suppression) that caused non-MOCR mediated changes in the SFOAE. A "notched-noise" elicitor was chosen. The notch, a 2.1 octave spectral notch centered at the probe-tone frequency, was designed to remove all energy at those frequencies that produced measurable two-tone suppression effects (Appendix D).

1. No systematic dependence of MOCR time-course on elicitor laterality

Ipsilaterally, contralaterally, and bilaterally activated MOCR responses produced waveform curvatures (Fig. 8) for which neither the delays nor the single time constants

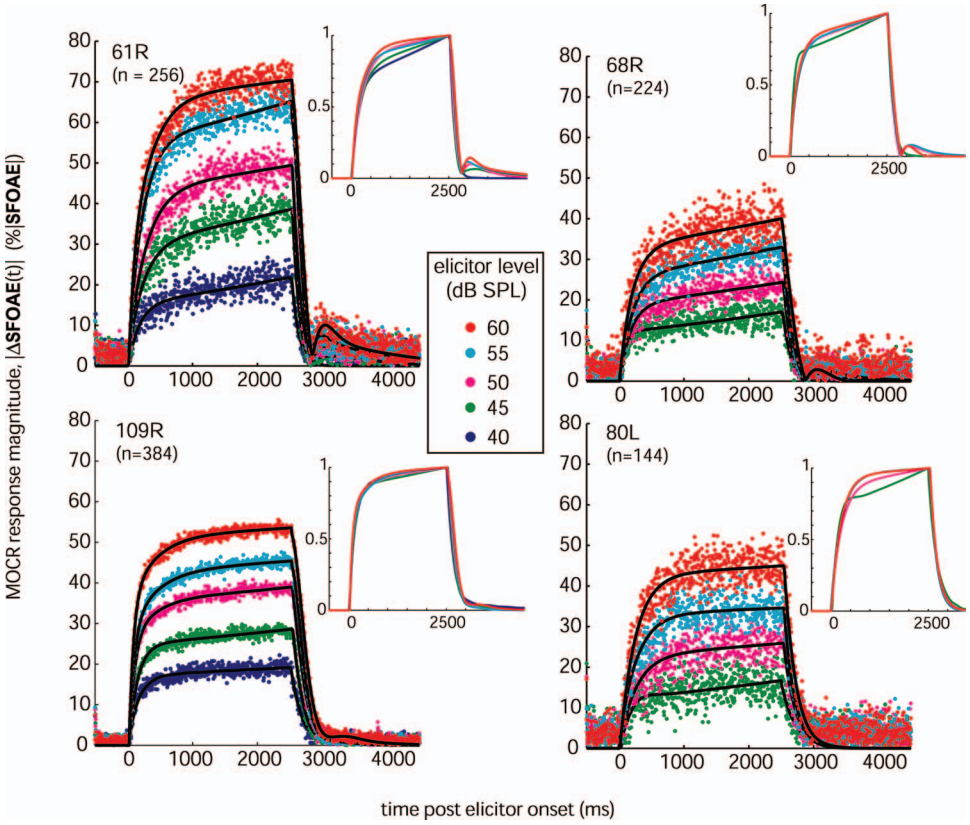


FIG. 5. MOCR responses for wideband contralateral elicitors at various elicitor levels show that while response amplitudes changed in a consistent and systematic way with elicitor level the time courses were qualitatively similar across levels (four subjects). Repetition interval was 5 s and n is the number of responses that were averaged for each measurement. Elicitor was on at 0 ms and off at 2500 ms. Solid lines are sum of complex exponentials characterizations of the waveforms. Insets show the characterizations normalized and overlaid to allow a comparison of time course vs. elicitor level (coded by color).

showed any systematic dependence on elicitor laterality (Fig. 9). Subjects 68R and 109R produced greater responses during contralateral activation than during ipsilateral activation; the opposite was true for subject 82L.

2. MOCR can exhibit binaural interaction

How do the ipsilateral and contralateral limbs of the MOCR combine when activated together? One hypothesis is that the two effects add. To test this, we added the three-term characterizations of the ipsilaterally and contralaterally elicited responses (complex addition at each time point) and compared the result with bilaterally elicited responses. Two of three ears (68R and 82L) agreed with the addition hypothesis. Ear 109R, however, departed significantly from the hypothesis producing a lower bilateral response (by a factor of 0.8) and a greater phase (+20 degrees, not shown) than would be predicted by simple addition. Although the magnitude of 109R's response was not predicted by the simple addition hypothesis, the "shape" of this response on the complex plane was predicted by the simple addition hypothesis, and this was true for all three subjects.

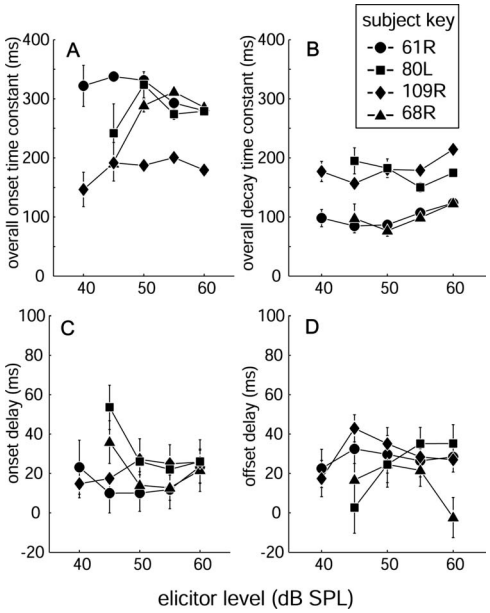


FIG. 6. MOCR overall onset and decay time constants (panels A and B) and onset and offset delays (panels C and D) vs contralateral elicitor level for four subjects show no consistent systematic dependence on elicitor level. Error bars=1 SD.

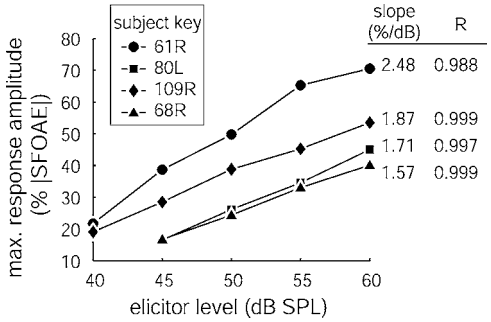


FIG. 7. Growth of MOCR, response amplitude vs. contralateral elicitor level shows an almost linear growth (correlation coefficient, R, from linear least squares regression, is close to 1). Standard deviations are smaller than the symbols.

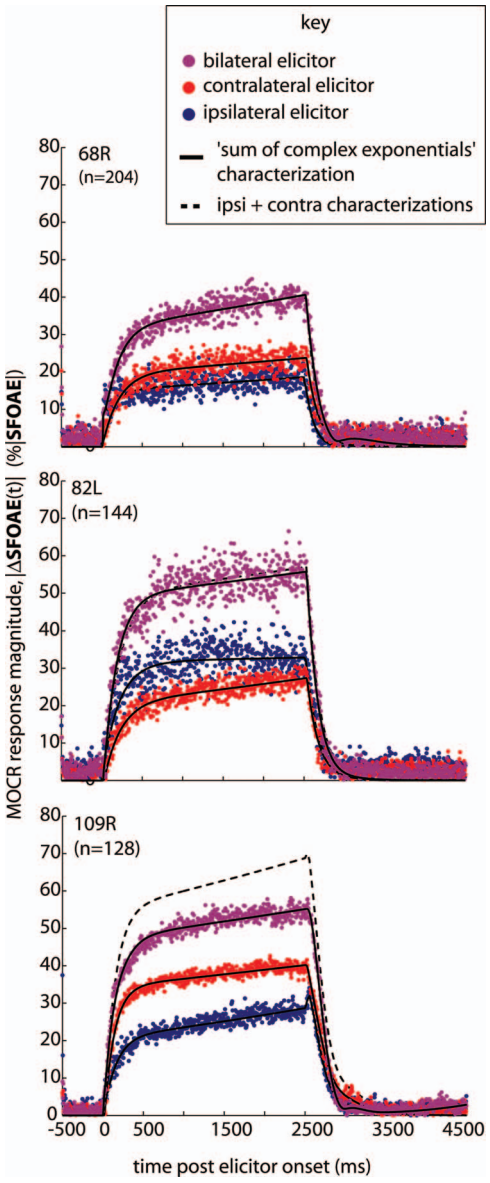


FIG. 8. MOCR response waveforms for notched-noise elicitors presented ipsilaterally, contralaterally, and bilaterally in three subjects show that the waveforms were similar across laterality. Complex addition of the ipsilaterally and contralaterally activated responses (dashed lines) predicted the bilaterally activated response for subjects 68R and 82L, but not for subject 109R. Elicitor was on between 0 and 2500 ms, repetition interval was 5 s and n is the number of responses that were averaged for each measurement.

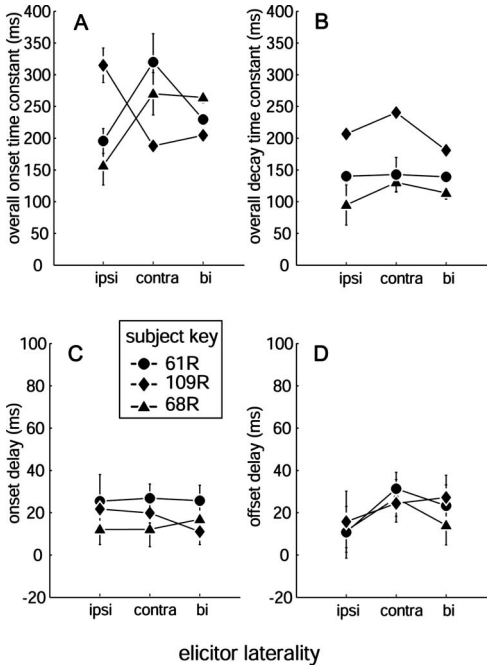


FIG. 9. MOCR overall onset and decay time constants (panels A and B) and onset and offset delays (panels C and D) vs elicitor laterality for three subjects show no systematic dependence on elicitor laterality.

IV. DISCUSSION

A. MOCR time constants

1. Onset time constants compared with previous work

We observed similar MOCR time-courses across all ears including similar, but not identical, overall onset (277±62) and decay (159±54 ms) time constants. The subset of ears that provided high SNR (4 of 9 ears) enabled us to probe the details of their time-courses which revealed (a) the existence of 3 underlying onset time constants, τ_{fast}=60–80 ms, τ_{medium}=290–350 ms, and τ_{slow}=10s of seconds, and (b) that the responses decayed like under-damped sinusoids with an overshoot—the under-damped sinusoid accounted for >70% of the curvatures—and also included an exponentially decaying component with a long time constant, τ_{long} =400–900 ms.

Previous data on the human MOCR time-course come from a variety of OAE measurements. Early data are accessible from a study investigating changes in spontaneous otoacoustic emissions (SOAEs) due to contralateral acoustic stimulation (CAS) (Mott *et al.*, 1989). Here CAS-induced SOAE frequency shifts (which were more salient than concurrent amplitude changes) required 40–200 ms to reach 25% of their final value, corresponding to onset time constants of 139–695 ms, and consistent with our overall onset time constant values (Table II).

Another MOCR time-course measure comes from the effects of CAS (30 dB SL, 80 or 244 ms duration wideband noise) on transiently evoked otoacoustic emissions (TEOAEs) (Maison *et al.*, 2001b). In this study, MOCR effects appeared to grow roughly linearly from 0 to 60 ms post CAS onset and then remain almost constant thereafter. This time-course is different from that found by other studies. The discrepancy may be due to limitations of the experimental method and the fact that the 60 dB SPL click stimulus used has been shown to also be an effective MOCR elicitor (Guinan *et al.*, 2003).

The most extensive previous studies of the human MOCR time-course used distortion product otoacoustic emissions (DPOAEs). These studies measured changes in DPOAEs due to CAS or adaptation (Kim *et al.*, 2001, Bassim *et al.*, 2003). DPOAE adaptation time-courses were fit with 2 exponentials: One with a fast time constant, 10 ms$<\tau_{fast}<$350 ms (median ~70 ms), and another with a slow time constant, 350 ms$<\tau_{slow}<$5.5 s (medians 1.5 and 2.1 s from two studies). While the 70 ms time constant[6] matches our fast onset time constant (Table II), the 1.5–2 s

time constant is notably absent from our data. One explanation comes from a DPOAE adaptation study in cat where a ~1 s onset time constant was observed and persisted after complete OCB section showing that it was mediated by intrinsic cochlear processes, not the MOCR (Liberman *et al.*, 1996). Counter to this explanation, Bassim *et al.* (2003) reports roughly similar ~1 s time constant for changes in DPOAEs due to either CAS or adaptation.

Another difference between our data and the DPOAE data is that onset time constants of ~300 ms and ~10s of s are present in our data but were not noted in the DPOAE studies. Kim *et al.*, did observe a peak in their time constant histogram at the maximum obtained value of 5.5 s, and suggested that longer stimuli and recording durations would be necessary to determine these time constants accurately. A similar comment can be made about our slowest onset time constants, $\tau_{slow} \sim 10$s of s. Our fitting procedures showed that there were such slow time constants. We were able to put a lower bound, $\tau_{slow} > 5.9$ s, on the slow time constant so it cannot correspond to the 1 to 2 s time constant reported in the DPOAE data (Liberman *et al.*, 1996, Kim *et al.*, 2001, Bassim *et al.*, 2003). Histograms of the time constants obtained via DPOAEs give hints of more peaks than the 2 time constants put forward by Kim *et al.* (2001) and Bassim *et al.* (2003). It seems likely that higher SNR DPOAE measurements would have revealed more details of the time-course, including more than two onset time constants. However, pursuing the MOCR time-course using DPOAE adaptation is complicated because (1) DPOAEs are composed of two underlying OAE sources with different properties that may behave differently under MOCR activation and (2) DPOAE adaptation can be mediated by intrinsic cochlear effects in addition to the MOCR activation. Presumably it is these complexities that cause MOCR activation to sometimes increase DPOAE levels;[7] furthermore these complexities may explain the wide scatter in DPOAE-based measures of MOCR time constants. Until these complexities are understood, it is difficult to appreciate the MOCR time-course via a DPOAE method.

SFOAE-based methods are better suited to MOCR time-course investigation. They are able to measure the time-course in individual subjects, and are easier to interpret than those considered above because they require only a single low-level (40 dB SPL) probe-tone in the measurement ear, one that does not appear to activate the MOCR itself (Guinan *et al.*, 2003). While a suppressor was used to normalize the data, and could be considered a complication, normalization does not affect the time-courses of the measurements.

There are limits to the present study. Specifically, the MOCR time-course reported here applies only to our measurement conditions, i.e., bilateral 40 dB SPL probe tones near 1 kHz, wide-band noise elicitors (flat spectrum and notched), and subjects that were awake but with unmonitored arousal states. It is possible that other conditions, for example other elicitors, e.g., narrow-band noise elicitors, tones, clicks, etc., would produce different results. Another limitation is that we used only noise elicitors of moderate to low levels. At high-levels, where the MOCR appears to afford protection from acoustic trauma (Rajan, 1988, Maison and Liberman, 2000, Rajan, 2000), the time-course of efferent effects might be different. Finally, our study involved only 9 ears.

2. Onset time constants compared with cellular processes in outer hair cells (OHCs)

The fast (~70 ms) onset time constant we report was also found in many other cochlear measurements in animals involving MOC activation: Basilar membrane motion, otoacoustic emissions, cochlear microphonic, endolymphatic potentials, compound action potentials, and single auditory nerve fiber recordings (Fex, 1962, Desmedt *et al.*, 1971, Warren and Liberman, 1989, Cooper and Guinan, 2003). Such universal presence indicates that the 70 ms time constant is a fundamental time constant of the MOCR.

Evidence for one possible cellular correlate for the ~70 ms time constant comes from measurements of the calcium (Ca^{2+})-mediated potassium (K^+) current found in OHCs (Housley and Ashmore, 1991, Fuchs and Murrow, 1992, Sridhar *et al.*, 1997). Animal work on guinea pig isolated outer hair cells (OHCs) has shown that upon application of acetylcholine (ACh), an outward K^+ current develops with a time constant of 70 ± 14 ms (Evans *et al.*, 2000) but these measurements were done at room temperature and the actual K^+ current time constant at body temperature is likely to be faster. The K^+ current is a result of the change in intracellular calcium concentrations due to an influx of Ca^{2+} through MOC synapses. Could this be what gives rise to the universal ~70 ms time constant?

Ca^{2+} concentrations in the basal end of OHCs following the application of ACh have also been seen to increase on a ~300 ms time scale (Evans *et al.*, 2000) presumably through a different mechanism. It is possible that the ~330 ms time constant we report is related to Ca^{2+} release from local subsynaptic cisternae involving ryanodine receptors (Sridhar *et al.*, 1997).

Evidence of a slow (10s of s) onset time constant (Table II) has been previously reported in guinea pig ($\tau_{slow} = 34 \pm 9$ s) (Sridhar *et al.*, 1995, Sridhar *et al.*, 1997). In addition, a study on guinea pig basilar-membrane (BM) motion found that BM motion is inhibited by MOC activation on fast (10–100 ms) and slow (10–100 s) time scales and these data indicate that the slow effect has a different mechanical origin (Cooper and Guinan, 2003) than the fast effect.

A possible OHC cellular correlate for τ_{slow} comes from the observations that (1) application of ACh to isolated OHCs results in a decrease of OHC axial stiffness on a slow time scale (~20 s), and (2) increases in intracellular Ca^{2+} cause the OHC cell body to elongate on the slow time scale (Sziklai *et al.*, 1996, Dallos *et al.*, 1997, Frolenkov *et al.*, 2003). The decrease in OHC stiffness appears to involve OHC regions far from MOC-OHC synapses. Perhaps the slow time constant is due to the time it takes for Ca^{2+} to migrate up the OHC[8] and affect the cytoskeleton (Evans *et al.*, 2000).

Whether, the above postulated origins for these onset time constants are correct or not, the rough correspondence between the fast, medium, and slow time constants we measured and those found in OHCs suggest that OHC cellular

processes could account for the majority of the MOCR time-course. Other possibilities do not look as promising. Post stimulus time histograms (PSTH) from recordings of single MOC neurons responding to tone bursts showed that MOC firing rate decreased over time (i.e., MOC neurons adapted) with a time constant of ~68 ms (Brown, 2001). The MOCR effects we measured *increased* monotonically with continued stimulation, thus we did not see any effect of adaptation in our MOCR time-courses. The MOCR time-course we observed could conceivably be caused by recruiting more and more efferent fibers as the stimulus persists. However, since OHCs have been shown to have processes that evolve with time-courses similar to the MOCR time-course, a more likely explanation is that a majority of the time-course is generated downstream of MOC neurons.

B. Onset and offset delays

Our best estimates of onset and offset delays (i.e., the time between elicitor onset or offset and a change in the observed response) put these values at ~25 ms for both. Previous studies found a large range of contralateral reflex latencies (Mott *et al.*, 1989, Lind, 1994, Giraud *et al.*, 1997, Hill *et al.*, 1997, Maison *et al.*, 2001b) though the methods used in these studies varied greatly in their ability to accurately detect the latency.

Our measured reflex delays can be divided into three components: (1) A neural delay,[9] defined as the delay from sound onset (or offset) to the firing (or decrease in firing) of MOC neurons at their OHC synapses; (2) an MOC-OHC synaptic delay, which includes the delays within OHCs; (3) a reverse otoacoustic delay, the delay for the MOC-induced change to appear as a change in the SFOAE in the ear canal. One theory puts the reverse otoacoustic delay at approximately half of the overall SFOAE delay (Zweig and Shera, 1995) or ~5 ms, another theory puts it close to 0 (Siegel *et al.*, 2005). The minimum neural delay in experimental animals is approximately 5 ms (reviewed by Brown *et al.*, 2003) but the latency of most neurons varies with sound level and is typically ~20 ms at 10 dB above the neuron's threshold. Thus, there is between 0 and 20 ms left for the MOC-OHC synaptic delay although the true synaptic delay is probably much less than 20 ms. There could be as little as only a few ms synaptic delay for MOC activity to begin to increase the flow of calcium into the OHC postsynaptic region and produce a measurable MOC effect.

The accuracy of our latency measures (SDs=15 ms) combined with the small range of levels tested (20 dB range) prevented us from determining whether MOCR latencies were level dependant as has been reported for individual MOC fibers (Brown, 1989). However, a strong systematic (e.g., slope=30 ms/10 dB) level dependence is unlikely.

C. MOCR decay and onset time-courses appear to be governed by a single second-order system

Both the onset and decay of MOCR responses can be fit to a second-order system and only one system parameter needs to be changed to account for the elicitor-on/elicitor-off difference (Appendix C). The onset time-course, by itself, can be produced by separate, single-time-constant processes. The decay time-course, on the other hand, shows an overshoot like a decaying sinusoid that cannot be produced by separate, single-time-constant processes. The finding that a second-order system can account for 70%–90% of both time-courses (Table II and Appendix C) by allowing one parameter to change by a factor of 2 is evidence that the MOCR's onset and decay are governed by the same second-order system. In short, one system can account for the bulk of the time-course, giving rise to the two faster onset time constants as well as the decay overshoot. Considering the similarity of the MOCR time constants with those from various OHC measurements, this "one system" seems likely to be in OHCs.

D. Time-course evidence for binaural interaction

Two of three subjects produced bilaterally activated MOCR responses that were well approximated by the sum of the ipsilaterally and contralaterally activated responses, i.e., they showed no binaural interaction of the reflexes (Fig. 8). In contrast, ear 109R produced a lower bilateral response than the sum of the unilateral responses, indicating a binaural inhibition. Possible sources for this binaural inhibition include central reflex inhibition and/or a saturation of the MOC effect in the periphery.

Binaural interactions that were a function of elicitor interaural time difference have been recently reported (Kemp *et al.*, 2005). Our experiments used the same elicitor noise in both ears and, therefore, could not detect binaural interactions that required different signals in the two ears.

E. Implications of the MOCR time-course for the role of the reflex in hearing

The human MOCRs time-course is best suited to operate on acoustic changes that persist for 100's of ms or longer. This conclusion is based upon the fact that overall MOCR effects in the cochlea build up and decay over a time scale of 100's of milliseconds. Time constants of 10s of seconds are also present. Consequently, MOCR is not fast enough to be operating on the information-carrying envelope of speech or to provide any protection from isolated loud impulse noises, such as gunshots (Wagner *et al.*, 2005). The time-course is, however, consistent with other roles that have been postulated for the MOCR: Protection (from long-lasting, >100 ms loud sounds) (Rajan and Johnstone, 1983, Reiter and Liberman, 1995), dynamic range adjustment (Geisler, 1974, Winslow and Sachs, 1988), signal detection in noise (Winslow and Sachs, 1988, Kawase *et al.*, 1993), and aiding selective attention (Hernandez-Peon, 1956, Meric and Collet, 1994).

The MOCR and the MEMR respond differently to increasing acoustic elicitor levels. MEMR response onset latencies and onset time constants decrease with increasing elicitor level (Moller, 1961), but MOCR latencies and time constants appear to depend very little on elicitor level (Fig. 6). In addition, MEMR responses wane in the presence of continuing low-level elicitors, but do not adapt for high-level elicitors (Dallos, 1964). Furthermore, the MEMR has a high threshold. These MEMR time-course properties are useful

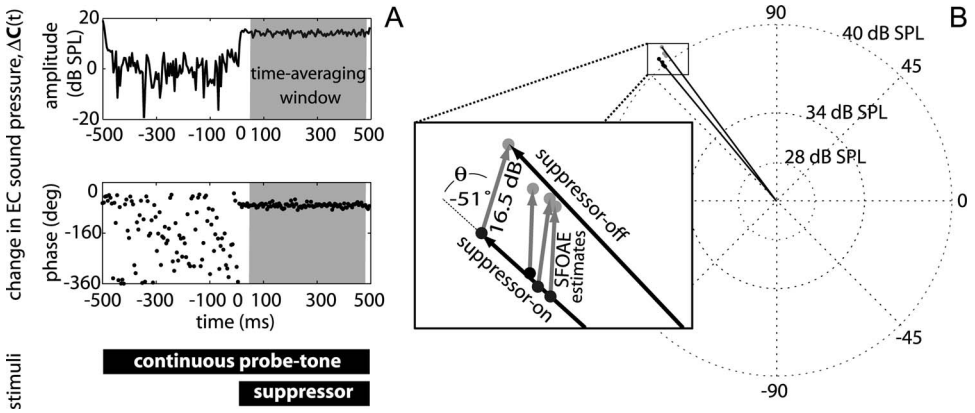


FIG. 10. Panel A shows how the change EC sound pressure (relative to the average pre-suppressor EC pressure across $t=-450$ to -50 ms) during a two-tone suppression measurement is used to estimate an SFOAE amplitude. Vector time-averaging the change in EC sound pressure due to the suppressor-tone (i.e., across shaded region, 50–450 ms) produced an estimate of the native SFOAE. Panel B shows the repeatability of the estimate of SFOAE by plotting four such SFOAE estimates (grey vectors) made at different times for the same subject, ear, and probe-tone frequency. The SFOAE vector labeled 16.5 dB SPL in panel B was obtained from the measurement shown in panel A. The raw suppressor-off and suppressor-on vectors (from time averaging over the same shaded window prior to subtraction, and not shown in panel A) are also shown. Stimulus parameters were: probe-tone frequency=1.04 kHz, probe-tone level =40 dB SPL, suppressor-tone frequency=930 Hz, suppressor-tone level=60 dB SPL, subject 87L.

for a reflex serving a protective role, the primary role of the MEMR. The MOCR time-course, on the other hand, shows little or no dependence on elicitor level (Fig. 5) and has a lower threshold so that it is acting throughout most of the range of hearing. In other words, over the measurable range of elicitor levels for both reflexes, there is a fundamental difference in the way the two reflexes respond to increasing elicitor levels. This difference broaches the possibility of different primary functions for the two reflexes.

The pupillary light reflex is similar to the MOCR in that it operates over 100's of milliseconds. The function of the pupillary light reflex is to adjust the amount of light entering the eye in order to optimize visual acuity (Campbell and Gregory, 1960). Perhaps the MOCR does a similar thing for sound and the ear.

ACKNOWLEDGMENT

This work was supported by NICDC R01 DC005977, P30 DC005209.

ABBREVIATIONS

BM = basilar membrane
$\mathbf{C}(t)$ = time varying complex compound-tone (measured)
CAS = contralateral acoustic stimulation
CF = characteristic frequency
CM = cochlear microphonic
DPOAE = distortion-product otoacoustic emission
EC = ear-canal
EP = endolymphatic potential
MEM = middle-ear-muscle
MEMR = middle-ear-muscle reflex
MOC = medial olivocochlear
MOCR = medial olivocochlear reflex
OCB = olivocochlear bundle
OHC = outer hair cell
P = significance probability ($P<0.05$ is considered statistically significant)
PSTH = post-stimulus time histogram
R = correlation coefficient
SNR = signal-to-noise ratio
SFOAE = stimulus-frequency otoacoustic emission
$\Delta\mathbf{SFOAE}(t)$ = time varying change in complex stimulus frequency emission (calculated)
SOAE = spontaneous otoacoustic emission
TEOAE = transiently evoked otoacoustic emission
VS = vector strength

APPENDIX A: MEASUREMENT NORMALIZATION

SFOAE-based MOCR measurements produce low-amplitude responses when the SFOAE has a low-amplitude; however, MOCR strength is not necessarily weak in these cases. To compare MOCR strengths across measurements with varying SFOAE amplitudes, we report responses in units relative to an estimate of the amplitude of the SFOAE being monitored

$$|\Delta\mathbf{SFOAE}(t)| = \frac{|\Delta\mathbf{C}(t)|}{|\mathbf{SFOAE}|}. \quad \text{(A1)}$$

To estimate $|\mathbf{SFOAE}|$ we suppressed the SFOAE via two-tone suppression. Ideally, if suppression of the SFOAE is complete, the amplitude of the vector difference in EC sound pressure between the suppressed and nonsuppressed conditions would equal the native SFOAE amplitude (i.e., $|\Delta\mathbf{C}|=|\mathbf{SFOAE}|$). We used a 60 dB SPL ipsilateral suppressor-tone stimulus 110 Hz below the probe-tone to suppress the SFOAE. This tone did not completely suppress the SFOAE (see Appendix D for more details) but gave a good approximation. $|\mathbf{SFOAE}|$ was taken to be the average change in EC sound pressure amplitude caused by the suppressor-tone (the average $|\Delta\mathbf{C}(t)|$, see Fig. 10). The $|\mathbf{SFOAE}|$ estimate was stable and repeatable across sessions and days even though the pre-suppressor baseline wandered.

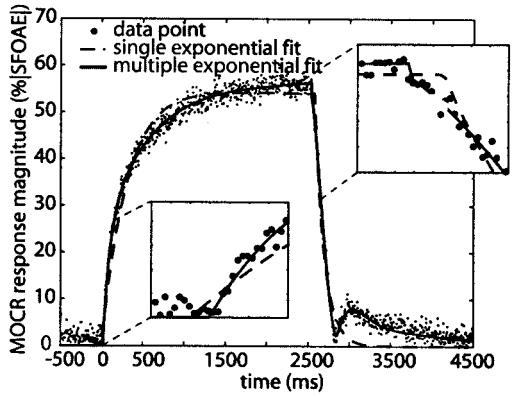


FIG. 11. Two characterizations of the MOCR time course evoked by contralateral sound. Data (points) show MOCR mediated change in the SFOAE induced by a 60 dB SPL contralateral wideband noise elicitor (on at 0 ms and off at 2500 ms). The two curves represent two mathematical characterizations of the data: a fit using a single complex exponential term for the rising and decaying portions (dashed) and one using three such terms for each portion (solid). Exploded views show transition regions. Since the rising and falling portions were fitted separately, "jumps" (less than the measurement noise) could occur (ear S61R, test tone 1.1 kHz, 816 averages).

In other words the SFOAE rode on baseline variations (Fig. 10). One source of these baseline variations is movement of the earphone assembly within the ear canal.

APPENDIX B: MATHEMATICAL CHARACTERIZATIONS OF THE MOCR RESPONSE WAVEFORM

In order to quantify the MOCR responses, two mathematical characterizations were made for each waveform (Fig. 11). First, single complex exponential characterizations were made to quantify MOCR time constants for the rising and decaying portions of each response. Second, sum of complex exponentials characterizations were made to quantify onset and offset delays, extract multiple time constants (when possible), and to describe the rising and decaying curvatures more accurately. Complex exponentials were always used because these produced curvatures similar to the measured responses and because complex terms could fit magnitude and phase data simultaneously.[10]

1. Single complex exponential characterization

For the single complex exponential characterization a piecewise continuous function $\mathbf{y}_1(t)$ [Eq. (B1)] with 5 free parameters (t_r, $\mathbf{Z}_r$, $\mathbf{s}_r$, t_d, and $\mathbf{s}_d$) was used to characterize each response waveform, $\Delta\mathbf{SFOAE}(t)$.

$$\mathbf{y}_1(t) = \begin{cases} t < t_r, & \mathbf{y}_0 = 0 + 0i \\ t_r \leqslant t \leqslant t_d, & \mathbf{y}_r(t) = \mathbf{Z}_r(1 - e^{\mathbf{s}_r(t-t_r)}) \\ t_d < t, & \mathbf{y}_d(t) = \mathbf{Z}_d e^{\mathbf{s}_d(t-t_d)} \\ \textit{with } \mathbf{Z}_d \textit{ set to } \mathbf{y}_r(t_d) & \end{cases} . \tag{B1}$$

Where:

t_r=time when the response onset begins (re: elicitor onset, i.e., the onset delay)

t_d=time when the response offset begins (re: elicitor *onset*)

$\mathbf{s}_r$=complex exponent of the rise[11]

$\mathbf{s}_d$=complex exponent of the decay

$\mathbf{Z}_r$=eventual (saturated) maximum (complex-valued)

$\mathbf{Z}_d$=complex value from which the decay begins (not a free parameter, determined by $\mathbf{Z}_r$, $\mathbf{s}_r$, and t_d)

All parameter values were determined simultaneously using a non-linear parametric least-squares search algorithm that minimized the error between $\mathbf{y}_1(t)$ and a response waveform, $\Delta\mathbf{SFOAE}(t)$. From this "single-complex-exponential" characterization, single onset and offset time constants were calculated [Eqs. (B2) and (B3)]

$$\tau_r = -1/\mathrm{Re}\{\mathbf{s}_r\}, \tag{B2}$$

$$\tau_d = -1/\mathrm{Re}\{\mathbf{s}_d\}. \tag{B3}$$

2. "Sum of complex exponentials" characterization

For the "sum of complex exponentials" characterization, the piecewise function, $\mathbf{y}_2(t)$, [Eq. (B4)] was used to characterize each response waveform, $\Delta\mathbf{SFOAE}(t)$

$$\mathbf{y}_2(t) = \begin{cases} t < t_r, & \mathbf{y}_0(t) = 0 \\ t_r \leqslant t \leqslant t_d, & \mathbf{y}_r(t) = \mathbf{Z}_{DC} + \sum_{n=1}^{2 \text{ or } 3} \mathbf{Z}_{r_n}(1 - e^{\mathbf{s}_{r_n} t}) \\ t_d < t, & \mathbf{y}_d(t) = \sum_{n=1}^{2 \text{ or } 3} \mathbf{Z}_{d_n} e^{\mathbf{s}_{d_n} t} \end{cases} . \tag{B4}$$

The number of terms (and, therefore, parameters) was determined by an algorithm.[12] The roles of the parameters in this characterization are analogous to those in the single complex exponential characterization described above except: (1) The real and imaginary parts of the complex exponents cannot necessarily be separated when being interpreted (e.g., complex conjugate exponents add to form sinusoids[13]); (2) the eventual maximum value of the response is represented by the sum of $\mathbf{Z}_{r_n}$ and a constant, $\mathbf{Z}_{DC}$.

To carry out each characterization, the waveform was split into three regions: (a) Pre-response, in which the response was considered to be zero; (b) rising response (including the plateau); and (c) decaying response. Every data point was ascribed to one of the three regions. Two time-boundaries, t_r, and t_d, were chosen in order to separate the regions (this was re-done for all reasonable t_r, t_d pairs). For each t_r, t_d pair, all remaining parameters in each region were determined by using a matrix pencil method (Hua, 1990, Sarkar, 1995). The parameters of the rising portion were not constrained to make the $\mathbf{y}_2(t)$ curve go though the point 0. Neither were the rising and decaying portions of the characterization constrained to coincide. When discontinuities appeared at these boundaries they were within the measurement noise.

The number of terms to use was determined by a cost/benefit approach. Terms were added, $\mathbf{Z}_{r_n}(1-e^{\mathbf{s}_{r_n}t})$ for the rising region or $\mathbf{Z}_{d_n}e^{\mathbf{s}_{d_n}t}$ for the decaying region, until the additional term failed to improve the goodness of fit (R^2 value) in that region or did so only incrementally.[14] Once the optimal number of terms and the associated parameter values were determined for each region, the three regions (i.e., the rising, falling, and the pre-elicitor region assumed to contain no signal) were conjoined to form a curve. The "fit" between the resulting composite curve, $\mathbf{y}_2(t)$, and the entire response waveform being characterized, $\Delta\mathbf{SFOAE}(t)$, was evaluated using the standard goodness of fit value, R^2, evaluated over all three regions. The process was repeated for all reasonable $[t_r, t_d]$ pairs, and the $\mathbf{y}_2(t)$ curve that gave the highest R^2 value was selected as the winning sum of exponentials characterization (this included the t_r and t_d parameters that segregated the winning curve as well as the associated parameters determined from the matrix pencil method: $\mathbf{Z}_{r_n}, \mathbf{s}_{r_n}, \mathbf{Z}_{d_n}, \mathbf{s}_{d_n}$)

The sum of exponentials characterization was used to extract multiple time constants where possible. Conventional time constants are defined by real-valued exponents but our exponents were complex-valued. Consequently, the calculation of multiple conventional time constants was complicated by interactions between $\mathbf{s}_{r_n}$ or $\mathbf{s}_{d_n}$ terms. However, if the imaginary parts of the parameters were small relative to their real parts (i.e., the exponents were mostly real), multiple onset and decay time constants were approximated, respectively, as

$$\tau_{r_n} = -1/\mathrm{Re}\{\mathbf{s}_{r_n}\}, \tag{B5}$$

$$\tau_{d_n} = -1/\mathrm{Re}\{\mathbf{s}_{d_n}\}. \tag{B6}$$

In order to quantify onset and offset delays with greater time resolution than our sample rate (5 ms/point), onset and offset delays were calculated from the winning sum of exponentials characterization. The response onset time (the time of earliest MOCR response after elicitor inception) was determined by the time of intersection, or the time of nearest miss when there was no intersection,[15] of the rising portion of the characterization and 0 (the origin of the complex plane). The response-offset time was similarly determined by the time of intersection, or the time of nearest miss when there was no intersection, of the rising and decaying portions of the characterization. The onset delay and the offset delay were then calculated by subtracting elicitor onset and offset times from response onset and offset times respectively.

C. Calculating the best estimates of parameters and estimating their errors

To provide the best estimate for each parameter and to quantify the error of those best estimates, two simulations were done for every waveform characterized. The first simulation was used to quantify the error caused by measurement noise; the second simulation was used to quantify biases introduced by the characterization process.

The first simulation calculated means and variances at each time point from individual trials prior to averaging. These statistics were used to define independent Gaussian distributions for the real and imaginary parts at each time point from which new samples could be picked. By sampling from each time point's distribution 500 times, 500 waveforms that were statistically similar to the original waveform were generated. These 500 waveforms were mathematically characterized in the same way as the original response waveform, and they produced 500 estimates for each parameter so that each parameter had a distribution.

The second simulation added noise taken at random with replacement from a subject's pre-elicitor response region (−450 to −50 ms) to a signal with preset parameters. 500 waveforms were made this way; these produced 500 estimates of each known parameter. The known parameter values were then subtracted from the estimates to reveal any bias. This analysis showed that onset and decay time constants were free from significant bias, but that onset and offset delays were underestimated by 0–10 ms depending on the noise level.

The distribution of values of a parameter from the first simulation was convolved with the bias distribution from the second simulation. The resulting distribution was used to obtain the best estimate for that parameter (the distribution mean) and the standard deviation of the best estimate (the distribution standard deviation).[16] These are the values reported in this manuscript.

APPENDIX C: A SIMPLE PHENOMENOLOGICAL MODEL FOR THE MOCR RESPONSE

Although the rising time-course of the MOCR response appears fundamentally different in shape than its decay, a simple model shows how the main features of both time-courses may be due to the same underlying system. To better understand how these time-courses might be connected, we attempted to model both using a simple second-order system, Eq. (C1). Symbols common for a mechanical system were used but the actual underlying system is more likely neurochemical.

$$M\ddot{x}(t) + B\dot{x}(t) + Kx(t) = F_{\mathrm{in}}. \tag{C1}$$

For the system of Eq. (C1), the displacement time waveform, $x(t)$, in response to a step displacement input, F_{in}, can be found by invoking a Laplace transformation and solving which gives Eq. (C2).

$$x(t) = Z_0 + \sum_{n=1}^{2} \mathbf{Z}_n e^{\mathbf{s}_n t}. \tag{C2}$$

The system's behavior is determined by the parameters M, B, and K or equivalently by the loci of the complex-valued "poles" [$\mathbf{s}_1$ and $\mathbf{s}_2$ after solving Eq. (C2) for $\mathbf{s}$ or the roots of Eq. (C1)]. If the poles lie on the real axis, pure exponential curvatures exist; if the poles are complex conjugates, a sinusoidal component exists. The intensity independence of the time-course means that multiplication of Eq. (C1) by any constant will continue to produce the same time-course, therefore, it is the ratios of the parameters that determine the time-course thus the problem can be reduced to two parameters K/M and B/M.

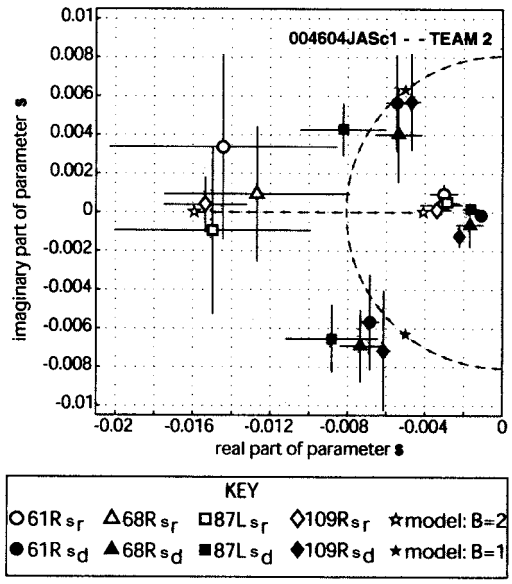


FIG. 12. Exponent parameters or poles for data from four ears that had high SNR compared with the poles of a simple second order model (for an explanation of parameters see Appendix B). The rising portion of the responses (unfilled symbols) were well characterized by two real "poles" with two corresponding time constants. The decaying portion of the responses (filled symbols) were well characterized by 2 complex conjugate poles (accounting for >70% of the decay curvature) plus a single real pole with a long time constant. The dashed lines with arrows show how the two poles of a simple second order system move when the value of the damping element is changed from 0 to 2 (while keeping chosen mass and spring elements constant: $M=100$ kg, $K=0.0065$ N/m). Particular pole locations for two values of the damping element are designated by stars. The bulk of the observed wave-shapes can be accounted for by changing a single parameter, the damping element (or first derivative coefficient) in a simple second order model. Open symbols are onset poles and filled symbols are decay poles and different symbol shapes are used for different ears as shown in the key.

The loci of four sets of $\mathbf{s}_{r_n}$ and $\mathbf{s}_{d_n}$ parameters (see Appendix B) from the sum of exponentials characterizations of our four highest SNR MOCR responses were compared with the loci of $\mathbf{s}_1$ and $\mathbf{s}_2$ poles from the simple second-order model (Fig. 12). With M and K fixed ($M=100$ kg and $K=0.0065$ N/m) we matched the locations of the rising $\mathbf{s}_{r_n}$ parameters with a B value of 2 Ns/m. Keeping M and K at their fixed values, we closely matched two of the three[17] $\mathbf{s}_{d_n}$ parameters solely by reducing B by a factor of two to $B=1$ Ns/m (Fig. 12). Thus, for these subjects, the transition from the rising curvature to the decaying curvature can be largely achieved by changing a single parameter (the first derivative coefficient) in a second-order model. This suggests that a single second order process rather than multiple unconnected processes are largely responsible for the MOCR's time-course.

The agreement between these data and the simple second-order model is not perfect. The "conjugate poles" were not true conjugates in the sense that they did not form a vertical line and in that they did not have equal weighting (generally the term with the lower pole had a larger coefficient magnitude). And the data constellations did not strictly overlay the second-order system's path (dashed line in Fig. 12). Nonetheless, the model provides a good approximation to the measured data and shows that much of the difference between the rising and falling time-courses can be accounted for by a single underlying second-order system.

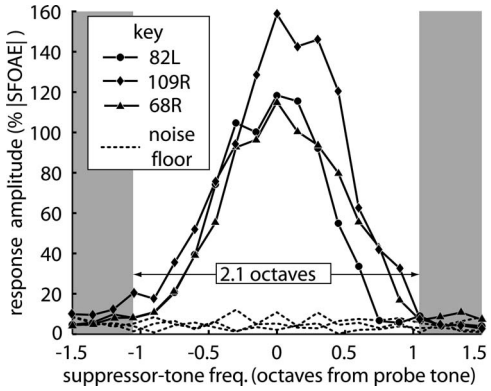


FIG. 13. Bandwidth of two-tone suppression measured via the average change in SFOAE magnitude between suppressor-on and suppressor-off conditions for 60 dB SPL tone suppressors at various frequencies relative to the probe tone. 16 one-second responses were averaged for each point. Ipsilateral suppressor tones (500 ms duration) produced two-tone suppression effects (solid lines), while contralateral "suppressor tones" produced no effects (dashed lines) and were used to asses the noise floor. The edges of the gray bands demark the chosen 2.1 octave span for the spectral notch incorporated into elicitors when comparing ipsilaterally, contralaterally, and bilaterally elicited MOCR responses.

APPENDIX D: TWO-TONE SUPPRESSION BANDWIDTH

Measuring ipsilaterally or bilaterally activated MOCR responses using our SFOAE-based method required a probe-tone and an elicitor-noise be presented simultaneously in the measurement ear. In this situation, two-tone suppression by the elicitor can reduce the SFOAE amplitude, obfuscating MOCR responses. To avoid such contamination of our MOCR responses, we used an elicitor-noise with a spectral notch centered at the probe-tone frequency. The notch width was designed to be wide enough to not produce two-tone suppression effects. To determine the appropriate notch width, we measured the bandwidth of two-tone suppression of the SFOAE (Fig. 13).[18] For these measurements, a 40 dB SPL probe tone was fixed at a subject's probe frequency (near 1 kHz) while a 60 dB SPL suppressor-tone was presented at various frequencies (21 logarithmically spaced frequencies spanning three octaves and centered at the probe tone frequency were used). Since the MOCR does not respond well to pure tone stimuli (Guinan *et al.*, 2003) as demonstrated in Fig. 13 by the lack of any effect of contralateral "suppressor-tones" dashed lines), all resulting changes in the SFOAE from adding ipsilateral tone suppressors were presumed to be to non-MOCR, two-tone suppression effects. The bandwidth of suppression was found to be ~2.1 octaves.

When comparing ipsilaterally, contralaterally, and bilaterally activated MOCR time-courses (Sec. III C), a 2.1 octave spectral notch centered at the probe-tone frequency was used for all noise-elicitors. In retrospect, since subject 109R showed some suppression at −1.05 octaves, perhaps the notch should have been extended for this subject. This oversight, however, is unlikely to have much of an effect. The total level of the notched-noise elicitor was 60 dB SPL,

therefore the elicitor would introduce much less energy at the −1.05 octave place in the cochlea than the 60 dB SPL suppressor-tone used for Fig. 13 that concentrates all its energy there.

[1]Bilateral probe tones were used because our previous work used them. For these experiments, however, bilateral probes do not provide any benefit aside from allowing comparison with our prior work.

[2]Because spontaneous emissions can be entrained (Van Dijk and Wit, 1988), they may shift toward and interact with the probe-tone stimulus confounding SFOAE-based measures (Burns *et al.*, 1984). To resolve this issue, once a potential measurable probe-tone frequency was identified, two buffers (2621 ms each) were acquired from the microphones in quiet and their spectra were averaged. If any spontaneous emission >-5 dB SPL was detected within 50 Hz of the prospective probe-tone frequency, a new probe frequency was chosen.

[3]The criterion for acceptable SNR was 95% phase coherence, defined by a vector strength (VS) >0.95 in a time window during the elicitor (i.e., 1950–2450 ms, see Table I). Phase coherence was used because coherent phase is a strong indicator of signal presence. Vector strength was calculated by adding vectors of unity magnitude but with phase values taken from the data points in the time window (Goldberg and Brown, 1969) (n in the equation below represents indexes to these $\sim$500 consecutive samples). The resultant sum vector was then divided by the total number of vectors included in the sum, and its length had to be >0.95:

$$VS = \frac{1}{N}\sum_{n=1}^{N} e^{i\angle \mathbf{C}(n)} > 0.95.$$

[4]The term "long" was used for this time constant to differentiate it from the three other time constants discussed in this paper, i.e., fast, medium, and slow. All four time constants are numerically distinct.

[5]The idea that allows us to put a lower bound on the slow time constant is, "if the time constant were shorter, we would have seen the curvature produced by it." Since the early rise of an exponential is approximately linear, when the data displayed a very slow (or effectively linear) curvature, we could put a lower bound on the time constant that depended on the SNR of that parameter and the analysis window length. The lower bound was set at 95% confidence.

[6]70 ms was the reported *median*, the average is larger due to significant scatter in the data.

[7]*In the present dataset* the MOCR always decreased the estimated SFOAE.

[8]There is a system of Ca^{2+} stores (sub-surface cisternae) located along OHC lateral cell-membranes that could be involved in migration of Ca^{2+}.

[9]The neural delay contains both forward-otoacoustic and neural delays. One reason these two delays are combined is because it is hard to calculate what the forward-otoacoustic delay should be for a wideband stimulus—delays in the cochlea are different for different frequencies. Time "0" in all figures was when the elicitor stimulus's electrical signal was sent to the sound source, therefore, all otoacoustic delays include acoustic delays of <0.1 ms between the earphone and the tympanic membrane.

[10]This strategy removes the bias that is introduced when fitting the magnitude data alone. When noise exists in a complex-valued measurement (zero mean noise in x and y) the resulting magnitude is always $>$ zero. Therefore, an estimate of the magnitude of the signal based on only magnitude data will be biased and the amount of the bias will increase with increasing noise.

[11]For a single term complex exponential, the complex-valued exponent can be thought of in its component parts. The real part of the exponent is inversely related to a conventional time constant and determines the speed of the magnitude changes over time, while the imaginary part determines the speed with which the phase changes over time.

[12]The number of parameters was increased until the least squares fit failed to improve or did so only incrementally. Higher SNR measurements produced more terms. The maximum number of parameters generated was 14. This happened for the four highest SNR subjects: 87L, 109R, 68R, and 61R, who's rising and decaying response curvatures were described by the sum of 3 complex exponential terms.

[13]No restrictions were placed on the parameters; consequently, the functions may defy interpretation as combinations of familiar functions.

[14]Over-fitting manifested when sinusoidal oscillations appeared in the characterization curves. Such oscillations were obvious in the curves when they occurred, and were not present in the data. These oscillations were excluded to insure smooth curves and their exclusion set an upper bound of 3 on the number of complex exponential terms that could be used in the characterization.

[15]The parameters of the rising portion were not constrained to make the $\mathbf{y}_2(t)$ curve go though the point 0. Neither were the rising and decaying portions of the characterization constrained to coincide. When discontinuities appeared at these boundaries they were within the measurement noise.

[16]This is an over-estimate of the standard deviation because both simulations used measurement noise (albeit different types). Measurement noise only appears once in the actual data.

[17]The third decay parameter $\mathbf{s}_{d_3}$ represented a long exponential time constant that was not captured by this model. It was ignored because it accounted for $<30\%$ of the decay curvature.

[18]Our standard suppression measurement used to estimate $|\mathbf{SFOAE}|$ for normalization of MOCR responses uses a suppressor tone presented 110 Hz below the probe frequency. However, the excellent cancellation afforded by alternating elicitor polarity rendered this offset unnecessary and created a situation where a suppressor tone presented nearer or even at the probe tone frequency yielded a response greater than the nominal 100%.

Bassim, M. K., Miller, R. L., Buss, E., and Smith, D. W. (**2003**). "Rapid adaptation of the 2f1-f2 DPOAE in humans: binaural and contralateral stimulation effects," Hear. Res. **182**, 140–152.

Brown, M. C. (**1989**). "Morphology and response properties of single olivocochlear fibers in the guinea pig," Hear. Res. **40**, 93–109.

Brown, M. C. (**2001**). "Response adaptation of medial olivocochlear neurons is minimal," J. Neurophysiol. **86**, 2381–2392.

Brown, M. C., de Venecia, R. K., and Guinan, Jr., J. J. (**2003**). "Responses of medial olivocochlear neurons. Specifying the central pathways of the medial olivocochlear reflex," Exp. Brain Res. **153**, 491–498.

Burns, E. M., Strickland, E. A., Tubis, A., and Jones, K. (**1984**). "Interactions among spontaneous otoacoustic emissions. I. Distortion products and linked emissions," Hear. Res. **16**, 271–278.

Campbell, F. W., and Gregory, A. H. (**1960**). "Effect of size of pupil on visual acuity," Nature (London) **187**, 1121–1123.

Cooper, N. P., and Guinan, J. J. J. (**2003**). "Separate mechanical processes underlie fast and slow effects of medial olivocochlear efferent activity," J. Physiol. (London) **548**, 307–312.

Dallos, P. (**1964**). "Dynamics of the Acoustic Reflex: Phenomenological Aspects," J. Acoust. Soc. Am. **36**, 2175–2183.

Dallos, P., He, D. Z., Lin, X., Sziklai, I., Mehta, S., and Evans, B. N. (**1997**). "Acetylcholine, outer hair cell electromotility, and the cochlear amplifier," J. Neurosci. **17**, 2212–2226.

Desmedt, J. E., La Grutta, V., and La Grutta, G. (**1971**). "Contrasting effects of centrifugal olivo-cochlear inhibition and of middle ear muscle contraction on the response characteristics of the cat's auditory nerve," Brain Res. **30**, 375–384.

Dolan, D. F., Guo, M. H., and Nuttall, A. L. (**1997**). "Frequency-dependent enhancement of basilar membrane velocity during olivocochlear bundle stimulation," J. Acoust. Soc. Am. **102**, 3587–3596.

Evans, M. G., Lagostena, L., Darbon, P., and Mammano, F. (**2000**). "Cholinergic control of membrane conductance and intracellular free Ca2+ in outer hair cells of the guinea pig cochlea," Cell Calcium **28**, 195–203.

Fex, J. (**1962**). "Auditory activity in centrifugal and centripetal cochlear fibres in cat. A study of a feedback system," Acta Physiol. Scand. **(Suppl 189)**, 1–68.

Frolenkov, G. I., Mammano, F., and Kachar, B. (**2003**). "Regulation of outer hair cell cytoskeletal stiffness by intracellular Ca2+: underlying mechanism and implications for cochlear mechanics," Cell Calcium **33**, 185–195.

Fuchs, P. A., and Murrow, B. W. (**1992**). "A novel cholinergic receptor mediates inhibition of chick cochlear hair cells," Proc. R. Soc. London, Ser. B **248**, 35–40.

Geisler, C. D. (**1974**). "Letter: Hypothesis on the function of the crossed olivocochlear bundle," J. Acoust. Soc. Am. **56**, 1908–1909.

Giraud, A. L., Wable, J., Chays, A., Collet, L., and Chery-Croze, S. (**1997**)."Influence of contralateral noise on distortion product latency in humans: is the medial olivocochlear efferent system involved?," J. Acoust. Soc. Am. **102**, 2219–2227.

Goldberg, J. M., and Brown, P. B. (**1969**). "Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization," J. Neurophysiol. **32**, 613–636.

Guinan, J. J. (**1996**). "Physiology of Olivocochlear Efferents," in *The Co-*

chlea, edited by P. Dallos *et al.* (Springer, New York), Vol. **8**, pp. 436–502.
Guinan, J. J., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (**2003**). "Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," J. Assoc. Res. Otolaryngol. **4**, 521–540.
Hernandez-Peon, H. S., and Micheal, Jouvet (**1956**). "Modification of electric activity in cochlear nucleus during "attention" in unanesthetized cats," Science **123**, 331–332.
Hill, J. C., Prasher, D. K., and Luxon, L. M. (**1997**). "Latency of contralateral sound-evoked auditory efferent suppression of otoacoustic emissions," Acta Oto-Laryngol. **117**, 343–351.
Housley, G. D., and Ashmore, J. F. (**1991**). "Direct measurement of the action of acetylcholine on isolated outer hair cells of the guinea pig cochlea," Proc. R. Soc. London, Ser. B **244**, 161–167.
Hua, Y. S., and T., K. (**1990**). "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," IEEE Trans. Acoust., Speech, Signal Process. **38**, 814–824.
Kawase, T., Delgutte, B., and Liberman, M. C. (**1993**). "Antimasking effects of the olivocochlear reflex. II. Enhancement of auditory-nerve response to masked tones," J. Neurophysiol. **70**, 2533–2549.
Kemp, D., Hsueh, J.-Y., and Slightam-Wearne, J. (**2005**). "*Effferent OAE Suppression and Binarual Interactions*," Proceedings of the ARO, New Orleans, **1379**.
Kemp, D. T. (**1978**). "Stimulated acoustic emissions from within the human auditory system," J. Acoust. Soc. Am. **64**, 1386–1391.
Kim, D. O., Dorn, P. A., Neely, S. T., and Gorga, M. P. (**2001**). "Adaptation of distortion product otoacoustic emission in humans," J. Assoc. Res. Otolaryngol. **2**, 31–40.
Liberman, M. C., Puria, S., and Guinan, Jr., J. J. (**1996**). "The ipsilaterally evoked olivocochlear reflex causes rapid adaptation of the 2f1-f2 distortion product otoacoustic emission," J. Acoust. Soc. Am. **99**, 3572–3584.
Lind, O. (**1994**). "Contralateral suppression of TEOAE. Attempts to find a latency," Br. J. Audiol. **28**, 219–225.
Maison, S., Durrant, J., Gallineau, C., Micheyl, C., and Collet, L. (**2001a**). "Delay and temporal integration in medial olivocochlear bundle activation in humans," Ear Hear. **22**, 65–74.
Maison, S., Micheyl, C., and Collet, L. (**2001b**). "Influence of focused auditory attention on cochlear activity in humans," Psychophysiology **38**, 35–40.
Maison, S. F., and Liberman, M. C. (**2000**). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," J. Neurosci. **20**, 4701–4707.
Meric, C., and Collet, L. (**1994**). "Differential effects of visual attention on spontaneous and evoked otoacoustic emissions," Int. J. Psychophysiol **17**, 281–289.
Moller, A. R. (**1961**). "Bilateral contraction of the tympanic muscles in man examined by measuring acoustic impedance-change," Ann. Otol. Rhinol. Laryngol. **70**, 735–752.
Mott, J. B., Norton, S. J., Neely, S. T., and Warr, W. B. (**1989**). "Changes in spontaneous otoacoustic emissions produced by acoustic stimulation of the contralateral ear," Hear. Res. **38**, 229–242.
Mountain, D. C. (**1980**). "Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics," Science **210**, 71–72.
Murugasu, E., and Russell, I. J. (**1996**). "The effect of efferent stimulation on basilar membrane displacement in the basal turn of the guinea pig cochlea," J. Neurosci. **16**, 325–332.
Rajan, R. (**1988**). "Effect of electrical stimulation of the crossed olivocochlear bundle on temporary threshold shifts in auditory sensitivity. I. Dependence on electrical stimulation parameters," J. Neurophysiol. **118**, 2434–2443.
Rajan, R. (**2000**). "Centrifugal pathways protect hearing sensitivity at the cochlea in noisy environments that exacerbate the damage induced by loud sound," J. Neurosci. **20**, 6684–6693.
Rajan, R., and Johnstone, B. M. (**1983**). "Crossed cochlear influences on monaural temporary threshold shifts," Hear. Res. **9**, 279–294.
Reiter, E. R., and Liberman, M. C. (**1995**). "Efferent-mediated protection from acoustic overexposure: Relation to slow effects of olivocochlear stimulation," J. Neurophysiol. **73**, 506–514.
Sarkar, T. K., and P., O. (**1995**). "Using the matrix pencil method to estimate the parameters of a sum of complex exponentials," IEEE Antennas Propag. Mag. **37**, 48–55.
Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., and Ruggero, M. A. (**2005**). "Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering," J. Acoust. Soc. Am. **118**, 2434–2443.
Siegel, J. H., and Kim, D. O. (**1982**). "Efferent neural control of cochlear mechanics? Olivocochlear bundle stimulation affects cochlear biomechanical nonlinearity," Hear. Res. **6**, 171–182.
Sridhar, T. S., Brown, M. C., and Sewell, W. F. (**1997**). "Unique postsynaptic signaling at the hair cell efferent synapse permits calcium to evoke changes on two time scales," J. Neurosci. **17**, 428–437.
Sridhar, T. S., Liberman, M. C., Brown, M. C., and Sewell, W. F. (**1995**). "A novel cholinergic "slow effect" of efferent stimulation on cochlear potentials in the guinea pig," J. Neurosci. **15**, 3667–3678.
Sziklai, I., He, D. Z., and Dallos, P. (**1996**). "Effect of acetylcholine and GABA on the transfer function of electromotility in isolated outer hair cells," Hear. Res. **95**, 87–99.
Van Dijk, P., and Wit, H. P. (**1988**). "Phase-lock of spontaneous oto-acoustic emissions to a cubic difference tone," in *Basic Issues in Hearing*, edited by H. Duifhuis, J. W. Horst, H. P. W., Basic Issues in Hearing (London Academic, London), pp. 101–105.
Wagner, W., Heppelmann, G., Kuehn, M., Tisch, M., Vonthein, R., and Zenner, H. P. (**2005**). "Olivocochlear activity and temporary threshold shift-susceptibility in humans," Laryngoscope **115**, 2021–2028.
Warren, III, E. H., and Liberman, M. C. (**1989**). "Effects of contralateral sound on auditory-nerve responses. II. Dependence on stimulus variables," Hear. Res. **37**, 105–121.
Winslow, R. L., and Sachs, M. B. (**1988**). "Single-tone intensity discrimination based on auditory-nerve rate responses in backgrounds of quiet, noise, and with stimulation of the crossed olivocochlear bundle," Hear. Res. **35**, 165–189.
Zweig, G., and Shera, C. A. (**1995**). "The origin of periodicity in the spectrum of evoked otoacoustic emissions," J. Acoust. Soc. Am. **98**, 2018–2047.

Asynchrony and the grouping of vowel components: Captor tones revisited

Brian Roberts[a] and Stephen D. Holmes
School of Psychology, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom and Psychology, School of Life and Health Sciences, Aston University, Birmingham B4 7ET, United Kingdom

(Received 27 October 2005; revised 21 February 2006; accepted 4 March 2006)

Asynchrony is an important grouping cue for separating sound mixtures. A harmonic incremented in level makes a reduced contribution to vowel timbre when it begins before the other components. This contribution can be partly restored by adding a captor tone in synchrony with, and one octave above, the leading portion of the incremented harmonic [Darwin and Sutherland, Q. J. Exp. Psychol. A **36**, 193–208 (1984)]. The captor is too remote to evoke adaptation in peripheral channels tuned to the incremented harmonic, and so the restoration effect is usually attributed to the grouping of the leading portion with the captor. However, results are presented that contradict this interpretation. Captor efficacy does not depend on a common onset, or harmonic relations, with the leading component. Rather, captor efficacy is influenced by frequency separation, and extends to about 1.5 oct above the leading component. Below this cutoff, the captor effect is equivalent to attenuating the leading portion of the incremented harmonic by about 6 dB. These results indicate that high-level grouping does not govern the captor effect. Instead, it is proposed that the partial restoration of the contribution of an asynchronous component to vowel timbre depends on broadband inhibition within the central auditory system. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2190164]

PACS number(s): 43.66.Ba, 43.66.Jh, 43.66.Lj, 43.71.Es [AJO] Pages: 2905–2918

I. INTRODUCTION

Acoustic elements arising from independent sources are unlikely to begin or end at the same time. Indeed, onset asynchrony is arguably the most powerful cue for the perceptual segregation of concurrent sounds (Bregman, 1990; Darwin and Carlyon, 1995). In general, an onset asynchrony of 30–50 ms greatly reduces the contribution of a single component to judgments of the timbre (e.g., Bregman and Pinker, 1978; Green and Dai, 1992) or location (e.g., Hill and Darwin, 1996) of a complex tone, and causes it to be heard out as a separate sound (e.g., Bregman and Pinker, 1978; Dannenbring and Bregman, 1978). A longer onset asynchrony, of about a few hundred milliseconds, is needed to exclude a component from the computation of global pitch (e.g., Darwin and Ciocca, 1992; Hukin and Darwin, 1995), or to exclude a whole formant from the computation of vowel timbre (Darwin, 1981).

Listeners are highly practiced at categorizing speech sounds, and so judgments of vowel quality provide a convenient way of assessing the contribution of an asynchronous component to the timbre of a complex tone. Darwin (1984a) pioneered this approach by measuring the effects on perceived vowel quality of manipulating the relative onset/offset time and level of individual harmonics in the first-formant (F1) region. These harmonics are resolved, and so the F1 frequency must be abstracted in some way from the individual peaks in this region of the excitation pattern. Darwin (1984a) noted that the perceived F1 frequency of a vowel does not correspond simply to the frequency of the highest local peak in the excitation pattern, but rather corresponds to the weighted mean of the harmonics in this spectral region (Karnickaya *et al.*, 1975; Carlson *et al.*, 1975; Assmann and Nearey, 1987). Darwin (1984a) boosted the level of a single harmonic in the F1 region and found that the extra energy changed the perceived vowel quality in a way consistent with the change in the physical position of the formant peak. However, the contribution of the extra energy was greatly reduced when it began before the other components. Darwin (1984b) demonstrated that this exclusion by asynchrony occurs even when the level of a harmonic is not raised above the vowel's original spectral envelope. Subsequent work has shown that listeners are generally able to subtract from perceived vowel quality the amplitude of that part of a harmonic that begins before vowel onset (Darwin, 1995). This result can be understood in terms of the old-plus-new heuristic, which states that any part of a current sound that can plausibly be interpreted as the continuation of a previous sound should be removed perceptually from the mixture (Bregman, 1990).

What physiological mechanism might mediate the effect of onset asynchrony on judgments of timbre? Darwin (1984a) pointed out that peripheral adaptation provides a possible alternative to an explanation based on a higher-level mechanism for grouping. The response of an auditory nerve fiber to a pure tone of fixed level is greatest at stimulus onset, decreases gradually, and reaches a steady state after a few tens of milliseconds (Kiang *et al.*, 1965; Westerman and Smith, 1984; see Palmer, 1995). Therefore, the neural response to the leading component will have adapted substan-

[a]Electronic mail: b.roberts@aston.ac.uk

tially by the time the vowel begins. This might be equivalent perceptually to reducing the level of the leading component. Darwin's (1984a) demonstration that part of the effect of leading energy persists when a silent gap of 30 ms is introduced between it and vowel onset is consistent with an explanation in terms of peripheral adaptation. However, two lines of evidence suggest that the effect of onset asynchrony depends, at least partly, on a higher-level grouping mechanism. These are considered in turn.

Darwin (1984a) pointed out that perceptual grouping can operate retroactively, whereas adaptation cannot. Therefore, if the contribution of a component to vowel quality can be reduced by making that component end after the vowel, then the effect cannot be explained in terms of peripheral adaptation. The effect of offset asynchrony is greatest for vowels of short duration (Darwin, 1984b; Darwin and Sutherland, 1984; Roberts and Moore, 1991), giving a maximum effect of about half that found for onset asynchrony. Roberts and Moore (1991) considered and rejected the idea that the effect of offset asynchrony on vowel quality is a consequence of backward masking. In particular, they noted that backward masking (*i*) tends to be weak or absent in practiced listeners (Miyazaki and Sasaki, 1984; Oxenham and Moore, 1994); (*ii*) only occurs when the masker is much higher in level than the signal (Elliott, 1971); and (*iii*) largely disappears after about 10 ms (Zwislocki, 1978), whereas the effect of offset asynchrony takes about 80 ms to approach asymptote. Although the evidence is indirect, the influence of offset asynchrony on judgments of vowel quality strongly implies that a substantial part of the effect of onset asynchrony is attributable to perceptual grouping rather than to peripheral adaptation.

Darwin and Sutherland (1984) suggested that a leading component may make a reduced contribution to vowel quality because the leading portion of the component forms, with its continuation into the vowel, a separate perceptual group. If so, they reasoned that it should be possible to restore the contribution of the leading component to vowel quality by weakening the grouping between its leading portion and its continuation into the vowel. Their approach was to introduce an additional tone that was intended to form a perceptual group with the leading portion of the asynchronous component, and thus to free its continuation to integrate into the vowel percept. To encourage this *captor* tone to group with the leading portion, it was set to begin at the same time as the leading component and to end at vowel onset. The frequency of the captor tone was set to be one octave higher than the frequency of the leading component. This value was chosen so that any peripheral adaptation, or two-tone suppression (Houtgast, 1974), caused by the captor tone would be negligible at the frequency of the leading component. Although not explicitly stated, it is implied that the choice of a harmonic relationship between the captor and the leading component was also intended to encourage their perceptual fusion.

Darwin and Sutherland (1984) found that the presence of the captor tone partly reversed the reduction in the contribution of the leading component to vowel quality. This part-reversal cannot be explained in terms of peripheral adaptation. Furthermore, the fact that the extent of the reversal was limited to about one third does not necessarily imply that the rest of the effect of onset asynchrony is attributable to adaptation. This is because the physical continuity of the leading portion of the asynchronous component and its continuation into the vowel is likely to support some perceptual integration of the leading and continuing portions, despite the presence of the captor tone. Darwin and Sutherland's (1984) elegant study has since become a textbook example of how the perceptual effects of onset asynchrony reflect the operation of specific grouping mechanisms (see, e.g., Bregman, 1990; Darwin and Carlyon, 1995; Darwin, 1997; McAdams and Drake, 2002). However, the underlying assumption that the effect of the captor tone arises from its grouping with the leading portion of the asynchronous component, on the basis of their common onset time and their harmonic relationship, was not tested by Darwin and Sutherland and has not been tested since. The experiments reported here test this assumption explicitly and demonstrate that the effect of the captor tone does *not* depend on either of these grouping cues. Instead, it is proposed that the captor effect depends on inhibitory processes acting within the central auditory system.

II. GENERAL METHOD

A. Overview

Darwin has developed a sensitive paradigm for measuring the extent to which a harmonic in the F1 region is perceptually integrated into a vowel (Darwin, 1984a, b; Darwin and Sutherland, 1984). Darwin's paradigm is based on an approach pioneered by Carlson *et al.* (1975), and we have used it in all the experiments reported here. The British English vowels /ɪ/ and /ɛ/ do not differ greatly in their second and higher formants (see, e.g., Henton, 1983). Therefore, it is possible to construct a continuum of vowels, differing only in their F1 frequencies, that spans the /ɪ/-/ɛ/ phoneme boundary. Stimuli from this continuum can be presented to listeners in a forced-choice categorization task to provide an estimate of the position of the phoneme boundary.

The F1 frequency used to synthesize a vowel is defined as its *nominal* F1 value. If the effect of adding energy to a harmonic *raises* the perceived F1 frequency, then the boundary will move to a *lower* nominal F1 value on the continuum. Similarly, a reduction in perceived F1 frequency will cause the boundary to rise. For a given increase in the level of a harmonic in the F1 region, the extent to which a particular factor affects the perceptual integration of the extra energy can be measured by comparing the size of the boundary shifts across conditions.

Darwin (1984a, b) used vowels with an F0 frequency of 125 Hz and manipulated their third (375 Hz) and fourth (500 Hz) harmonics. Darwin (1984a) found that the extent to which the original vowel quality was recovered when the additional energy began before and continued after the vowel was greater when the fourth harmonic was manipulated. This effect was attributed to differences in masking. Specifically, it was argued that boosting the third harmonic masked the fourth harmonic for sounds near the phoneme boundary, and so prevented the fourth harmonic from reemerging as a fac-

tor influencing perceived F1 frequency. Darwin (1984b) used linear predictive coding to estimate perceived F1 frequency for vowels with a range of increments in level applied to the third or fourth harmonic. Discrepancies between these estimates and listeners' judgments indicated that the extra energy was not always fully integrated into the vowel percept, even when it was added in synchrony. This effect was quite marked when the energy was added at 375 Hz, below the F1 frequency at the phoneme boundary, but the 500-Hz component had to be boosted by 12 dB before any of the extra energy was excluded from the vowel percept. Darwin (1984b) argued that this difference may reflect a specific constraint on vocal-tract transfer functions. For these reasons, all the experiments reported here manipulated the level of the fourth harmonic.

B. Stimuli and conditions

The spectrum of each vowel was computed by adding harmonics with the amplitudes and phases given by the equations for the vocal-tract transfer function using cascade formant synthesis (Klatt, 1980). Vowel stimuli in the basic F1 continuum were synthesized with an F0 frequency of 125 Hz, and an F1 bandwidth of 90 Hz. All harmonics were present up to the fortieth (5 kHz). Formants 2–5 were set to frequencies of 2300, 2900, 3800, and 4600 Hz, with bandwidths of 120, 170, 1000, and 1000 Hz, respectively. There were 20 vowels per continuum, with F1 values spaced in equal steps of 10 Hz and ranging either from 350 to 540 Hz (experiments 1–3) or from 360 to 550 Hz (experiments 4 and 5). Vowel stimuli were 56 ms in duration (seven pitch periods), including linear onset and offset ramps of 16 ms each. To ensure roughly constant loudness across the basic continuum, all vowels were normalized to give the same root-mean-square (rms) voltage. The normalized basic continuum constituted the *vowel-alone condition*.

The continuum for the *incremented-fourth condition* was generated by adding in phase a pure tone at the frequency of the fourth harmonic (500 Hz) to each vowel in the normalized basic continuum. The additional tone had the same rise, steady-state, and fall times as the vowel and was added in synchrony. Its level was set to be 6 dB higher than the fourth harmonic, and so the 500-Hz component for each vowel was boosted by 9.5 dB relative to the value specified by the normalized transfer function. The continuum for the *leading-fourth condition* was generated by extending the additional tone so that it began before the vowel. The lead times used ranged from a few tens to a few hundreds of milliseconds across experiments. Figures 1(a) and 1(b) show schematic spectrograms representing stimuli from the incremented-fourth and leading-fourth conditions, respectively.

Captor conditions were generated by accompanying each vowel of the leading-fourth condition with a higher-frequency pure tone that was set to the same level as the additional 500-Hz tone. The onset and offset ramps for the captor were 16 ms each. Captor offset always began at vowel onset, and so its amplitude envelope reached zero at the same moment as the envelope of the vowel reached steady state. The captor tone began in synchrony with the leading fourth harmonic in some cases, but began before or after it in other cases. Captor-control conditions differed only in that the vowels were taken from the incremented-fourth condition, and so the leading portion of the fourth harmonic was absent. These controls were included to measure the extent to which perceived vowel quality is influenced by the captor tones in the absence of asynchronous energy at 500 Hz. Phoneme boundaries for these control conditions should not differ from the incremented-fourth condition if the captor produces no perceptual change in the vowel. Figures 1(c) and 1(d) show schematic spectrograms representing stimuli from the synchronous-captor and captor-control conditions, respectively.

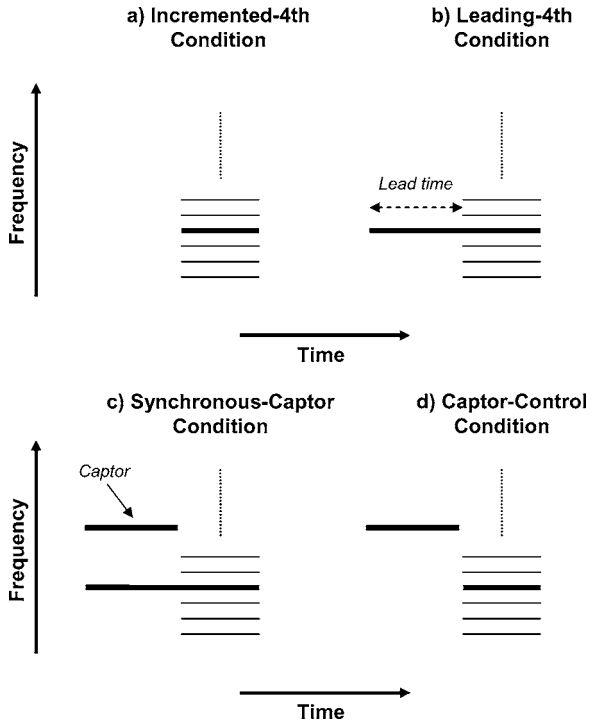


FIG. 1. Schematic spectrograms representing the stimulus arrangements for the following conditions: (a) incremented-fourth; (b) leading-fourth; (c) synchronous captor; (d) captor control. Each part displays harmonics in the F1 region of a vowel (the continuation of the spectrum is shown by a vertical dotted line). In each case, a tone has been added to the vowel to increase the level of its fourth harmonic. The perceptual consequences of this manipulation depend on the context (Darwin and Sutherland, 1984): (a) adding the tone in synchrony with the vowel changes its quality; (b) this change is largely abolished by an onset asynchrony; (c) adding the captor partly reverses the effect of making the tone asynchronous; (d) adding the captor does not alter the phonetic change caused by adding the tone in synchrony with the vowel.

All stimuli were generated using MITSYN software (Henke, 1997). Stimuli were synthesized at a sampling rate of 16 kHz, stored on disk, and played back via a 16-bit digital-to-analog converter (Data Translation DT2823). They were low-pass filtered (Fern Developments EF16X module; corner frequency = 5.2 kHz, roll-off = 100 dB/oct) and presented diotically over Sennheiser HD 480-13II earphones. Stimuli were checked using a spectrum analyzer (Stanford Research Systems SR780). The overall output level was set using a programmable attenuator (0.25-dB steps) and calibrated using a sound-level meter (Brüel and Kjaer, type 2209) coupled to the earphones by an artificial ear (type

4153). The presentation level for vowels from the basic continuum was about 69 dB SPL. The maximum presentation level for a vowel from the incremented-fourth continuum was about 76 dB SPL.

C. Listeners

Twelve listeners took part in each experiment. All volunteers were native speakers of British English who reported no hearing impairment. They were paid for their assistance and had earlier successfully completed a screening procedure. Initial screening was undertaken to ensure that each listener was able to distinguish between the two vowel categories and to produce a monotonic change in the identification function as the F1 value increased across a continuum. Listeners first identified two randomized blocks of the full set of 20 stimuli in the vowel-alone condition to familiarize themselves with the task. These could be repeated if required. Listeners then identified ten randomized blocks of tokens from the vowel-alone and incremented-fourth conditions (400 trials). To pass the screening, listeners also had to reach a criterion shift in phoneme boundary for the incremented-fourth condition compared with the vowel-alone condition. The criterion was a minimum downward shift of 20 Hz, confirming that boosting the 500-Hz component increased the perceived F1 frequency of vowels along the continuum. About 75% of volunteers passed the screening and were free to take part in as many or as few experiments as they wished. An extended screening, including an additional and stricter criterion, was used for listeners who took part in experiments 4 and 5. This is described in Sec. VI A.

D. Procedure

Stimuli were played to the listeners in a double-walled sound-attenuating chamber (Industrial Acoustics 1201A). On each trial, listeners responded using a keyboard to indicate whether the vowel token was perceived as /ɪ/ or /ɛ/. The printed instructions for listeners gave the words "hid" and "head" as examples in "h-d" context of the two phonemic categories. Except for practice blocks, described below, listeners were not allowed to repeat a stimulus before making their identification. The next trial began 0.5 *s* after the return key was pressed to confirm the choice. Listeners were asked to note the trial number, displayed on screen, for any accidental key presses. These were then corrected manually in the data file before analysis. Listeners were asked to focus on the vowel and to try to ignore any additional sounds that might be present. No feedback was provided.

A block of stimuli consisted of all 20 tokens for every condition in that experiment. The number of trials per block varied from 140 to 340 trials, depending on the number of conditions. The stimulus order was randomized anew for each repetition of a block, and ten repetitions were presented to listeners over the course of an experiment. As stimuli from all conditions were intermingled in a block, any difference in the position of the phoneme boundary between conditions cannot be attributed to range effects (Brady and Darwin, 1978). The use of randomized order also precludes the possibility of systematic contrast effects influencing the boundary position.

Each experiment was run as one or more sessions that lasted between 40 and 70 min. Every session began with a practice block that consisted of a subset of the stimuli from the main experiment. The subset included one token of either every second or every fourth F1 step from the full set, depending on the number of conditions in each experiment, and comprised between 70 and 110 randomized trials. A typical session comprised the practice block and about 1800 experimental trials. Listeners typically took breaks after every few hundred trials.

E. Data analysis

The data for each listener comprised the number of /ɪ/ responses out of ten for each nominal F1 value and condition. Following previous practice (e.g., Darwin, 1984a; Roberts and Moore, 1990), each data set was fitted individually by probit analysis (Finney, 1971). The nominal F1 frequency at the /ɪ/-/ɛ/ phoneme boundary was defined as the mean of the probit function (50% point). The effect on vowel timbre of the additional energy at 500 Hz was estimated for each condition by measuring the shift in boundary position relative to the reference condition (vowel alone). The probit function also provided an estimate of the slope of the identification function. In all experiments, these slopes did not change substantially across conditions and so are not considered further.

Occasionally, a misclassification would occur for a stimulus clearly some distance from the crossover region of the identification function, giving a "9-1" or "1-9" pattern of /ɪ/-/ɛ/ responses for that stimulus, rather than a "10-0" or "0-10" pattern. Although this did not greatly affect the probit estimate of boundary position, it could affect the estimate of slope. Therefore, following Roberts and Moore (1990, 1991), these response patterns could be changed to "10-0" or "0-10," as appropriate, before performing the probit analysis. The criterion for making the change was that at least the neighboring two continuum members on both sides gave a "10-0" or "0-10" response pattern. Overall, about 20% of the listeners who completed an experiment were replaced because they obtained, for one or more conditions, either a nonmonotonic identification function or a boundary estimate that fell outside the range of nominal F1 values tested.

Once data collection was completed, a within-subjects analysis of variance (ANOVA) was performed on the boundary estimates. Given the small number of pairwise comparisons required from the large set of possible two-tailed comparisons in each experiment, they were computed using the restricted least-significant-difference (LSD) test (Snedecor and Cochran, 1967). In most cases, the only pairwise comparisons quoted are for the following conditions: vowel-alone versus incremented-fourth, incremented-fourth versus leading-fourth, and vowel-alone versus leading-fourth.

Of central interest in experiments 1–4 is the extent to which the *asynchrony-induced reduction* in the effect of the added 500-Hz tone is reversed by the presence of a captor.

This is called the *restoration effect* and its value, in Hz, is defined as the difference between the boundaries for the captor condition and its control subtracted from the difference between the boundaries for the leading-fourth and incremented-fourth conditions. The computation compares each captor condition with its control (rather than with the incremented-fourth condition) to take account of any effect that the captor may have on perceived vowel quality, other than through its modulating influence on the phonetic effect of the leading fourth harmonic.

To establish whether or not different captor conditions produced different amounts of restoration, a within-subjects ANOVA was performed on the transformed data. As before, any subsequent pairwise comparisons were performed using the restricted LSD test. It was also important to demonstrate that the restoration effects obtained were significantly greater than zero. Unless the ANOVA revealed a significant effect of captor condition on the restoration values, this was achieved by calculating a mean value across conditions for each listener and performing a single one-sample t test (two tailed) to compare the overall mean with zero. Otherwise, a separate one-sample t test was performed for each condition.

III. EXPERIMENT 1

Darwin and Sutherland (1984) demonstrated that accompanying the leading portion of additional energy in the F1 region with a captor tone partly reversed the effect of the asynchrony. They argued that the effect of the captor depends primarily on its common onset with the leading 500-Hz tone. Experiment 1 tested this hypothesis by replicating Darwin and Sutherland's experiment and extending it to include two additional kinds of captor tone, one which began before the leading 500-Hz tone (advanced-onset captor) and one which began after it (delayed-onset captor). In terms of the nominal F1 value along the /ɪ/-/ɛ/ continuum, it was expected that the phoneme boundary would be lower in the incremented-fourth condition than in the vowel-alone condition, that it would shift back towards its original value in the leading-fourth condition, and that (relative to its control) the extent of this upward shift would be reduced in the synchronous-onset captor condition. If the asynchronous-onset captors are found to be as effective as the synchronous-onset captor at restoring the contribution of the leading 500-Hz tone, then captor efficacy must depend on factors other than the grouping of the captor with the leading tone on the basis of common onset.

A. Method

There were nine conditions in total. These were the three standard conditions (vowel alone, incremented fourth, leading fourth), plus three captor conditions and their controls. The lead time on the added 500-Hz tone was 240 ms (cf. Darwin and Sutherland, 1984). The captor conditions were created by adding a pure tone at 1000 Hz to the stimuli from the leading-fourth continuum. For each member of the continuum, the level of the 1-kHz tone was the same as that of the added 500-Hz tone. These parameters are identical to those used by Darwin and Sutherland (1984). The captor either began at the same time as the leading 500-Hz component (synchronous onset), 160 ms before it (advanced onset), or 160 ms after it (delayed onset). Hence, the durations of the captors for these conditions were 240, 400, and 80 ms, respectively. In all cases, captor amplitude was ramped down to zero at the same time as vowel amplitude was ramped up to steady state. Each captor condition had a corresponding control condition, for which the stimuli differed only in that the leading portion of the added 500-Hz tone was absent. The number of trials per block in the main experiment was 180 (9 conditions $\times$ 20 F1 values).

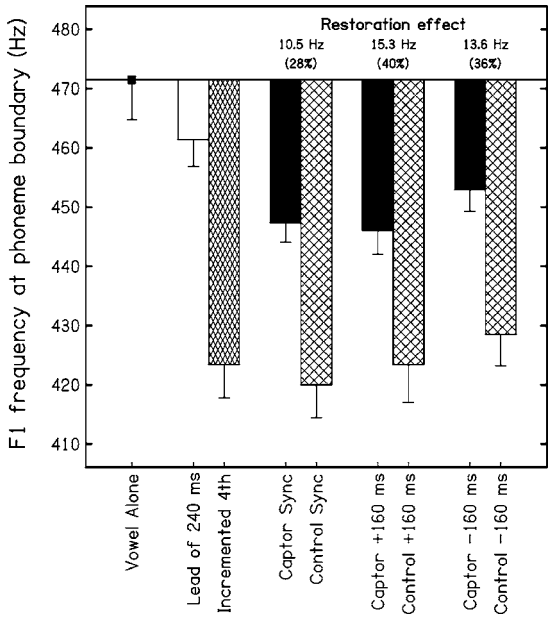


FIG. 2. Results for experiment 1—effect of captor asynchrony for a long lead-time on the fourth harmonic (240 ms). Mean phoneme boundaries are expressed as nominal first-formant frequencies for 12 listeners (with intersubject standard errors). The boundary for the reference condition (vowel alone) is indicated by the horizontal line and the square symbol. The boundaries for the leading-fourth condition and the incremented-fourth condition are shown by the open bar and narrow-hashed bar, respectively. The boundaries for the captor conditions and for their controls are shown by the solid bars and wide-hashed bars, respectively. The extent to which each captor condition restores the effect of the leading fourth on vowel quality, relative to its own control, is shown above each pair of bars. Each restoration effect is quoted as an F1 boundary shift in Hz and as a percentage of the boundary difference between the leading-fourth and incremented-fourth conditions.

B. Results and discussion

Figure 2 displays the mean phoneme boundaries (and intersubject standard errors) across conditions in terms of nominal F1 frequency. Mean restoration effects for the captor conditions are shown above the corresponding histogram bars. Each value is quoted as an F1 boundary shift in Hz and also as a percentage of the boundary difference between the leading-fourth and incremented-fourth conditions. The effect of condition on the position of the phoneme boundary was highly significant $[F(8,88)=27.732,\ p<0.001]$. The boundary for the incremented-fourth condition was lower than the boundary for the vowel-alone condition $[t(11)=13.358,\ p<0.001]$, indicating that perceived F1 frequency was raised when extra energy was added in synchrony to the fourth harmonic. The size of the mean bound-

ary shift was 48 Hz. The boundary for the leading-fourth condition was higher than the boundary for the incremented-fourth condition [$t(11)=6.315$, $p<0.001$], but did not differ from the boundary for the vowel-alone condition [$t(11)=1.511$, $p=0.159$]. This indicates that the extra energy at 500 Hz was excluded from the perceptual computation of F1 frequency when it began 240 ms before the vowel.

For ease of comparison, our preliminary analysis of the effects of the captor conditions followed the approach of Darwin and Sutherland (1984). This analysis showed that all captor conditions gave significantly lower boundaries than the leading-fourth condition [synchronous onset: $t(11)=3.908$, $p=0.002$; advanced onset: $t(11)=3.923$, $p=0.002$; delayed onset: $t(11)=2.941$, $p=0.013$]. There was one captor-control condition for which the boundary position was significantly different from the incremented-fourth condition [delayed-onset: $t(11)=3.122$, $p=0.010$]. This occurrence indicates the importance of including captor-control conditions in the experiment.

The efficacy of the three types of captor was compared using the computed restoration values (see Sec. II E). Restoration values describe the extent of the boundary shift induced by the captor when the effect of the corresponding captor control has been factored out. Analysis of the restoration effects revealed that there was no significant difference between the three captor conditions [$F(2,22)=1.149$, $p=0.335$], and that the overall magnitude of the restoration effect was significantly above zero [$t(11)=4.084$, $p<0.001$]. This suggests that the captors were equally effective in weakening the effect of applying a lead time of 240 ms to the added 500-Hz tone, irrespective of the onset time of the captor. The overall mean restoration effect was about 35%. Note that the failure to find a significant reduction in captor efficacy when the onset of the captor leads or lags the onset of the 500-Hz component cannot be attributed to a lack of statistical power, because the observed magnitude of the restoration effect was actually smallest for the synchronous-captor condition.

The magnitude of the boundary shift induced by boosting the fourth harmonic of the vowel by 9.5 dB, the near-complete return of the boundary when the additional energy begins 240 ms before the vowel, and the extent to which the captor reverses the boundary return are all very similar to the findings of Darwin and Sutherland (1984). However, the absence of an effect of captor onset time does not support their proposal that the restoration induced by the captor depends on a common onset for the captor and the leading 500-Hz component. Note that captor duration covaried with captor onset time, ranging from 400 ms (advanced-onset case) to 80 ms (delayed-onset case). Hence, the comparable restoration values for the three captor types suggests that temporal integration of captor energy is complete by 80 ms.

IV. EXPERIMENT 2

The absence of an effect of relative captor onset time in experiment 1 is consistent with the idea that captor efficacy does not depend on grouping by common onset. However, the use of a long lead time (240 ms) for the added 500-Hz tone leaves open the possibility that the perceptual consequences of a difference in onset time had "worn off" long before the vowel began. If so, the captor may have grouped with the leading portion of the added tone on the basis of their octave relationship. Indeed, if one listens to a harmonic complex for which one component begins before the others, the salient pitch heard for that component within the complex begins to fade as time passes. Therefore, as a precaution, we compared the efficacy of synchronous and asynchronous captors in the context of a short lead time (40 ms) on the added 500-Hz tone. If these captors are similarly effective under these conditions, then we can rule out the alternative proposal and conclude that grouping by common onset does not mediate the captor effect.

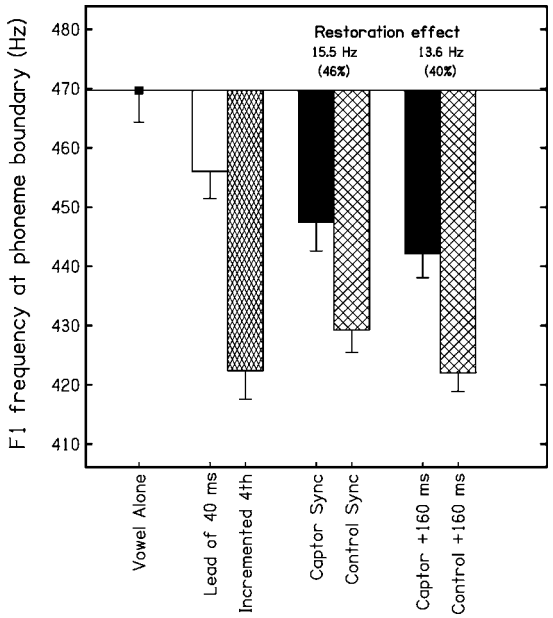


FIG. 3. Results for experiment 2—effect of captor asynchrony for a short lead-time on the fourth harmonic (40 ms). Phoneme boundaries and restoration effects are displayed as for Fig. 2.

A. Method

There were seven conditions—the standard three plus two captor conditions and their controls. The lead time on the added 500-Hz tone was set to 40 ms. The 1-kHz captor either began at the same time as the leading 500-Hz component (synchronous onset), or 160 ms before it (advanced onset). Hence, the durations of these captors were 40 and 200 ms, respectively. The number of trials per block in the main experiment was 140 (7 conditions $\times$ 20 F1 values).

B. Results and discussion

Figure 3 displays the mean phoneme boundaries for each condition, with intersubject standard errors, and restoration effects for each captor condition. The effect of condition on the position of the phoneme boundary was highly significant [$F(6,66)=32.543$, $p<0.001$]. The boundary for the incremented-fourth condition was lower than the boundary for the vowel-alone condition [$t(11)=9.694$, $p<0.001$], indicating that perceived F1 frequency was raised when extra

energy was added in synchrony to the fourth harmonic. The size of the mean boundary shift was 47 Hz. The boundary for the leading-fourth condition was higher than the boundary for the incremented-fourth condition [$t(11)=5.242$, $p<0.001$], but was still lower than the boundary for the vowel-alone condition [$t(11)=2.577$, $p=0.026$]. This indicates that most, but not all, of the extra energy at 500 Hz was excluded from the perceptual computation of F1 frequency when it began 40 ms before the vowel.

Analysis of the restoration effects revealed that there was no significant difference between the two captor conditions [$F(1,11)=0.499$, $p=0.495$], and that the overall magnitude of the restoration effect was significantly above zero [$t(11)=4.672$, $p<0.001$]. This suggests that both the synchronous-onset and the advanced-onset captors were equally effective in weakening the effect of applying a lead time of 40 ms to the added 500-Hz tone. The overall mean restoration effect was about 43%, which is at least as large as the overall mean for experiment 1.

An asynchrony of 40 ms is almost as effective at excluding the additional energy at 500 Hz from the perceptual estimation of F1 frequency as the lead time of 240 ms used in experiment 1. This is consistent with the findings of Roberts and Moore (1991). They added tones to the F1 region, at frequencies other than harmonics of F0, and showed that the reduction in the effect on vowel quality when the extraneous energy began before the vowel approached asymptote around a lead time of 40 ms. Darwin (1984b) and Darwin and Sutherland (1984) found that an onset-time difference of only 32 ms was sufficient largely to abolish the phonetic effects of boosting the fourth harmonic (500 Hz) of a vowel. Furthermore, Rasch (1978) found that a lead time of 30 ms was sufficient for a signal tone to be perceived as continuing through a masking tone. Clearly, the auditory system can utilize small differences in the onset times of acoustic events to segregate them perceptually.

Darwin and Sutherland's (1984) study did not include any short-duration captors. The size of the restoration effect brought about by the synchronous-onset captors (duration =40 ms) was comparable with that brought about by the advanced-onset captors (duration=200 ms) and also by the long captors (duration=up to 400 ms) used in experiment 1. This extends the findings of experiment 1 by suggesting that temporal integration of captor energy is essentially complete by 40 ms. However, the most important outcome is the replication of the finding that relative onset-time does not affect captor efficacy. The later of the two onsets determines the point at which the auditory system can detect an asynchrony between the 500-Hz tone and the captor tone. Only 40 ms elapsed from this point to the onset of the vowel in the current experiment, whereas, in experiment 1, 80 ms elapsed for the delayed-onset captors and 240 ms elapsed for the advanced-onset captors. The short time from captor onset to vowel onset in the current experiment makes it highly unlikely that the effect of the asynchrony had decayed away by the time the vowel began. Therefore, we can conclude that the similar efficacy of synchronous and asynchronous captors is a genuine effect.

V. EXPERIMENT 3

Darwin and Sutherland's (1984) choice of a captor frequency one octave above the leading 500-Hz tone means that their captor tone may have formed a perceptual group with the leading portion of the added tone on the basis of harmonic relations. Experiments 1 and 2 manipulated the relative onset time of the captor and the leading tone, but maintained this octave relationship. Therefore, it is possible that captor efficacy in all of these experiments depended on grouping by harmonic relations rather than on grouping by common onset. Experiment 3 explored this possibility by including some captors that were harmonically related to the leading tone and some that were not. The range of captor frequencies used exceeded two octaves, and so also allowed the effect of frequency separation between the captor and the leading tone to be measured. The lowest-frequency captor used was 900 Hz, which is about four ERBs [equivalent rectangular bandwidths (Glasberg and Moore, 1990)] above the leading tone. This lower limit was set to ensure that any spread of excitation from the captor would not produce any significant adaptation in auditory nerve fibers tuned to 500 Hz.

A. Method

There were 17 conditions—the standard three plus seven captor conditions and their controls. The lead time on the added 500-Hz tone was set to 240 ms, and the captor always began at the same time as the leading 500-Hz component (synchronous onset). Captor frequencies were set to be multiples of either 450 Hz or 500 Hz, only the latter of which are harmonically related to the leading fourth harmonic. The set of values used was 900, 1000, 1350, 1500, 1800, 2000, and 2250 Hz. The range of frequencies used ensured that each harmonic value tested was bracketed by inharmonic values. The 1-kHz captor case is identical to the synchronous-onset captor case in experiment 1. The number of trials per block in the main experiment was 340 (17 conditions $\times$ 20 F1 values).

B. Results

Figure 4(a) displays the mean phoneme boundaries for each condition, with intersubject standard errors. The "H" symbols indicate cases where the captor frequency was harmonically related to the leading 500-Hz component. The effect of condition on the position of the phoneme boundary was highly significant [$F(16,176)=23.058$, $p<0.001$]. The boundary for the incremented-fourth condition was lower than that for the vowel-alone condition [$t(11)=10.144$, $p<0.001$], indicating that perceived F1 frequency was raised when extra energy was added in synchrony with the fourth harmonic. The size of the mean boundary shift was 50 Hz. The boundary for the leading-fourth condition was higher than that for the incremented-fourth condition [$t(11)=5.501$, $p<0.001$], but did not differ from the boundary for the vowel-alone condition [$t(11)=0.308$, $p=0.764$]. As for experiment 1, this indicates that the extra energy at 500 Hz was excluded from the perceptual computation of F1 frequency when it began 240 ms before the vowel.

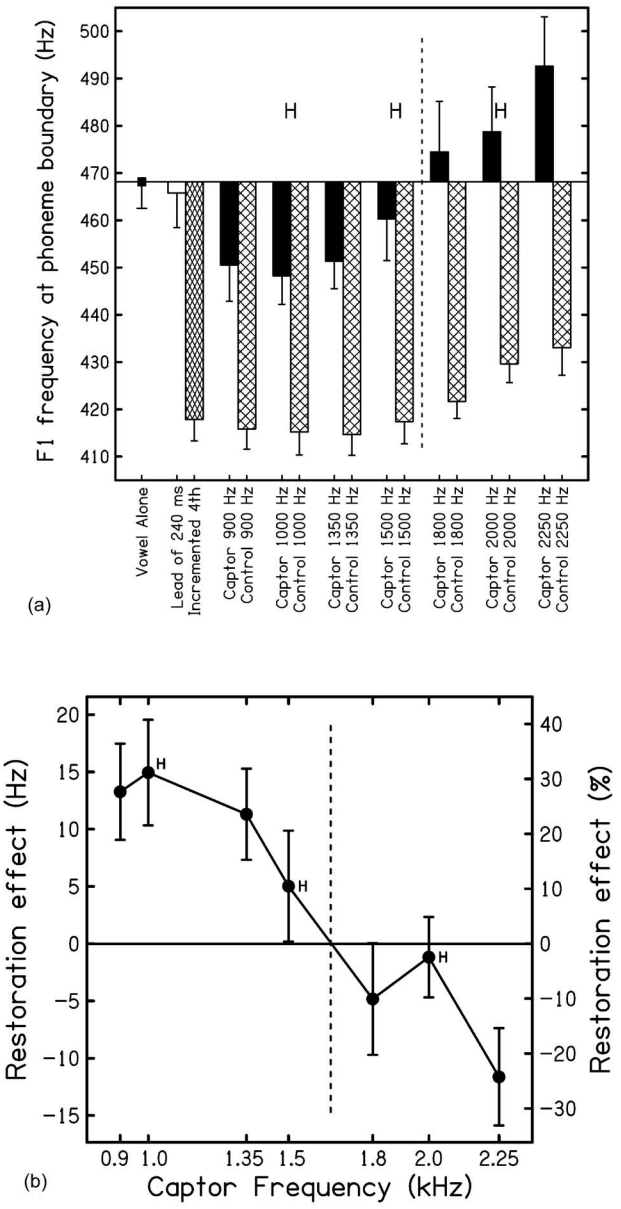


FIG. 4. Results for experiment 3—effect of captor frequency and harmonic relations with respect to the leading fourth harmonic. Part (a) displays the phoneme boundaries as for Fig. 2. The H symbols indicate captor conditions for which the captor frequency was harmonically related to the leading fourth harmonic. The vertical dotted line divides the captor conditions for which the control boundaries are close to the incremented-fourth boundary (on the left) from those for which the control boundaries drift towards higher F1 values (on the right). See text for a full discussion of the underlying basis for this distinction. Part (b) displays the restoration effects across captor frequency as F1 boundary shifts in Hz (left axis) and as a percentage of the boundary difference between the leading-fourth and incremented-fourth conditions (right axis). The H symbols and the vertical dotted line correspond to their counterparts in part (a).

The boundaries for the captor controls were very similar to the boundary for the incremented-fourth condition for frequencies up to 1500 Hz, but drifted upwards for frequencies of 1800 Hz and above. This distinction is indicated by the vertical dotted line in Fig. 4(a). The upward drift in boundary for the captor controls was significant for the highest two frequencies tested [2000 Hz: $t(11)=2.770$, $p=0.018$; 2250 Hz: $t(11)=2.444$, $p=0.033$]. Furthermore, in contrast with the usual pattern, the boundary shifts for the corresponding captor conditions were *negative* (i.e., upward) for frequencies of 1800 Hz and above.

Figure 4(b) displays the mean restoration effects for each captor condition, with intersubject standard errors. The "H" symbols and the vertical dotted line correspond to their counterparts in Fig. 4(a). Analysis of the restoration values revealed a highly significant effect of captor frequency [$F(6,66)=8.285$, $p<0.001$]. Restoration values were significantly above zero for captor frequencies below 1500 Hz [900 Hz: $t(11)=3.161$, $p=0.009$; 1000 Hz: $t(11)=3.244$, $p=0.008$; 1350 Hz: $t(11)=2.842$, $p=0.016$]. This indicates that the ability of the captor partly to reverse the effect of starting the added 500-Hz tone before the vowel is frequency dependent. The restoration effect was about 25%–30% for captor frequencies up to 1350 Hz, but declined to zero as the captor frequency was increased to about 1500–1800 Hz. Pairwise comparisons showed that the restoration values for the lower-frequency captors (900, 1000, and 1350 Hz) did not differ from one another, but they all differed ($p<0.05$) from the restoration values for the higher-frequency captors (1800, 2000, and 2250 Hz). The restoration effect for the highest-frequency captor (2250 Hz) was the only value to fall significantly below zero [$t(11)=-2.735$, $p=0.019$]. Captor efficacy does not appear to depend on a harmonic relationship between the captor and the added tone.

C. Discussion

First, we consider the effects of captor frequencies in the range 900–1500 Hz, for which the boundaries for the corresponding captor controls were very close to the boundary for the incremented-fourth condition. Next, we consider the effects of captor frequencies in the range 1800–2250 Hz, for which the corresponding control boundaries are shifted upwards relative to the incremented-fourth condition. Finally, we consider the extent to which harmonic relations between the captor and the leading fourth harmonic may influence captor efficacy.

The findings for the lower-frequency captors indicate that frequency separation is a key factor governing the magnitude of the restoration effect. The effect of the captor appears to be broadband with an upper cutoff frequency of about 1500 Hz. This indicates that the response to the captor extends up to about one-and-a-half octaves above the frequency of the leading fourth harmonic. Note that this estimate of the bandwidth of the captor response is not affected by the atypical boundary shifts observed for the higher-frequency captors, because captor efficacy clearly declines in the 1350–1500-Hz region, before the control boundaries show any clear sign of upward drift with respect to the incremented-fourth condition. The significance of the observed bandwidth effect is considered further in Sec. VIII.

The findings for the higher-frequency captors illustrate the importance of including captor controls in the experiment. The upward drift in the boundaries of the captor controls for frequencies of 1800 Hz and above indicates that the influence of these captors on perceived vowel quality was not solely through modulating the effect of asynchrony on the contribution of the boosted fourth harmonic. How else

might these captors have affected vowel quality? The test range used was chosen to avoid the spread of adaptation to frequencies close to F1, but the higher-frequency captors used will have adapted the lower skirt of the second formant, which began around 1800 Hz for our stimuli. The degree of adaptation may have been considerable, given that the captors were set to be 6 dB above the original level of the fourth harmonic, and that the intensity of F2 was below that of F1.

We used a fixed F2 frequency of 2300 Hz for all our stimuli, but spoken examples of /ɛ/ typically have a lower F2 frequency than spoken examples of /ɪ/ (see, e.g., Henton, 1983). Therefore, the rise in perceived F2 frequency resulting from adapting the lower skirt of F2 may have made our stimuli sound more /ɪ/-like, leading to a rise in phoneme boundary. The finding that the highest-frequency captor (2250 Hz) was associated with the greatest upward shift in the control boundary seems to challenge this idea, because here the perceptual effect of the adaptation would have been to flatten the F2 peak rather than to raise the F2 frequency. However, the cluster of higher formants in front vowels is typically integrated perceptually into a single energy peak called F2-prime (Chistovich *et al.*, 1979; Chistovich, 1985). The center-of-gravity of this cluster would be raised by flattening the F2 peak. The finding of a negative restoration effect for the highest-frequency captor indicates that the inclusion of captor controls is not always entirely successful in partialling out unwanted effects. Perhaps the influence of the captor on perceived F1 (via modulation of the effect of asynchrony) and its influence on perceived F2-prime (via adaptation) are not strictly additive effects.

Darwin and Sutherland's (1984) choice of an octave relationship between the captor and the leading portion of the added tone in principle may have provided another cue for grouping them together. However, the results of the current experiment do not support this hypothesis. Not only is it clear that the captor does not need to be harmonically related to the leading component to produce a restoration effect, but there is also little or no evidence that captor efficacy is increased by this kind of harmonic relationship. Rather, the influence of frequency on captor efficacy appears to depend only on spectral separation. Nonetheless, it should be acknowledged that the use of pure-tone captors does not provide a very strong test of the effect of harmonic relations. Bregman (1990) has argued that at least three components must be present before their simultaneous grouping can be influenced by harmonic relations. In support of this idea, he noted that Bregman and Doehring (1984) had shown that a partial could be captured from a three-component complex into a sequential stream more easily when it was mistuned relative to the others, whereas Bregman and Pinker (1978) had not found such an effect for two-tone complexes. A more rigorous test of whether or not harmonic relations can influence captor efficacy would require the use of multicomponent captors. This idea is considered further in Sec. VIII.

VI. EXPERIMENT 4

Experiments 1–3 have shown that captor efficacy does not depend on the formation of a perceptual group between the captor and the leading portion of the 500-Hz tone on the basis of either common onset or harmonic relations. However, in all of these experiments, one or other of these grouping factors was left undisrupted. Therefore, to rule out the possibility that the remaining factor favoring the grouping of the captor and the leading tone was sufficient to maintain captor efficacy, we included a condition in which the captor had neither a common onset nor harmonic relations with the leading tone.

A. Method

There were 11 conditions—the standard three plus four captor conditions and their controls. The lead time on the added 500-Hz tone was set to 160 ms. The four captor conditions were derived from a combination of two binary-state variables: (a) captor began at the same time as the leading 500-Hz component (synchronous onset, captor duration $=160$ ms) or 80 ms earlier (advanced onset, captor duration $=240$ ms) and (b) captor frequency was set to 1000 Hz (harmonic case) or 900 Hz (inharmonic case). The number of trials per block in the main experiment was 220 (11 conditions $\times$ 20 F1 values).

The procedure used to screen for listeners to take part in experiments 4 and 5 was more stringent than that used in the earlier experiments. The original screening used only the vowel-alone and incremented-fourth continua. Experiments 1–3 established that asynchrony can produce a clear reduction in the effect of added energy at 500 Hz, and that the effect of the extra energy can be partly restored by accompanying the leading portion with a captor tone. Therefore, the screening and training was extended to include a leading-fourth condition (lead time$=240$ ms), a captor condition (1 kHz, synchronous onset), and its control. Listeners identified ten randomized blocks of tokens from these five conditions (1000 trials). The additional criterion was that a minimum restoration effect of 5 Hz was required for listeners to proceed to the main experiment. This criterion further excluded about 10% of those volunteers who would have passed the original screening. However, it largely eliminated the subsequent exclusion of data from the main experiment based on obtaining a nonmonotonic identification function, or a boundary estimate that fell outside the test range, for one or more of the conditions.

B. Results and discussion

Figure 5 displays the mean phoneme boundaries for each condition, with intersubject standard errors, and restoration effects for each captor condition. The effect of condition on the position of the phoneme boundary was highly significant [$F(10,110)=49.971$, $p<0.001$]. The boundary for the incremented-fourth condition was lower than the boundary for the vowel-alone condition [$t(11)=15.031$, $p<0.001$], indicating that perceived F1 frequency was raised when extra energy was added in synchrony to the fourth harmonic. The size of the mean boundary shift was 51 Hz. The boundary for the leading-fourth condition was higher than the boundary for the incremented-fourth condition [$t(11)=8.4000$, $p<0.001$], but did not differ from the boundary for the vowel-

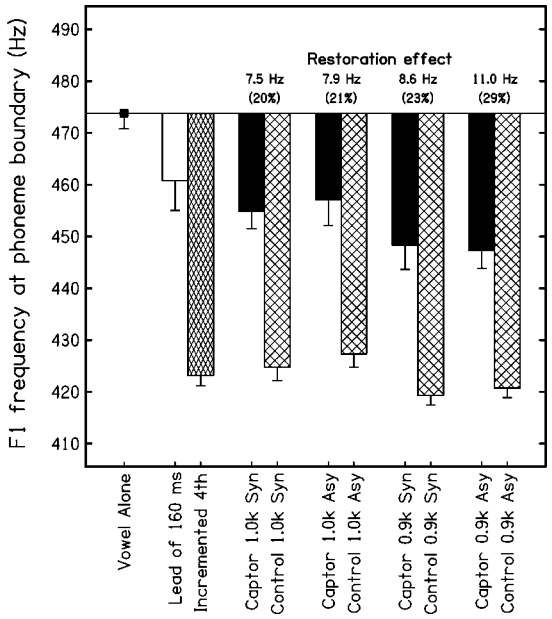


FIG. 5. Results for experiment 4—effect of captor asynchrony and inharmonicity. Phoneme boundaries and restoration effects are displayed as for Fig. 2.

alone condition [$t(11)=2.027$, $p=0.068$]. This indicates that the extra energy at 500 Hz was excluded from the perceptual computation of F1 frequency when it began 160 ms before the vowel.

Analysis of the restoration effects revealed that there was no significant difference between the four captor conditions [$F(3,33)=0.470$, $p=0.706$]. The overall magnitude of the restoration effect (about 8.7 Hz, or 23%) was smaller than that observed in experiments 1 and 2 and was not quite significantly greater than zero [$t(11)=2.003$, $p=0.071$]. However, an inspection of the individual data identified one listener who obtained negative restoration values for all four captor conditions, owing to an unusually small boundary return for the leading-fourth condition. For the other 11 listeners, the overall magnitude of the restoration effect was significantly greater than zero [$t(10)=3.034$, $p=0.013$].

The results suggest that the four types of captor were equally effective in weakening the effect of applying a lead time of 160 ms to the added 500-Hz tone. Note that the failure to find a significant reduction in efficacy when the captor is both asynchronous and inharmonic cannot be attributed to a lack of statistical power, because the observed magnitude of the restoration effect was actually greatest for this condition. Therefore, we can rule out the possibility that captor efficacy was maintained in the earlier experiments because the captors used retained *at least* either a common onset or a harmonic relationship with the leading 500-Hz tone. It is clear that captor efficacy, both in our study and in the study of Darwin and Sutherland (1984), does *not* depend on the grouping of the captor with the leading 500-Hz tone on the basis of common onset or harmonicity.

VII. EXPERIMENT 5

Whatever the basis of the restoration effect, it is clear that asynchrony is less effective at excluding an added 500-Hz tone from perceived vowel quality when the leading portion is accompanied by a captor tone. Darwin (1995) has shown that asynchrony is also less effective at perceptually excluding the added tone when its leading portion is reduced in level. The greater the attenuation, the greater the downward shift in boundary towards the value for the incremented-fourth condition. The extent of this *boundary return* provides a measure of the effect of the attenuation that is closely analogous to the magnitude of the restoration effect as a measure of captor efficacy. Hence, finding the attenuation required to produce a boundary return of similar size to a typical restoration effect (about a third) allows the effect of the captor to be expressed as equivalent to applying that level of attenuation to the leading portion of the fourth harmonic. Darwin's (1995) stimuli were similar to ours and included instances equivalent to attenuating the leading portion of the fourth harmonic by 6 dB and by 15 dB. Interpolating between these values suggests that the effect of the captors that we used in experiments 1–4 was equivalent to attenuating the leading portion of the fourth harmonic by about 9 dB. Our informal observations were generally consistent with Darwin's, but suggested that the boundary return produced by a given level of attenuation may have been greater than was apparent in his data. Therefore, we tested a full set of listeners using parameters for the vowel stimuli that were identical to those used in experiment 4.

A. Method

There were seven conditions—the standard three plus four conditions in which the level of the *leading portion* of the added 500-Hz tone was attenuated by 6, 12, 18, or 24 dB, with respect to the leading-fourth condition. The linear increase in amplitude at the end of the leading portion of the added tone, to reach its steady-state value, was simultaneous with the onset ramp for the vowel. No captor tones were present in any of the conditions. The lead time on the added 500-Hz tone was set to 160 ms. The number of trials per block in the main experiment was 140 (7 conditions conditions $\times$ 20 F1 values).

B. Results and discussion

Figure 6 displays the mean phoneme boundaries for each condition, with intersubject standard errors. It can be seen that attenuating the leading portion of the added 500-Hz tone causes the boundary to shift downwards from its value for the leading-fourth condition towards its value for the incremented-fourth condition. Mean boundary returns are displayed for each level of attenuation above the corresponding histogram bars. Each value is quoted as an F1 boundary shift in Hz and also as a percentage of the boundary difference between the leading-fourth and incremented-fourth conditions.

The effect of condition on the position of the phoneme boundary was highly significant [$F(6,66)=48.136$, $p<0.001$]. The boundary for the incremented-fourth condition was lower than the boundary for the vowel-alone condition [$t(11)=16.837$, $p<0.001$, indicating that perceived F1 frequency was raised when extra energy was added in syn-

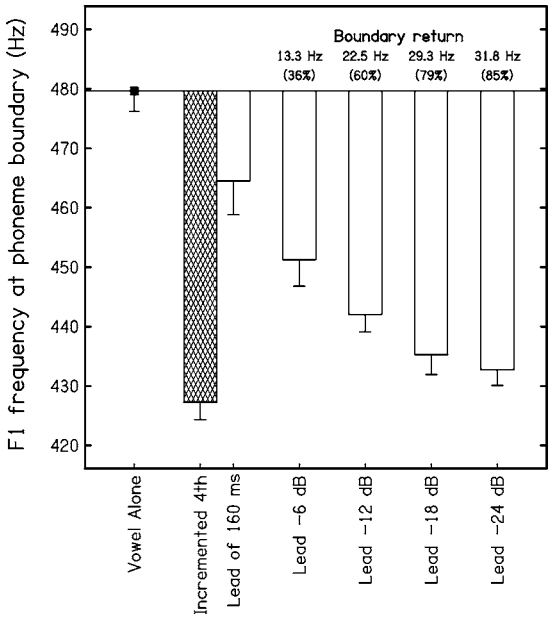


FIG. 6. Results for experiment 5—effect of attenuating the leading portion of the fourth harmonic. Phoneme boundaries for the reference condition (vowel alone) and for the incremented-fourth condition are displayed as for Fig. 2. From left to right, the open bars show the boundaries for the leading-fourth conditions as the attenuation of the leading portion is increased from 0 to 24 dB in 6-dB steps. The extent to which attenuating the leading portion returns the boundary to that for the incremented-fourth condition is quoted as an F1 boundary shift in Hz and as a percentage of the boundary difference between the leading-fourth (0 dB) and incremented-fourth conditions.

chrony to the fourth harmonic. The size of the mean boundary shift was 52 Hz. The boundary for the leading-fourth condition was higher than the boundary for the incremented-fourth condition [$t(11)=7.005$, $p<0.001$], but was still lower than the boundary for the vowel-alone condition [$t(11)=2.354$, $p=0.038$]. This indicates that most, but not all, of the extra energy at 500 Hz was excluded from the perceptual computation of F1 frequency when it began 160 ms before the vowel.

To assess the effect of attenuation on the extent of the boundary return, a within-subjects ANOVA was performed on the transformed data. Analysis of the boundary returns revealed a highly significant effect of level of attenuation [$F(3,33)=22.488$, $p<0.001$]. Pairwise comparisons with the leading-fourth condition indicated that even the smallest attenuation used (6 dB) caused a significant shift in the boundary back towards its position for the incremented-fourth condition [$t(11)=3.977$, $p=0.002$]. Pairwise comparisons with the incremented-fourth condition indicated that the boundary return, though substantial, was incomplete even for the largest attenuation used (24 dB) [$t(11)=3.771$, $p=0.003$]. These findings indicate that reducing the level of the leading portion of the added 500-Hz tone progressively weakens the effect of the asynchrony, but that a small effect remains even when the leading portion has been attenuated to far below the level of any harmonic in the F1 region of the vowel. Darwin (1995) found a similar relationship, except that the boundary return for a given level of attenuation was about 10% greater in our study than in his. Nonetheless, it should be noted that this difference is small enough for our data to be compatible with Darwin's conclusion that listeners are generally able to subtract perceptually the amplitude of the leading portion of the added 500-Hz tone from the overall amplitude at that frequency during the vowel.

The extent of restoration brought about by the captors used in the earlier experiments averaged about a third. Therefore, the boundary returns observed in the current experiment suggest that accompanying the leading portion of the fourth harmonic with a captor tone of equal level is equivalent to reducing its physical level by about 6 dB. This is a little less than the reduction of 9 dB suggested by Darwin's (1995) data. It seems plausible that this degree of "gain reduction" in the neural response might arise from an inhibitory process in the central auditory system. Note that this does not necessarily imply that the captor decreases the perceived level of the leading portion of the fourth harmonic.

VIII. FINAL DISCUSSION

Darwin and Sutherland (1984) proposed that the partial restoration of the contribution of the asynchronous fourth harmonic to vowel quality depends on the perceptual grouping of the captor tone with the leading portion of that component. They envisaged the common onset time for the captor and the leading portion as the primary factor governing their grouping. Harmonic relations between them may have provided an additional basis for their grouping. However, our results indicate that the restoration effect produced by a pure-tone captor does not depend either on a common onset, or harmonic relations between the captor and the leading component. In the absence of any other obvious factors encouraging their grouping, a different kind of explanation is required. Given the results of experiment 5, we propose that the effect of the captor is mediated by a reduction in the *effective* level of the leading portion of the fourth harmonic that arises from an inhibitory interaction between the two components. Before this proposal is discussed in more detail, it is important to consider the implications of using single- (rather than multi-) component captors, in relation to assessing the possible contribution to the restoration effect of grouping by harmonicity.

A. Captor efficacy and harmonic relations

Darwin and Ciocca (1992) showed that a slightly mistuned component made a smaller contribution to the global pitch of an otherwise harmonic complex tone when it began 160 ms or more before the other components. Ciocca and Darwin (1993) found that this effect of onset asynchrony could be reduced by accompanying the leading portion of the mistuned harmonic with a simultaneous multi-component captor. The target complex had an F0 of 155 Hz and the leading fourth component was set to either 640 or 600 Hz, corresponding to a mistuning of about +3% or −3%, respectively. The captor comprised harmonics 1–2 and 4–5 of either an F0 of 200 Hz or of 213.3 Hz. These values were chosen so that the leading component either corresponded to the third harmonic of the captor F0 or was mistuned from it by about 6%. Ciocca and Darwin (1993) found that the captor was effective at restoring the contribution of the leading

component to global pitch only when it was harmonically related to that component and when the onset asynchrony on the leading component was 160 ms or more. This result suggests that grouping by harmonicity primarily determines captor efficacy in this context, such that an inharmonic relationship between the leading component and the captor prevents grouping by common onset.

Bregman (1990) has argued that a minimum of two components is needed to define a harmonic frame with which another component may fuse perceptually on the basis of sharing a common F0. This requirement was not met by Darwin and Sutherland (1984), or in our experiments, because pure-tone captors were used. However, there are potential difficulties in using captors comprising more than one component when judgments of vowel quality are required. To avoid significant adaptation in the region of F1 and F2, the captor components would need to be restricted to fall in the spectral trough between these two formants. To ensure that the captor components and the leading fourth component (500 Hz) also form a set of consecutive harmonics, the only option would be to use a two-component captor with frequencies of 1000 and 1500 Hz. The efficacy of this captor could be compared with that of others whose partials did not form a harmonic series with the leading fourth component. However, it should be noted that pitch and timbre are distinct perceptual properties of a complex tone (cf. Hukin and Darwin, 1995), and so the factors that govern the ability of a captor to reduce the effects of onset asynchrony on judgments of these different attributes are not necessarily the same. Whatever the outcome of the proposed experiment, we can be confident that the significant restoration effect produced by pure-tone captors, in the context of judgments of vowel quality, cannot be explained in terms of harmonic relations.

B. How might the restoration effect arise from inhibitory interactions in the central auditory system?

Our proposal that the restoration effect arises from inhibitory interactions between the captor and the leading portion of the fourth harmonic poses the question of how and where this happens in the auditory system. Inhibitory interactions between cells occur at many levels in the central auditory system, and so the psychophysical restoration effect may arise at one of several (or even multiple) locations along the auditory pathway. Given this, our approach has been to propose a neural processing scheme that is simple, based on known physiology, and located as early in the auditory pathway as is plausible. On this basis, a cochlear-nucleus circuit is presented that has the right kind of response properties to account for the restoration effect, but which also serves a more general illustrative purpose. Even if the restoration effect is eventually shown to arise from inhibitory interactions elsewhere in the auditory system, these interactions are very likely to produce responses with the same general form as those described here.

The key to the circuit is broadband inhibition in the ventral cochlear nucleus (Palombi and Caspary, 1992) applied by onset chopper (OC) cells to transient chopper (CT) cells, as shown schematically in Fig. 7. Each CT cell receives excitation from auditory nerve fibers arising from the same place along the cochlea, and so has a narrow-band receptive field. CT cells are known to receive inhibitory inputs from OC cells (Ferragamo *et al.*, 1998), and these connections are assumed to occur between cells sharing a common "best frequency" or BF (e.g., Pressnitzer *et al.*, 2001). Each OC cell receives excitation from auditory nerve fibers arising from different places along the cochlea, and so has a wideband receptive field. The receptive field of an OC cell extends on average from about two octaves below the cell's BF to about one octave above (Palmer *et al.*, 1996). Note that the effect of frequency separation on captor efficacy that was observed in experiment 3 is broadly consistent with the expected upper cutoff frequency for OC cells with a BF equal to the frequency of the leading fourth harmonic (500 Hz). The output of the circuit is transmitted via the excitatory projections from the CT cells to the dorsal cochlear nucleus and inferior colliculus (e.g., Adams, 1983).

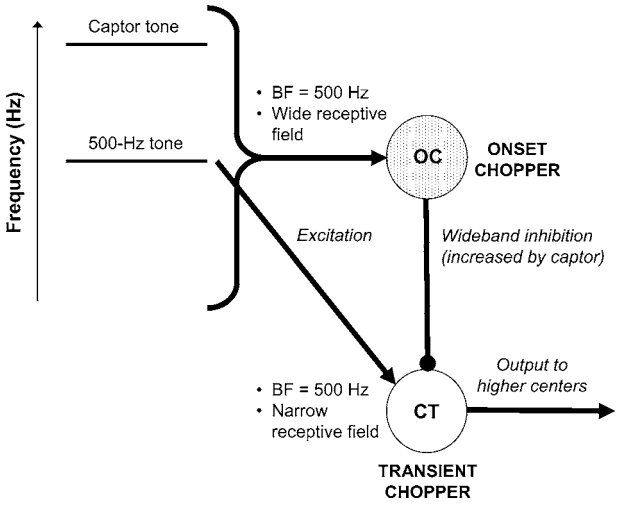


FIG. 7. Schematic of a hypothetical neural circuit in the cochlear nucleus to account for the restoration effect. BF refers to the best frequency of a cell. It is proposed that the captor tone reduces the excitatory response of transient choppers (CTs) tuned to the leading portion of the 500-Hz tone by increasing the wideband inhibitory input to these cells from onset choppers (OCs) with the same BF. This is equivalent to reducing the level of the leading portion, which has been shown to reduce the effect of asynchrony on the contribution of the 500-Hz tone to perceived vowel quality (experiment 5). See text for a full discussion.

Consider the response of CT and OC cells with a BF of 500 Hz to the leading portion of the fourth harmonic in our stimuli. Both types of cell will receive excitation from auditory nerve fibers tuned to 500 Hz, but the response of the CT cells will depend on the balance of this excitation and the weaker inhibition received via the OC cells. Now consider how the output of these CT cells will change when the leading portion of the 500-Hz component is accompanied by a 1-kHz captor tone. The captor tone falls outside the narrow receptive field of the CT cells, but falls within the wide receptive field of the OC cells. Hence, the direct excitatory input to these CT cells will not change, but the indirect inhibitory input will increase, leading to a reduced response to the 500-Hz component. This net inhibition may be regarded as reducing the *effective* level of the leading portion of the

fourth harmonic, such that the presence of the captor has a similar effect to that of reducing its *physical* level, as observed in experiment 5. However, there is another means by which the increased inhibition evoked by the captor tone might lead to a restoration effect. Both we and Darwin and Sutherland (1984) have used captor tones that ended at the same moment as the components of the vowel began. The release from inhibition at captor offset may have produced a transient rebound in the excitatory response of CT cells tuned to 500 Hz, and hence an increased response to the fourth harmonic at vowel onset.

Bleeck *et al.* (2005) have recently observed responses in the ventral cochlear nucleus of the guinea pig that are consistent with the two aspects of the proposed mechanism. They used a two-component complex tone in which the higher component ended first. The lower component was chosen to match the BF of the unit being recorded and the higher component was set to be spectrally remote. This arrangement is analogous to that of our fourth harmonic (which continues through the vowel) and captor tone (which ends at vowel onset). A number of units, some of which were classified as CT cells, showed both a reduced response to the lower tone when the upper tone was present (sustained inhibition) and a rebound burst when the upper tone terminated (release from inhibition). Of course, the response to stimuli like those used in our psychophysical studies is complicated by the onset of all the other components of the vowel when the captor ends, many of which may act to inhibit the response to the continuing 500-Hz component. However, two points are worth noting. First, the fourth harmonic is the most intense component of the vowel, because it is both close to the F1 frequency and incremented by 9.5 dB relative to the spectral envelope. Second, the key comparison is between the response to the fourth harmonic at vowel onset in the presence or absence of the captor. Compared to the no-captor case, the termination of the captor would still have produced a *relatively* greater response to the fourth harmonic at vowel onset, even if the other components of the vowel acted to reduce the absolute magnitude of the response.

In conclusion, our proposed processing scheme provides a good qualitative account of our psychophysical findings. Although there is insufficient evidence to conclude that the restoration effect arises specifically in the cochlear nucleus, it seems very probable that it arises somewhere in the central auditory system from broadband inhibition of a similar form.

C. Implications for asynchrony as a cue for auditory grouping and concluding remarks

Our findings contradict Darwin and Sutherland's (1984) proposal that the restoration effect depends on the perceptual integration of the captor and the leading portion of the four harmonic. However, this outcome does *not* undermine the idea that the reduced contribution to vowel quality made by a component that begins before the others depends on asynchrony acting as a perceptual grouping cue for its exclusion. It was noted earlier that an offset asynchrony, produced by extending a component beyond the end of a vowel, can reduce its contribution to vowel quality (Darwin and Sutherland, 1984; Darwin, 1984b; Roberts and Moore, 1991). Peripheral adaptation cannot account for this perceptual effect. What remains for conjecture is the functional significance of the restoration effect and the reason for its insensitivity to the established grouping cues of common onset-time and harmonic relations.

We suggest that the restoration effect produced by a captor arises as a by-product of a mechanism that has evolved to enhance the representation of spectral features in broadband stimuli. The simple neural circuit that we have proposed to account for our data provides the basis for a network of lateral inhibition in which the more intense components will be more effective than the less intense ones at reducing the excitation evoked by the other components. For our stimuli, only two elements preceded the vowel (the captor and the leading portion of the fourth harmonic) and these were equally intense, most probably leading to a comparable degree of mutual inhibition. Hence, the idea that accompanying the leading portion of the fourth harmonic with a captor tone reduces the effective level of the leading portion seems a reasonable one. Furthermore, it seems plausible that a mechanism designed to enhance spectral features in a stimulus is fundamental enough to operate *before* grouping cues are used to decide whether the stimulus represents one or more acoustic events in the environment.

Further research is required to assess how well our proposed neural circuit in the cochlear nucleus can explain why onset asynchrony is less effective at excluding an added 500-Hz tone from perceived vowel quality when it is accompanied by a captor. However, it can be concluded from our psychophysical results that the *modulating* influence of a pure-tone captor on the consequences of onset asynchrony for timbre perception does not depend on grouping cues. Whatever the details of the underlying neural processes eventually prove to be, it seems very likely that the restoration effect arises from some kind of inhibitory interaction between the captor and the leading portion of the asynchronous component of the vowel. More generally, researchers should be aware of the potential for the presence of broadband inhibitory effects to produce patterns of behavior confusable with those produced by more cognitive mechanisms of auditory grouping.

ACKNOWLEDGMENTS

This research was supported by Research Grant No. 6/S17805 from the Biotechnology and Biological Sciences Research Council (UK) to the first author. We are grateful to Ian Winter, Stefan Bleeck, Jesko Verhey, and Neil Ingham for their collaborative research on the neural mechanisms in auditory brainstem that may mediate the perceptual effects reported here. Our thanks go to Al Bregman for stimulating discussion about the possible functional significance of the restoration effect produced by the captor tones. We are also grateful to the reviewers, Chris Darwin and Christophe Micheyl, for their helpful comments on an earlier version of this manuscript.

Adams, J. C. (**1983**). "Multipolar cells in the ventral cochlear nucleus project to the dorsal cochlear nucleus and the inferior colliculus," Neurosci. Lett. **37**, 205–208.

Assmann, P. F., and Nearey, T. M. (**1987**). "Perception of front vowels: The role of harmonics in the first formant region," J. Acoust. Soc. Am. **81**, 520–534.
Bleeck, S., Ingham, N., Verhey, J., and Winter, I. M. (**2005**). "Wideband suppression in cochlear nucleus: A role in grouping by common onset?" in ARO Abstracts **28**, 236.
Brady, S. A., and Darwin, C. J. (**1978**). "Range effect in the perception of voicing," J. Acoust. Soc. Am. **63**, 1556–1558.
Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).
Bregman, A. S., and Doehring, P. (**1984**). "Fusion of simultaneous tonal glides: The role of parallelness and simple frequency relations," Percept. Psychophys. **36**, 251–256.
Bregman, A. S., and Pinker, S. (**1978**). "Auditory streaming and the building of timbre," Can. J. Psychol. **32**, 19–31.
Carlson, R., Fant, G., and Granstrom, B. (**1975**). "Two-formant models, pitch and vowel perception," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London), pp. 55–82.
Chistovich, L. A. (**1985**). "Central auditory processing of peripheral vowel spectra," J. Acoust. Soc. Am. **77**, 789–805.
Chistovich, L. A., Sheikin, R. L., and Lublinskaja, V. V. (**1979**). "Centres of gravity and spectral peaks as the determinants of vowel quality," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Ohman (Academic, London), pp. 143–157.
Ciocca, V., and Darwin, C. J. (**1993**). "Effects of onset asynchrony on pitch perception: Adaptation or grouping?" J. Acoust. Soc. Am. **93**, 2870–2878.
Dannenbring, G. L., and Bregman, A. S. (**1978**). "Streaming vs. fusion of sinusoidal components of complex tones," Percept. Psychophys. **24**, 369–376.
Darwin, C. J. (**1981**). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Q. J. Exp. Psychol. A **33**, 185–207.
Darwin, C. J. (**1984a**). "Auditory processing and speech perception," in *Attention and Performance X: Control of Language Processes*, edited by H. Bouma and D. G. Bouwhuis (Erlbaum, Hillsdale, NJ), pp. 197–210.
Darwin, C. J. (**1984b**). "Perceiving vowels in the presence of another sound: Constraints on formant perception," J. Acoust. Soc. Am. **76**, 1636–1647.
Darwin, C. J. (**1995**). "Perceiving vowels in the presence of another sound: A quantitative test of the 'old-plus-new' heuristic," in *Levels in Speech Communication: Relations and Interactions*, edited by C. Sorin, J. Mariani, H. Meloni, and J. Schoentgen (Elsevier, Amsterdam), pp. 1–12.
Darwin, C. J. (**1997**). "Auditory grouping," Trends in Cognitive Science **1**, 327–333.
Darwin, C. J., and Carlyon, R. P. (**1995**). "Auditory grouping," in *Hearing: Handbook of Perception and Cognition (2nd ed.)*, edited by B. C. J. Moore (Academic, London), pp. 387–424.
Darwin, C. J., and Ciocca, V. (**1992**). "Grouping in pitch perception: Effects of onset asynchrony and ear of presentation of a mistuned component," J. Acoust. Soc. Am. **91**, 3381–3390.
Darwin, C. J., and Sutherland, N. S. (**1984**). "Grouping frequency components of vowels: When is a harmonic not a harmonic?" Q. J. Exp. Psychol. A **36**, 193–208.
Elliott, L. L. (**1971**). "Backward and forward masking," Audiology **10**, 65–76.
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.
Finney, D. J. (**1971**). *Probit Analysis*, 3rd ed. (Cambridge U.P., Cambridge, UK).
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Green, D. M., and Dai, H. (**1992**). "Temporal relations in profile comparisons," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 471–478.
Henke, W. L. (**1997**). *MITSYN: A coherent family of high-level languages for time signal processing*, software package (Belmont, MA).
Henton, C. G. (**1983**). "Changes in the vowels of received pronunciation," J. Phonetics **11**, 353–371.
Hill, N. I., and Darwin, C. J. (**1996**). "Lateralization of a perturbed harmonic: Effects of onset asynchrony and mistuning," J. Acoust. Soc. Am. **100**, 2352–2364.
Houtgast, T. (**1974**). "Lateral suppression in hearing," doctoral thesis, Free University, Amsterdam.
Hukin, R. W., and Darwin, C. J. (**1995**). "Comparison of the effect of onset asynchrony on auditory grouping in pitch matching and vowel identification," Percept. Psychophys. **57**, 191–196.
Karnickaya, E. G., Mushnikov, V. N., Slepokurova, N. A., and S., Ja. Zhukov (**1975**). "Auditory processing of steady-state vowels," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London), pp. 37–53.
Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT, Cambridge, MA).
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
McAdams, S., and Drake, C. (**2002**). "Auditory perception and cognition," in *Stevens' Handbook of Experimental Psychology. Vol. I: Sensation and Perception (3rd ed.)*, edited by H. Pashler and S. Yantis (Wiley, New York), pp. 397–452.
Miyazaki, K., and Sasaki, T. (**1984**). "Pure-tone masking patterns in nonsimultaneous masking conditions," Jpn. Psychol. Res. **26**, 110–119.
Oxenham, A. J., and Moore, B. C. J. (**1994**). "Modeling the additivity of nonsimultaneous masking," Hear. Res. **80**, 105–118.
Palmer, A. R. (**1995**). "Neural signal processing," in *Hearing: Handbook of Perception and Cognition (2nd ed.)*, edited by B. C. J. Moore (Academic, London), pp. 75–121.
Palmer, A. R., Jiang, D., and Marshall, D. H. (**1996**). "Response of ventral cochlear nucleus onset and chopper units as a function of signal bandwidth," J. Neurophysiol. **75**, 780–794.
Palombi, P. S., and Caspary, D. M. (**1992**). "GABA-A receptor antagonist bicuculline alters response properties of posteroventral cochlear nucleus neurons," J. Neurophysiol. **67**, 738–746.
Pressnitzer, D., Meddis, R., Delahaye, R., and Winter, I. M. (**2001**). "Physiological correlates of comodulation masking release in the mammalian ventral cochlear nucleus," J. Neurosci. **21**, 6377–6386.
Rasch, R. A. (**1978**). "The perception of simultaneous notes such as in polyphonic music," Acustica **40**, 21–33.
Roberts, B., and Moore, B. C. J. (**1990**). "The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels," J. Acoust. Soc. Am. **88**, 2571–2583.
Roberts, B., and Moore, B. C. J. (**1991**). "The influence of extraneous sounds on the perceptual estimation of first-formant frequency in vowels under conditions of asynchrony," J. Acoust. Soc. Am. **89**, 2922–2932.
Snedecor, G. W., and Cochran, W. G. (**1967**). *Statistical Methods, 6th ed.* (Iowa U.P., Ames, IA).
Westerman, L. A., and Smith, R. A. (**1984**). "Rapid and short-term adaptation in auditory nerve responses," Hear. Res. **15**, 249–260.
Zwislocki, J. J. (**1978**). "Masking: Experimental and theoretical aspects of simultaneous, forward, backward, and central masking," in *Handbook of Perception. Vol. IV. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic, New York), pp. 283–336.

Spectral loudness summation for short and long signals as a function of level

Anne-Kristin Anweiler and Jesko L. Verhey[a)]
AG Neusorensorik, Institut für Physik, Fakultät 5, Carl von Ossietzky Universität, 26111 Oldenburg, Germany

(Received 15 June 2005; revised 8 February 2006; accepted 9 February 2006)

The equal-loudness-ratio hypothesis (ELRH) assumes that the loudness ratio between equal-level long and short signals with the same spectrum is independent of level and spectrum. Thus, it predicts the same amount of spectral loudness summation for equal-level short and long signals. To investigate how duration and level affect spectral loudness summation, loudness functions for 10- and 1000-ms bandpass-filtered noise centered at 2 kHz were obtained for normal-hearing listeners using categorical loudness scaling. The bandwidth was varied from 200 to 6400 Hz. From the loudness functions the level difference between the 3200-Hz-wide reference signal and the equally loud test signals was determined for reference levels in the range from 25 to 105 dB SPL. In addition, for two reference levels (45 and 65 dB SPL), this level difference was measured using a loudness-matching procedure. Both procedures showed similar results for the two reference levels. Over a wide level range spectral loudness summation is larger for the short than for the equal-level long signals. This is in contrast to the predictions of the ELRH. However, the data at moderate levels are consistent with the expectations of a modified ELRH which assumes a smaller loudness ratio for broadband than for narrow-band signals.
[DOI: 10.1121/1.2184224]

PACS number(s): 43.66.Cb, 43.66.Mk, 43.66.Ba [JHG] Pages: 2919–2928

I. INTRODUCTION

Many studies have shown that loudness depends on the spectrum of a sound. The loudness of a sound with constant intensity significantly increases with bandwidth when a critical bandwidth is exceeded (e.g., Fletcher and Munson, 1933; Zwicker and Feldtkeller, 1955; Zwicker *et al.*, 1957; Scharf, 1959; Cacace and Margolis, 1985; Schneider, 1988). Thus, the level of a narrow-band signal is greater than that of a broadband signal at the same loudness. This effect is commonly referred to as spectral loudness summation. It is believed to be a consequence of two properties of the auditory system: spectral filtering and compression (e.g., Zwicker and Fastl, 1999). In each frequency band, loudness is determined separately. Then total loudness is determined by summing partial loudnesses across frequency bands. Due to the compressive relation between intensity of the signal components falling within the passband of each frequency band and partial loudness, the sum of partial loudnesses for a broadband signal stretching across several frequency bands is larger than the loudness of an equal-SPL narrow-band signal processed within one frequency band.

Loudness depends not only on the spectral content but also on the duration of a sound. Many studies examining the effect of signal duration on loudness have shown that loudness increases with signal duration (Port, 1963; Ekman *et al.*, 1966; Zwicker, 1966; Poulsen, 1981; Takeshima *et al.*, 1988). This effect is commonly referred to as temporal integration of loudness. Temporal integration is usually quantified as the level difference between equally loud short and long signals with the same spectrum (e.g., Poulsen, 1981; Florentine *et al.*, 1996; Buus *et al.*, 1997). Two recent studies investigating the level dependence of temporal integration of loudness (Florentine *et al.*, 1996; Buus *et al.*, 1997) found that the temporal integration of loudness varied nonmonotonically with level. They concluded that the loudness function must have a different slope for short and long stimuli at the same loudness. Based on these findings, Florentine *et al.* (1996) proposed the equal-loudness-ratio hypothesis (ELRH), which assumes that the loudness ratio between equal-SPL long and short signals is independent of level. This hypothesis is in good agreement with the data of Epstein and Florentine (2005) who measured loudness functions for short (5 ms) and long (200 ms) 1-kHz tones using cross-modality matching. They obtained an approximately constant loudness ratio except for a slight increase at low levels (below 40 dB SPL). In addition, Buus *et al.* (1997) assumed that this ratio is also independent of the spectral content of the signal, i.e., the ratio between the loudnesses of two stimuli with different durations, but the same spectrum, is independent of stimulus level and spectrum. Therefore, the loudness versus intensity curves of short and long sounds are vertical transposes of one another, provided they have the same spectrum and loudness is expressed in sones on a logarithmic scale. The ELRH predicts the same magnitudes of spectral loudness summation for different signal durations when compared at the same reference level (Verhey and Kollmeier, 2002) (see also Fig. 2, right panel). However, recent loudness matching data with a broadband reference at medium levels (Verhey and Kollmeier, 2002) indicate that the magnitude of spectral loudness summation (i.e., the level difference between equally loud narrow- and broadband sig-

[a)]Electronic mail: jesko.verhey@uni-oldenburg.de

nals) at the same reference level is larger for short signals (10 ms) than for long signals (1000 ms). Verhey and Kollmeier (2002) argued that a modified version of the ELRH may account for the data if it is assumed that the loudness ratio is smaller for broadband signals than for narrow-band signals.

The effect of duration on the magnitude of spectral loudness summation has hitherto only been investigated over a limited level range and only using a loudness matching procedure (Verhey and Kollmeier, 2002; Fruhmann *et al.*, 2003). The present study investigates whether a similar effect of duration on the magnitude of spectral loudness summation can be found using a different procedure and whether it is only limited to the intermediate level range used in the previous studies. Another aim of the study is to investigate if a modified ELRH with a spectrum-dependent loudness ratio can predict the loudness data. The present study uses the adaptive categorical loudness scaling procedure to investigate spectral loudness summation as a function of duration over a wide level range. In addition, for two intermediate levels, data are obtained using a loudness matching procedure, so that results can be compared across methods.

II. METHOD

A. Procedure

1. Loudness scaling

Loudness functions for short and long bandpass noise signals with different bandwidths were measured using the adaptive categorical loudness scaling procedure (Brand and Hohmann, 2002). Presentation levels were chosen adaptively across the entire auditory dynamic range of the subject and presented in random order. Subjects were asked to rate the loudness of each signal on a response scale consisting of 11 response alternatives. Named loudness categories were "sehr leise" ("very soft"), "leise" ("soft"), "mittel" ("medium"), "laut" ("loud"), and "sehr laut" ("very loud"). Four unnamed intermediate response alternatives were represented by horizontal bars. In addition, the two extreme categories "unhörbar" ("inaudible") and "zu laut" ("too loud") could be selected.

Preliminary measurements showed that for most subjects judging the loudness of the short signals was very difficult when tracks of different bandwidth were interleaved. In order to reduce scattering, bandwidths were therefore presented sequentially. In addition, the shorter signals were presented twice with a silent interstimulus interval of 700-ms duration during each trial.

All subjects performed three repetitions of loudness scaling for each signal duration. The sequence of presentation of durations was randomized between subjects.

Loudness scaling can be biased by several factors including the inaccurate estimation of the limits of the auditory dynamic range and an uneven distribution of the levels within those limits. Brand and Hohmann (2002) showed that the adaptive procedure better covers the dynamic range than a nonadaptive loudness scaling procedure that roughly determined the auditory dynamic range in a premeasurement phase. The accuracy of the categorical scaling procedure is limited by the number of response alternatives. Brand and Hohmann (2002) reported that, because of the restriction to 11 response alternatives, the subjects were forced to respond less accurately than they could, especially at soft levels. Thus, in the present study only levels above 25 dB were further analyzed. At higher levels this effect is reduced. This is probably due to the different slopes of the loudness functions at high and low levels. Brand and Hohmann (2001, 2002) reported that loudness functions tend to be steeper at high than at low levels.

2. Loudness matching

Test stimuli with different bandwidths were matched in loudness to a reference signal with fixed bandwidth and level using a two-interval, two-alternative forced choice procedure with interleaved tracks (Verhey and Kollmeier, 2002). In each trial, test and reference signal were presented in random order separated by a silent interval of 500-ms duration. Subjects indicated which signal was louder by pressing the corresponding key on a keyboard. The level of the test signal was varied according to a one-up one-down procedure. Starting level was either +10 or −10 dB relative to the level of the reference signal. If the subject indicated that the test signal was louder, its level was decreased in the next trial, otherwise it was increased. Step size started at 8 dB. At each upper reversal of the level tracking procedure it was decreased by a factor of 2 until a minimum step size of 2 dB was reached. At this step size the track continued for four reversals and then stopped. The level difference between test and reference signal needed to obtain equal loudness was determined as the mean value of the levels at these last four reversals.

Matching was performed for the two signal durations separately, i.e., signals of different duration were not compared directly. Each subject completed four runs of the experiment for each combination of duration and reference level. The sequence of presentation of reference level and duration was randomized between subjects. Tracks of the five different bandwidths of the same duration were interleaved. The five starting levels of the first and third experimental run were selected randomly from the values 10 and −10 dB relative to the reference level. For the second and fourth runs the respective opposite starting levels were used. This choice of starting levels differs from Buus *et al.* (1997) and Verhey and Kollmeier (2002). However, Verhey (1999) showed that the starting level has a negligible effect on spectral loudness summation for bandpass-noise signals when using an adaptive procedure with interleaved tracks.

B. Data analysis

For analysis of loudness scaling, loudness categories were linearly transformed to numerical values (categorical units, cu) from 0 (inaudible) to 50 (too loud). A model loudness function similar to the one used in Brand and Hohmann (2002) was fitted to all responses of one track. The function consists of two linear parts with independent slope values. The transition region between these linear parts is smoothed using a Bezier fit. The exact equations are given in Appendix

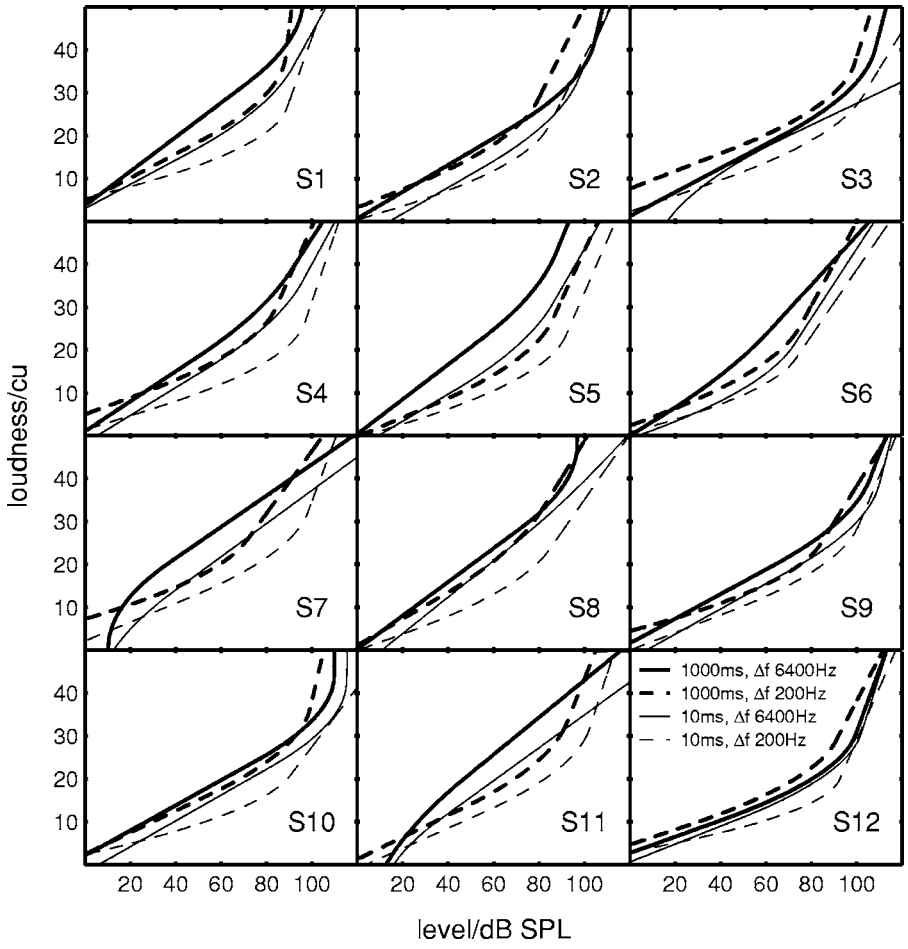


FIG. 1. Individual loudness functions in categorical units as a function of signal level in dB SPL for 12 normal-hearing listeners. Thick lines represent long signals (1000 ms) and thin lines short signals (10 ms). Data are shown for the widest (6400 Hz, solid lines) and narrowest (200 Hz, dashed lines) bandwidth used in the present study. Functions were averaged across three repetitions of the experiment.

A. In contrast to the function used by Brand and Hohmann (2002), the loudness of the intersection point of the linear parts is not fixed at 25 cu but is an additional parameter. A fixed loudness of the intersection point would only allow a horizontal shift along the axis of the physical level between the loudness functions of the short and long signals with the same spectral content. However, Florentine *et al.* (1996) argue that the loudness function for the short signal is a vertically shifted version (towards lower loudness values) of that for the long signal. Therefore the loudness of the intersection point was introduced as a parameter. The loudness function was fitted to the data for each run individually by minimizing the standard deviation. From the loudness functions, the level differences needed to produce the sensation of equal loudness were calculated individually using the inverse loudness function (see Appendix A).

For loudness scaling and matching the mean level difference of all repetitions of the experiment was calculated for each combination of bandwidth, duration, and reference level. To examine the statistical significance of the effects of the stimulus variables a four-way ANOVA for repeated measures was performed (SPSS Release 12.0). The dependent variable was the level difference between test and reference signals at equal loudness. Fixed factors were bandwidth, level, and duration. For loudness scaling five reference levels were chosen: 25, 45, 65, 85, and 105 dB SPL. Subject was a random factor. To investigate sources of significance *posthoc* Scheffé tests were performed.

Levels below 25 dB SPL were not analyzed because of the limited number of response alternatives at soft levels (see above). In addition, Blum *et al.* (2002) reported that the fitted loudness functions tend to underestimate quiet threshold.

C. Stimuli and apparatus

Stimuli for both procedures were bandpass noise signals centered geometrically around 2 kHz. In order to reduce envelope fluctuations, single-iterated low-noise noise was used, i.e., signals were divided by their Hilbert envelope and then restricted to their original bandwidth (Kohlrausch *et al.*, 1997; Verhey and Kollmeier, 2002). Signal durations were 10 and 1000 ms including 5-ms cosine-square ramps at signal on- and offset.

For the loudness scaling experiments signal bandwidths were 200, 400, 800, 1600, 3200, and 6400 Hz. For loudness matching the bandwidth of the reference signal was 3200 Hz. Test signal bandwidths were the remaining band-

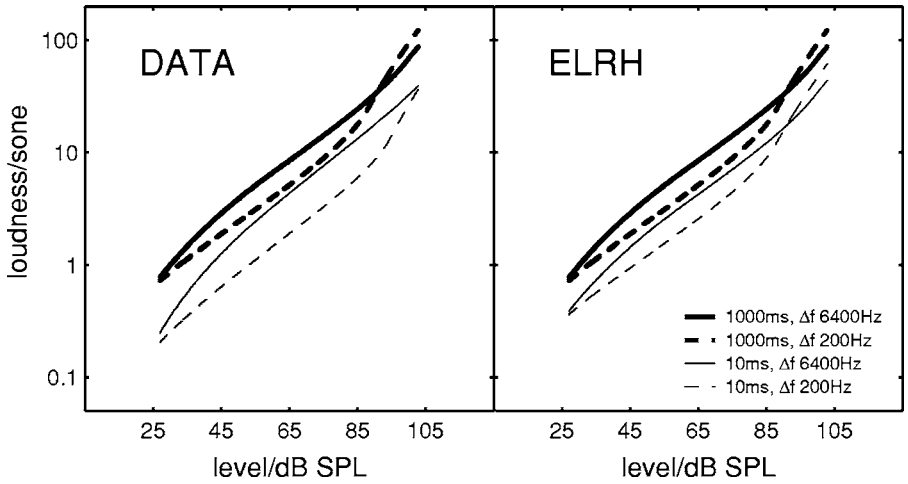


FIG. 2. Loudness functions in sones plotted as a function of signal level in dB SPL for two different bandwidths and two durations. Solid lines represent broadband signals (6400 Hz), dashed lines narrow-band signals (200 Hz). The left panel shows average loudness functions across 12 normal-hearing listeners calculated from loudness scaling data for long signals (1000 ms, thick lines) and short signals (10 ms, thin lines). The right panel shows the same functions as the left panel for long signals, whereas for short signals loudness functions are shifted by a constant factor of 2 relative to those for long signals independent of level and bandwidth as assumed by the equal-loudness-ratio hypothesis (ELRH). In contrast to the predicted loudness functions for the short tones (right panel), the data (left panel) show that the loudness ratio between short and long signals depends on the spectrum and may not be constant for very low and high levels.

widths used in the scaling experiment. Test and reference signal always had the same duration. Reference levels were 45 and 65 dB SPL.

Stimuli were generated digitally at a sampling rate of 44.1 kHz. A standard personal computer controlled stimulus generation and presentation and recorded results using a software package developed at the University of Oldenburg. Stimuli were D/A converted (RME TDIF-1), amplified (Tucker-Davis HB7), and presented monaurally via a Sennheiser HDA 200 headphone with free-field equalization. Subjects were seated in a sound-insulated booth. Responses were given by touching the corresponding button on a touch screen.

D. Subjects

Twelve normal-hearing subjects (ten male, two female; aged 21–29 years) participated in both the loudness scaling and loudness matching experiments. Two subjects were members of the research group, the others were paid volunteers. None of them had previous experience in loudness judgment experiments. Subjects were given no training on the task, but four subjects (S1, S7, S8, and S11) participated in preliminary loudness scaling experiments of half an hour each. All subjects had normal audiograms with hearing thresholds lower than or equal to 15 dB HL at the standard audiometric frequencies from 125 Hz to 8 kHz.

III. RESULTS

A. Loudness scaling

Figure 1 shows individual loudness functions averaged across three repetitions of the experiment for the narrowest and widest bandwidth used in the present study (200 Hz represented by dashed lines and 6400 Hz represented by solid lines). Long and short signals (1000 and 10 ms) are indicated by thick and thin lines, respectively. In general, loudness functions for all subjects show a similar behavior. Long signals are perceived as louder than short signals at the same level and, except for subjects S3 and S12, loudness functions for narrow-band signals curved more than those for broadband signals, i.e., for intermediate levels broadband signals are perceived as louder than narrow-band signals at the same level. However, there are large interindividual differences in the amount of spectral loudness summation, i.e., the horizontal distance between the solid and dashed lines with the same line width. It ranges from a few dB (e.g., subjects S8 and S10) up to 30 dB (subject S5) for the long signals. In general, the level difference between equally loud narrow- and broadband signals tends to be smaller for the long signals than for the short signals at the same level of the broadband signal.

The left panel of Fig. 2 shows average loudness functions across 12 normal-hearing listeners calculated from the individual data shown in Fig. 1. In contrast to Fig. 1, loudness is expressed in sones. To determine the average loudness functions, for each category from "very soft" to "very loud" the corresponding level was calculated for each subject from the individual loudness functions. Then the model loudness function was fitted to the mean of these levels and the loudness in categorical units was transformed to loudness in sones using the equation in Appendix B.

The loudness functions are shallower at moderate levels than at low and high levels. The vertical distance between loudness functions of the same bandwidth but different duration is approximately constant which supports the assumption of a level-independent loudness ratio. However, in qualitative agreement with Epstein and Florentine (2005) the ratio increases for levels below 40 dB SPL. In contrast to the assumptions of the equal-loudness-ratio hypothesis as proposed in Buus *et al.* (1997), the vertical distance between loudness functions of long and short duration is not independent of the spectrum of the sound. For the broadband signals,

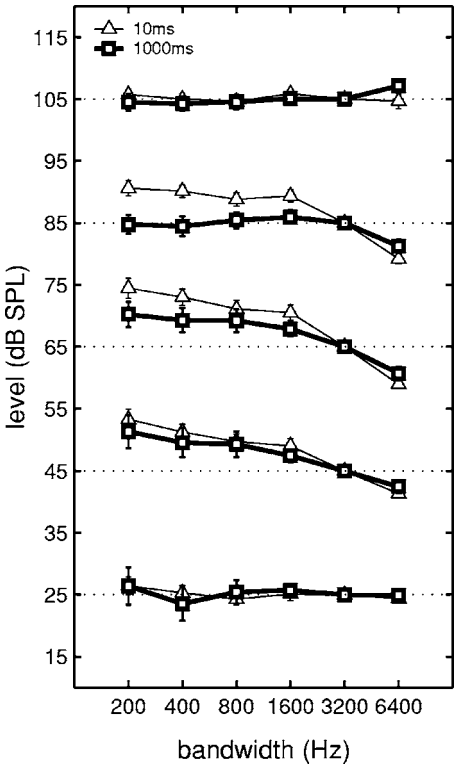


FIG. 3. Average spectral loudness summation as a function of signal bandwidth obtained from loudness scaling data of bandpass noise signals for 12 normal-hearing subjects. The level difference between the signal and the 3200-Hz-wide reference at equal loudness is plotted as a function of the signal bandwidth. Lines connect points of equal loudness for different reference levels (dotted horizontal lines): 25, 45, 65, 85, and 105 dB SPL. Signal durations were 1000 ms (squares, thick lines) and 10 ms (triangles, thin lines). Vertical bars show plus and minus one standard error across 12 listeners.

the vertical distance is smaller than for the narrow-band signals. This indicates that for signals with different spectra different level-independent loudness ratios between short and long signals exist.

To allow for a direct comparison with the predictions of the equal-loudness-ratio hypothesis, the right panel shows the same loudness scaling data as the left panel for long signals, whereas for short signals loudness functions are shifted by a constant factor of 2 relative to those for long signals independent of level and bandwidth. A factor of 2 was found by Zwicker and Fastl (1999) as the loudness reduction in sones between 2-kHz sinusoids with durations of 10 and 1000 ms at a SPL of 57 dB. In the present study, the average ratios between long and short signals of equal SPL are 3.0 ± 0.4 for 200-Hz-wide signals and 2.0 ± 0.2 for 6400-Hz-wide signals in the range from 45 to 105 dB SPL.

Figure 3 shows equal loudness contours and standard errors derived from the scaling data for 12 normal-hearing subjects. Lines connect signal levels for different bandwidths that were perceived as equally loud as the 3200-Hz-wide reference. Reference levels were 25, 45, 65, 85, and 105 dB SPL. Thick and thin lines indicate data for 1000 ms (squares) and 10 ms (triangles), respectively.

On average, a level dependence in the amount of spectral loudness summation can be observed. For the long signals, spectral loudness summation is largest for medium levels with a maximum level difference of 7 dB between equally loud 200- and 3200-Hz-wide stimuli at a level of the 3200-Hz-wide stimulus of 45 dB SPL. For the short signals, spectral loudness summation is also largest for medium levels but the maximum is now about 10 dB at a reference level of 65 dB. For very low and very high reference levels (25 and 105 dB SPL, respectively), no spectral loudness summation can be found for the long signal (<2 dB).

For medium to high levels (45 and 85 dB SPL), the level difference between the equally loud 200- and 3200-Hz-wide stimuli is up to 6 dB larger for the 10-ms than for the 1000-ms signals. At the lowest reference level, no duration dependence can be found. The effect of duration increases with increasing level up to a maximum difference in the amount of spectral loudness summation between short and long signals at 85 dB SPL and then decreases towards the highest level.

These effects are supported by a four-way ANOVA for repeated measures (see Table I). The stimulus variables bandwidth and reference level have significant effects on the level difference between the test and the reference signal ($p<0.001$). A Scheffé *posthoc* test for bandwidth shows no significant effect between 200 and 400 and between 400, 800, and 1600 Hz ($p>0.05$). For all other bandwidth pairs the level difference is significantly larger for the signal with the smaller bandwidth, i.e., there is a significant effect of spectral loudness summation. The significant interactions between bandwidth and subject and between duration and subject ($p<0.01$) indicate that the effects of the stimulus parameters on loudness depend on the subject. The significant interaction of bandwidth and duration ($p<0.05$) shows that the amount of spectral loudness summation depends on the duration. The significant interaction of bandwidth, duration,

TABLE I. Four-way analysis of variance for repeated measures of loudness scaling for normal-hearing subjects. The dependent variable is the level difference at equal loudness between the reference signal with 3200-Hz bandwidth and signals of all other bandwidths measured. Bandwidth (bw; five steps: 200, 400, 800, 1600, and 6400 Hz), signal duration (dur; two steps: 10 and 1000 ms), and level of the reference signal (lev; five steps: 25, 45, 65, 85, and 105 dB SPL) are fixed factors. Subject (sub; 12 steps: S1–S12) is a random factor.

Source	Sum of squares	df	Mean square	F	Probability
Const	4142.0	1	4142.0	4.8	0.051
bw	9210.0	5	1842.0	10.8	⩽0.0001
dur	601.1	1	601.1	1.3	0.275
lev	3560.4	4	890.1	14.9	⩽0.0001
sub	9470.1	11	860.9	1.6	0.203
bw*dur	1230.8	5	246.2	3.2	0.014
bw*lev	7403.6	20	370.2	21.1	⩽0.0001
dur*lev	604.8	4	151.2	2.6	0.046
bw*dur*lev	514.6	20	25.7	2.1	0.004
bw*sub	9415.5	55	171.2	2.1	0.003
dur*sub	5015.3	11	455.9	3.7	⩽0.0001
bw*dur*sub	4276.7	55	77.8	6.4	⩽0.0001
lev*sub	2634.5	44	59.9	1.0	0.562
bw*lev*sub	3857.9	220	17.5	1.5	0.003
dur*lev*sub	2522.2	44	57.3	4.8	⩽0.0001
bw*dur*lev*sub	2652.9	220	12.1	0.4	1.000

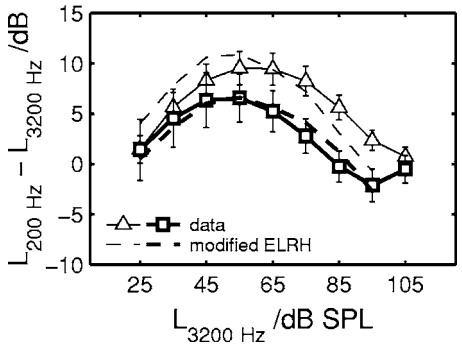


FIG. 4. Average spectral loudness summation as a function of level obtained from loudness scaling of bandpass noise signals for 12 normal-hearing subjects. The level difference between narrow-band (200 Hz) and broadband (3200 Hz) stimuli needed to produce the sensation of equal loudness is plotted as a function of level of the broadband stimulus. Results are shown for durations of 10 ms (triangles, thin lines) and 1000 ms (squares, thick lines). Vertical bars show plus and minus one standard error across 12 listeners. Dashed lines show model predictions assuming a bandwidth-dependent constant loudness ratio between long and short signals (thick and thin lines, respectively).

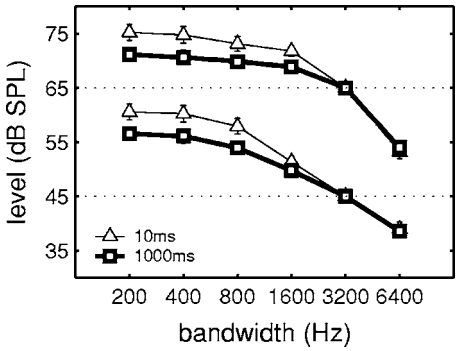


FIG. 5. Average spectral loudness summation as a function of signal bandwidth obtained from loudness matching of bandpass noise signals for 12 normal-hearing subjects. The level difference between the signal and the 3200-Hz-wide reference at equal loudness is plotted as a function of the signal bandwidth. Lines connect points of equal loudness for two different reference levels (dotted horizontal lines): 45 and 65 dB SPL. Signal durations were 1000 ms (squares, thick lines) and 10 ms (triangles, thin lines). Vertical bars show plus and minus one standard error across 12 listeners.

and subject ($p<0.001$) indicates that the effect of duration on the amount of spectral loudness summation depends on the subject. A Scheffé *posthoc* test for level shows no significant effect between 25, 85, and 105 dB SPL and between 45 and 65 dB SPL ($p>0.05$). For all other level pairs the level difference is significantly larger for medium levels than for very high and very low levels. As indicated by the significant interactions between bandwidth and level ($p<0.001$) and bandwidth, level, and subject ($p<0.01$), spectral loudness summation depends on the reference level and this interaction in turn depends on the subject. The interaction of duration, bandwidth, and reference level is significant ($p<0.01$). This indicates that within the investigated level range, the effect of duration on spectral loudness summation depends on the reference level. The significant interactions of duration and level ($p<0.05$) and duration, level, and subject ($p<0.001$) indicate that the duration effects depend on the level and that this interaction in turn depends on the subject.

Figure 4 shows the average level differences and standard errors for 12 normal-hearing listeners between a 200-Hz-wide signal and the 3200-Hz-wide reference as a function of level of the broadband signal. Data for the 10-ms signals are indicated by triangles. Squares denote the data for the 1000-ms signals. In addition to the measured data (solid line) model predictions are shown which will be discussed in Sec. IV. The figure gives a more detailed overview of the data already presented in Fig. 3 for the 200-Hz bandwidth. At low levels, the same amount of spectral loudness summation is found for short and long signals. At medium to high levels (>40 dB SPL), however, spectral loudness summation is larger for short than for long signals with a maximum difference of 7 dB at 80 dB SPL.

B. Loudness matching

Figure 5 shows average loudness matching data and standard errors for 12 normal-hearing subjects. Test signal levels that were perceived equally loud as the reference signal are plotted as a function of the bandwidth of the test signal. Lines connect points of equal loudness. Reference levels were 45 and 65 dB SPL and reference bandwidth was 3200 Hz. Signal durations were 1000 ms (squares, thick lines) and 10 ms (triangles, thin lines).

The average data show spectral loudness summation for both signal durations and both reference levels. The average level differences found for 1000-ms signals quantitatively agree with the data in Verhey and Kollmeier (2002): ΔL between equally loud signals with the smallest bandwidth (200 Hz) and the largest bandwidth (6400 Hz) is 17 dB for a reference level of 65 dB SPL [compared to 15 dB in Verhey and Kollmeier (2002)] and 18 dB for a reference level of 45 dB SPL [compared to 16 dB in Verhey and Kollmeier (2002)]. Spectral loudness summation is slightly larger for the lower reference level. The amount of spectral loudness summation depends on duration. For both reference levels, a 4-dB larger level difference for short than for long signals can be observed between the 200-Hz-wide test signal and the reference.

These effects are supported by a four-way ANOVA for repeated measures (see Table II). All stimulus variables (bandwidth, signal duration, and reference level) have significant effects on the level difference between the test and the reference signal ($p<0.001$ for bandwidth and reference level, $p<0.05$ for signal duration). A Scheffé *posthoc* test shows no significant effect between 200 and 400 Hz ($p>0.05$). For all other bandwidth pairs the level difference is significantly larger for the signal with the smaller bandwidth, i.e., there is a significant effect of spectral loudness summation. The significant interaction of bandwidth and subject ($p<0.05$) indicates that spectral loudness summation depends on the subject. The significant interactions and of bandwidth, duration, and subject ($p<0.01$) show that the amount of spectral loudness summation depends on the duration of the signal and that this interaction in turn depends on the subject. As indicated by the significant interaction between bandwidth and level ($p<0.001$), spectral loudness summation also depends on the reference level. The interaction of duration, bandwidth, and reference level is not significant ($p>0.05$). This indicates that within the investigated

TABLE II. Four-way analysis of variance for repeated measures of loudness matching for normal-hearing subjects. The dependent variable is the level difference between test and reference signal at equal loudness. Bandwidth of the test signal (bw; five steps: 200, 400, 800, 1600, and 6400 Hz), signal duration (dur; two steps: 10 and 1000 ms), and level of the reference signal (lev; two steps: 45 and 65 dB SPL) are fixed factors. Subject (sub; 12 steps: S1–S12) is a random factor.

Source	Sum of squares	df	Mean square	F	Probability
Const	27317.3	1	27317.3	52.2	≤0.0001
lev	4141.7	1	4141.7	61.8	≤0.0001
bw	51354.6	4	12838.7	178.8	≤0.0001
dur	1853.7	1	1853.7	6.0	0.032
sub	5758.6	11	523.5	1.6	0.227
lev*bw	945.8	4	236.4	22.2	≤0.0001
lev*dur	0.7	1	0.7	0.0	0.926
bw*dur	636.5	4	159.1	3.8	0.010
lev*bw*dur	43.7	4	10.9	1.1	0.388
lev*sub	737.3	11	67.0	0.9	0.546
bw*sub	3159.6	44	71.8	1.7	0.045
lev*bw*sub	468.0	44	10.6	1.0	0.460
dur*sub	3376.9	11	307.0	3.0	0.015
lev*dur*sub	788.6	11	71.7	7.0	≤0.0001
bw*dur*sub	1835.3	44	41.7	4.0	≤0.0001
lev*bw*dur*sub	453.8	44	10.3	1.2	0.162

level range, the effect of duration on spectral loudness summation does not depend on the reference level. The significant interactions of duration and subject ($p<0.05$) and of duration, subject, and level ($p<0.001$) show that duration effects depend on the subject and that those effects in turn depend on the level.

IV. DISCUSSION

On average, the amount of spectral loudness summation in normal-hearing listeners was found to depend both on level and signal duration. Similar results were obtained by both measuring procedures.

A. Comparison between loudness matching and loudness scaling

At a reference level of 65 dB SPL, the average level differences between equally loud 1000-ms signals are almost the same as those measured by loudness matching (except for a test signal bandwidth of 6400 Hz, which will be discussed later). The average ΔL between the 200-Hz-wide test signal and the reference is 5 dB for both procedures. In addition, both procedures exhibit a similar duration dependence of spectral loudness summation at this reference level.

At a reference level of 45 dB SPL, however, there are considerable differences between the average results of loudness matching and loudness scaling. For a duration of 1000 ms, the average ΔL between the 200-Hz-wide test signal and the reference is only 6 dB for loudness scaling compared to 12 dB for loudness matching. In addition, loudness scaling data show only a small difference between long and short signals at this level while a clear duration dependence of spectral loudness summation can be observed for the loudness matching data. The standard error of the loudness summation derived from loudness scaling data (Fig. 3) is considerably larger than that for the matching data (Fig. 5). Thus, the differences may be due to the greater standard error for loudness scaling which is a result of the large level range investigated and the limited number of loudness categories. In addition, at levels below about 40 dB the slope of the loudness function is very shallow and the number of loudness categories in this level range may be not sufficient (Brand and Hohmann, 2002), which would further increase the inaccuracy of the measurement.

A comparison between loudness matching and loudness scaling data shows large differences in the amount of spectral loudness summation between the 3200-Hz-wide and the 6400-Hz-wide signal. For loudness matching the amount of spectral loudness summation between these bandwidths is 11 dB at a level of 65 dB SPL compared to 4 dB for loudness scaling. This is in agreement with the results in Fruhmann *et al.* (2003) who used a 400-Hz-wide reference signal and found a spectral loudness summation between the 3200-Hz-wide and 6400-Hz-wide test signals of 5 dB at a reference level of 65 dB SPL. Thus, in contrast to Verhey (1999), the present data for the broadest test-signal bandwdith and a level of 65 dB SPL of the 3200-Hz-wide reference indicate that the loudness matching procedure may be slightly biased by the choice of the reference signal.

B. Comparison with data in the literature

Interindividual differences in the amount of spectral loudness summation such as those shown in Fig. 1 were already observed in previous loudness studies using categorical loudness scaling (Blum *et al.*, 2000; Brand and Hohmann, 2001). They were also found using other methods such as absolute magnitude estimation (e.g., Blum *et al.*, 2000) or loudness matching (e.g., Verhey and Kollmeier, 2002).

The average data resulting from the loudness scaling measurements show a level dependence in the amount of spectral loudness summation for the 1000-ms signals. A similar level dependence of spectral loudness summation for long signals is reported in the literature (e.g., Zwicker *et al.*, 1957; Wagner *et al.*, 2004). In qualitative agreement with the present loudness scaling data, Zwicker *et al.* (1957) and Wagner *et al.* (2004) found the largest spectral loudness summation for a level of the broadband signal of about 50 dB SPL and no spectral loudness summation for very high and very low levels (above 90 and below 30 dB SPL). For both procedures, the present data show a spectral loudness summation for the 1000-ms signals that was slightly larger for a reference level of 45 dB SPL than for a reference level of 65 dB SPL which qualitatively agrees with the data in Verhey and Kollmeier (2002).

Verhey and Kollmeier (2002) measured a 6 to 8 dB larger level difference between the 3200-Hz-wide reference and the 200-Hz-wide test signal for 10-ms signals than for 1000-ms signals at a reference level of 65 dB SPL. At the

same level, Fruhmann *et al.* (2003) measured a 5 dB larger level difference between the 400-Hz-wide reference and the 3200-Hz-wide test signal for the short signal. The effect of duration on the amount of spectral loudness summation in those two studies agrees well with the present data. In the present study, the difference in spectral loudness summation for short and long signals was 4 dB at a reference level of 65 dB SPL for both procedures.

C. Proposed modification of the equal-loudness-ratio hypothesis

The duration dependence of the amount of spectral loudness summation at the same reference level is at odds with the predictions of the equal-loudness-ratio hypothesis. In contrast to the data, it predicts the same amount of spectral loudness summation at the same reference level. The reason for the latter behavior is the assumption of a loudness ratio that is independent of the spectrum of the signal (see Fig. 2, right panel). In order to account for the duration dependence of spectral loudness summation found in the present study, a possible modification of the equal-loudness-ratio model might be the assumption that different level-independent loudness ratios exist for signals with different spectra (ratio increasing with decreasing bandwidth; see Fig. 2, left panel).

The dashed lines in Fig. 4 show model predictions assuming a bandwidth-dependent constant loudness ratio between long and short signals (thick and thin lines, respectively). Loudness functions for long signals were calculated from loudness scaling data averaged across 12 subjects. To obtain the loudness functions for short signals, those for long signals were shifted by a level-independent value. This value was derived from the data by averaging the ratio across levels between 45 and 105 dB SPL. Levels below 45 dB SPL were omitted because for these levels loudness ratios have been shown to increase [Epstein and Florentine (2005); see also Fig. 2, left panel]. The ratio between long and short signals was 3.0 ± 0.4 for a bandwidth of 200 Hz and 2.4 ± 0.2 for 3200-Hz signals. From the resulting loudness functions, level differences between narrow-band (200 Hz) and broadband (3200 Hz) stimuli needed to produce the sensation of equal loudness were calculated for both signal durations. Predictions and data for long signals slightly differ since the predictions are derived from average loudness functions. For both durations, model predictions are in reasonable agreement with the data for intermediate levels. For levels of the broadband stimulus $\geqslant$85 dB SPL a bandwidth-dependent loudness ratio cannot account for the data. This may indicate that for high levels the loudness ratio between long and short signals is not independent of level. Such a deviation of the constant loudness ratio above 85 dB seems to be inconsistent with the average data in Epstein and Florentine (2005). They found a nearly constant average ratio from 40 to 95 dB SL, although individual data show some deviation from this trend. However, they used a 1-kHz tone instead of a broadband noise. It is possible that the loudness ratio for narrowband signals is constant up to higher levels than for broadband signals. In general, a varying loudness ratio of only one of the stimuli considered in Fig. 3 will result in a mismatch between predictions and data in the present study.

V. SUMMARY AND CONCLUSION

The effect of duration on spectral loudness summation in normal-hearing listeners was investigated over a wide level range using categorical loudness scaling. In addition, loudness matching was used for two intermediate levels. In general, both procedures yielded similar results for these levels. Categorical loudness scaling is an effective method to investigate loudness over a large level range but is less accurate than loudness matching.

As already known from the literature, the amount of spectral loudness summation was found to be level dependent with a maximum effect at intermediate levels. On average, spectral loudness summation was larger for short signals (10 ms) than for long signals (1000 ms) at the same reference level. The duration dependence at the same level cannot be explained by the equal-loudness-ratio hypothesis.

The loudness ratio between long and short signals with the same bandwidth was found to be almost independent of level for levels greater than about 45 dB SPL. However, in contrast to the assumptions of the equal-loudness-ratio hypothesis, this loudness ratio was not found to be independent of the spectral content of the signal. For moderate levels, a corresponding modification of the equal-loudness-ratio hypothesis could in principle account for the duration dependence of spectral loudness summation in normal-hearing listeners.

ACKNOWLEDGMENTS

This study was supported by the Deutsche Forschungsgemeinschaft (DFG HO 1697/2-3).

APPENDIX A: MODEL FUNCTION FOR LOUDNESS SCALING DATA

The two linear parts with slopes m_l and m_h intersect at the point (L_c, c), where L_c denotes the level and c the categorical loudness value of the intersection point. The model function fitted to the data is then given by

$$f(L) = \begin{cases} \max[(c + abs(m_l)(L - L_c)), 0], & L \leqslant L_l, \\ (c - 10) + 20t(L), & L_{lo} < L < L_h, \\ \min[(c + abs(m_h)(L - L_c)), 50], & L \geqslant L_h, \end{cases} \tag{A1}$$

with

$$t(L)=\begin{cases}-\dfrac{x_1}{2x_2}-\sqrt{\dfrac{L-L_l}{x_2}+\dfrac{x_1^2}{4x_2^2}}, & \dfrac{L_c-(c-10)}{L_c-L_l}<\dfrac{c-(c+10)}{L_c+L_h},\\ -\dfrac{x_1}{2x_2}+\sqrt{\dfrac{L-L_l}{x_2}+\dfrac{x_1^2}{4x_2^2}}, & \dfrac{L_c-(c-10)}{L_c-L_l}>\dfrac{c-(c+10)}{L_c+L_h},\end{cases} \tag{A2}$$

and

$$x_1=2L_c-2L_l,$$

$$x_2=L_l-2L_c+L_h.$$

The equations for the linear parts ensure that slopes are always positive and that the loudness function has no values higher than 50 (corresponding to "too loud") or lower than 0 (corresponding to "inaudible").

Levels L_l and L_h at the beginning and end of the Bezier smoothing in the transition region between upper and lower linear part are given by

$$L_l=L_c-\frac{10}{m_l},$$

$$L_h=L_c+\frac{10}{m_h},$$

i.e., the Bezier interpolation stretches across an interval of 20 on the categorical unit axis centered around c. In Brand and Hohmann (2002) the parameter c was fixed at 25 cu.

To calculate level differences for a given loudness l the inverse of Eq. (A1) is needed:

$$f^{-1}(L)=\begin{cases}\dfrac{l-c}{abs(m_l)}+L_c, & 0\leqslant l\leqslant f(L_l),\\ x_2\left(t(l)+\dfrac{x_1}{2x_2}\right)^2-\dfrac{x_1^2}{4x_2^2}+L_l, & f(L_l)<l<f(L_h),\\ \dfrac{l-c}{abs(m_h)}+L_c, & f(L_h)\leqslant l\leqslant 50,\end{cases} \tag{A3}$$

with

$$t(l)=\frac{l}{20}-\frac{c-10}{20} \tag{A4}$$

and x_1, x_2 defined as above.

APPENDIX B: TRANSFORMATION OF CATEGORICAL UNITS INTO SONES

The present study uses the following transformations between sones and categorical units (cu) as suggested by Apell (2002):

$$cu'=0.3942\,\mathrm{lg}(\mathrm{sone})^3+3.2018\,\mathrm{lg}(\mathrm{sone})^2+9.7334\,\mathrm{lg}(\mathrm{sone})+11.2097, \tag{B1}$$

$$\mathrm{cu}=\max(0,\mathrm{cu}'). \tag{B2}$$

The equation assumes a cubic relation between the logarithmic sone scale and the categorical scale. It is a modified version of the transformation proposed in Blum *et al.* (2002). They assumed a quadratic relation between the logarithm of the sone values and the categorical units. Appell (2002) showed that the cubic transformation is a better approximation of the relation between categorical units and sones than the quadratic transformation [see Figure on p. 78 in Appell (2002)].

Appell, J.-E. (**2002**). "Loudness models for rehabilitative audiology," BIS-Verlag, Oldenburg, http://docserver.bis.uni-oldenburg.de/dissertation/2002/applou02/applou02.html

Blum, R., Hohmann, V., Appell, J. A., and Kollmeier, B. (**2002**). "Modelling categorical loudness perception for normal and hearing impaired listeners," Z. Audiol. **41**, 76–89.

Blum, R., Hohmann, V., and Kollmeier, B. (**2000**). "Loudness differences caused by narrow-band and broad-band stimulation (loudness summation) in normal and hearing impaired listeners," Z. Audiol. **39**, 119–127.

Brand, T., and Hohmann, V. (**2001**). "Effect of hearing loss, centre frequency, and bandwidth on the shape of loudness functions in categorical loudness scaling," Audiology **40**, 92–103.

Brand, T., and Hohmann, V. (**2002**). "An adaptive procedure for categorical loudness scaling," J. Acoust. Soc. Am. **112**(4), 1597–1604.

Buus, S., Florentine, M., and Poulsen, T. (**1997**). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am. **101**(2), 669–680.

Cacace, A. T., and Margolis, R. H. (**1985**). "On the loudness of complex stimuli and its relationship to cochlear excitation," J. Acoust. Soc. Am. **78**(5), 1568–1573.

Ekman, G., Berglund, B., and Berglund, U. (**1966**). "Loudness as a function of the duration of auditory stimulation," Scand. J. Psychol. **7**, 201–208.

Epstein, M., and Florentine, M. (**2005**). "A test of the equal-loudness-ratio hypothesis using cross-modality matching functions," J. Acoust. Soc. Am. **118**(2), 907–913.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**(3), 1633–1644.

Fruhmann, M., Chalupper, J., and Fastl, H. (**2003**). "Zum Einfluss von Innenohrschwer hörigkeit auf die Lautheitssummation," in *Abstract of the 29th Annual Meeting for Acoustics, Aachen*, p. 253.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence of the role of envelope fluctuations," Acust. Acta Acust. 83, 659–669.

Port, E. (**1963**). "Über die Lautstärke einzelner kurzer Schallimpulse," Acustica **13**, 212–223.

Poulsen, T. (**1981**). "Loudness of tone pulses in a free field," J. Acoust. Soc. Am. **69**(6), 1786–1790.

Scharf, B. (**1959**). "Loudness of complex sounds as a function of the number of components," J. Acoust. Soc. Am. **31**(6) 783–785.

Schneider, B. (**1988**). "The additivity of loudness across critical bands: A conjoint measurement approach," Percept. Psychophys. **43**, 211–222.

Takeshima, H., Suzuki, Y., Kono, S., and Sone, T. (**1988**). "Growth of the loudness of a tone burst with a duration up to 10 seconds," J. Acoust. Soc. Jpn. **9**(6), 295–300.

Verhey, J. L. (**1999**). "Psychoacoustics of spectro-temporal effects in masking and loudness perception," BIS-Verlag, Oldenburg (ISBN 3-8142-0662-

2), http://docserver.bis.uni-oldenburg.de/publikationen/bisverlag/verpsy99/verpsy99.html
Verhey, J. L., and Kollmeier, B. (**2002**). "Spectral loudness summation as a function of duration," J. Acoust. Soc. Am. **111**(3), 1349–1358.
Wagner, E., Florentine, M., Buus, S., and McCormack, J. (**2004**). "Spectral loudness summation and simple reaction time," J. Acoust. Soc. Am. **116**(3), 1681–1686.
Zwicker, E. (**1966**). "Ein Beitrag zur Lautstärkemessung impulshaltiger Schalle," Acustica **17**, 11–22.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics* (Springer, Berlin).
Zwicker, E., and Feldtkeller, R. (**1955**). "Über die Lautstärke von gleichförmigen Geräuschen," Acustica **5**, 303–316.
Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). "Critical band width in loudness summation," J. Acoust. Soc. Am. **29**(5), 548–557.

Discrimination of nonlinear frequency glides

Nick Thyer
School of Healthcare, University of Leeds, United Kingdom

Doug Mahar
School of Psychology and Counseling, Queensland University of Technology, Australia

(Received 26 May 2005; revised 6 March 2006; accepted 7 March 2006)

Discrimination thresholds for short duration nonlinear tone glides that differed in glide rate were measured in order to determine whether cues related to rate of frequency change alone were sufficient for discrimination. Thresholds for rising and falling nonlinear glides of 50-ms and 400-ms duration, spanning three frequency excursions (0.5, 1, and 2 ERBs) at three center frequencies (0.5, 2.0, and 6.0 kHz) were measured. Results showed that glide discrimination was possible when duration and initial and final frequencies were identical. Thresholds were of a different order to those found in previous studies using linear frequency glides where endpoint frequency or duration information is available as added cues. The pattern of results was suggestive of a mechanism sensitive to spectral changes in time. Thresholds increased as the rate of transition span increased, particularly above spans of 1 ERB. The Weber fraction associated with these changes was 0.6–0.7. Overall, the results were consistent with an excitation pattern model of nonlinear glide detection that has difficulty in tracking signals with rapid frequency changes that exceed the width of an auditory filter and are of short duration. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2191587]

PACS number(s): 43.66.Fe, 43.66.Ba [JHG] Pages: 2929–2936

I. INTRODUCTION

The discrete frequency and duration cues that define rate of change of frequency in linear frequency glides are confounded. This makes it difficult to differentiate between mechanisms responsible for glide discrimination that are sensitive to change *per se* and those related to simply duration or frequency coding. Research has generally concentrated on discrimination of either unidirectional frequency glides (Dooley and Moore, 1988a, b; Madden and Fire, 1996, 1997; Moore and Sek, 1998; Nabelek and Hirsch, 1969; Tsumura and Nimura, 1972) or periodic frequency modulated (FM) signals (Carlyon, 2000; Gardener and Wilson, 1979; Gockel, Moore, and Carlyon, 2001; Kay and Matthews, 1972; Moore and Sek, 1994; Moore and Sek, 1995b, 1996; Plack and Carlyon, 1994, 1995; Sek and Moore, 2000; Tansley and Suffield, 1983). Discrimination of both forms of signal might well require the detection of frequency rate change, but it does not follow that there is a common underlying mechanism for frequency glide and FM perception. In this paper we measured discrimination thresholds for nonlinear tone glides with identical end-point frequencies and duration where only rate of frequency change information was available so that participants' performance should reflect sensitivity to rate of change of frequency.

The perception of FM involves monitoring a continuous and cyclic modulation of frequency which, given sufficient periods of the modulating frequency, affords continuous rate of change information. Place coding theories (e.g., Zwicker, 1970) predict that FM discrimination is achieved by monitoring changes in the distribution of rate-place related activity on the basilar membrane and other structures in the auditory system where differences in the excitation patterns that arise from stimulation relate to perceived frequency changes. In this case threshold values would be linked to the more steeply sloping low frequency side of the excitation pattern where changes in level are greatest (Zwicker, 1970), or to multiple points from both sides of the excitation pattern (Moore and Sek, 1994). In either case the slopes of the excitation patterns for a particular center frequency are associated with the width of the auditory filter at that center frequency. Thus FM difference limens should be a constant proportion of the equivalent rectangular bandwidth (ERB) of the auditory filter at that carrier frequency (Glasberg and Moore, 1990). Thresholds at modulation rates of 10 Hz and above do show such a pattern (Moore and Sek, 1995b, 1996). At modulation rates below 5 Hz, and in particular where the carrier or center frequency is below 5 kHz, they do not and results are similar to those obtained from steady frequency difference limens (FDL). These results are more easily explained by a mechanism that utilizes information contained in temporal patterns of the auditory nerve (Moore and Sek, 1995b, 1996; Sek and Moore, 2000). This phase locking typically fails above 4–5 kHz and is unlikely to be involved over the entire frequency range of hearing which accounts for the elevated FDLs found at and above 6 kHz.

In contrast, unidirectional glide discrimination involves comparing two or more signals which only afford the listener discrete frequency change information over the signal duration. A possible mechanism underlying linear glide discrimination is one that allows a listener to perform a simple frequency or duration discrimination. Dooley and Moore (1988b) measured thresholds for static frequency discrimination and linear frequency glide detection where the rate of change of frequency changed as a function of *either* frequency *or* duration and thus either endpoint frequency or duration was available as a cue. Glide detection thresholds followed the general pattern expected for static FDLs and

were observed both as a function of frequency and duration at least up to 4 kHz. These results are more consistent with a temporal process related primarily to available static end-point frequency information and not differences in excitation pattern level or rate of change of frequency.

If excitation pattern monitoring was used in glide detection then a pattern of results like Zwicker's FM DL data would be expected. In order to test this, Madden and Fire (1996, 1997) used a roving design where the center frequencies of standard and target linear glides were uniformly varied to limit the use of end-point frequencies and the likelihood of invoking a simple frequency discrimination mechanism. They found that thresholds expressed as a constant proportion of the ERB varied little with center frequency and that duration had little effect on discrimination (the duration of their stimuli was 50 and 400 ms, frequency spans were 0.5, 1, and 2 ERBs, and their center frequencies were 0.5, 1.0, 2.0, 4.0, and 6.0 kHz). These data were taken as evidence for a rate of change processing mechanism that monitors excitation pattern changes. Thresholds were similar for the 0.5 and 1 ERB spans but worsened markedly for the 2 ERB spans. Performance more or less adhered to Weber's Law above transitions of 1 ERB.

Moore and Sek (1998) replicated and extended the Madden and Fire studies. Their data was not entirely consistent with an excitation pattern model nor was it sufficient to implicate the use of phase locking. To make it difficult to use cues based on monitoring changes in excitation levels, Moore and Sek added random glides in level to the glides in frequency and predicted that in these conditions glide discrimination thresholds would be raised. Small and moderate glides in level did not greatly affect performance. However, performance where large glides in level were present was actually better than theoretical limits would predict assuming an excitation-pattern model (Glasberg and Moore, 1990). They argued that phase locking might play a part in glide discrimination, particularly for slower transitions with lower center frequencies, but an excitation-pattern model could account for glide discrimination for larger frequency spans of 2 ERBs since the change in excitation pattern due to frequency change was large in comparison with changes they introduced in level.

In a further study Sek and Moore (1999) measured frequency DLs for pairs of different frequency tone bursts with center frequencies of 0.5, 1, 2, 4, and 6 kHz that were separated by either 5 ms of silence or were joined by glides of different durations between 0 and 500 ms. In one condition the center frequency was fixed in each trial and in another the center frequency was roved over trials with a range of ±2 ERBs. In a third condition cues based on changes of excitation level were disrupted by introducing random linear glides in level. Thresholds were similar across all center frequencies. Performance was poorer when the center frequency was roved. Sek and Moore interpreted this finding in two ways. First, that a fixed center frequency allowed listeners to monitor the start and end frequencies. Second, randomization of center frequencies was simply distracting or made the generation of an internal representation of the signals difficult. In either case thresholds in the roved condition were lower than the theoretical limit using only end-point information. Glide detection was also impaired by glides in level but was also better than would be predicted if listeners were simply using excitation pattern information. Importantly, when the 5 ms of silence were replaced with a 5-ms glide, thresholds improved significantly. They concluded that this was evidence for the additional involvement of a glide detection mechanism that was distinct from excitation pattern or end-point frequency monitoring.

Sek and Moore (1999) suggested that below 4 kHz there are two mechanisms available for glide discrimination and detection; a temporal mechanism that samples the end points combined with a glide detection mechanism based on excitation patterns. Since, in their study the glide duration lengthened at the cost of the steady state plateau duration, the prediction would be that the longer the glide the less end-point information would be available. Thresholds were poorer for long glides and better for short glides where both kinds of information were available. Nevertheless, linear glides that do not have start and finish plateaus such as the Sek and Moore 500-ms stimuli still have end-point cues available (Pollack, 1968, Dooley and Moore, 1988a, b; Madden and Fire, 1996, 1997) which makes it difficult to interpret their data in terms of how absolute rate of change of frequency is detected without the influence of end-point cues.

In the present study, glide discrimination was tested using pairs of stimuli that had identical end-point frequencies and durations. This allowed the direct measurement of listener's sensitivity to different glide rates over a range of frequency spans, durations, and center frequencies without the confounding availability of end-point cues. Previous studies have not convincingly done this and their conclusions regarding underlying mechanisms involved in glide detection are therefore questionable. It was hypothesized that participants would be able to detect differences in rate of change of frequency even though duration and end-point values were identical between and within trials. Further, it was predicted that nonlinear glide performance would be poorer than that reported by studies using linear glides because of the complete unavailability of end-point cues to augment performance. Finally, it was predicted that if an excitation pattern model is involved in processing rate of change of frequency then thresholds would generally be a constant proportion of the ERB irrespective of center frequency.

II. METHOD

A. Participants

All participants (9 females and 11 males) had hearing thresholds better than 20 dBHL for octave frequencies between 250 and 8000 Hz. The group mean age was 21 years [standard deviation (SD)=4.3].

B. Stimuli

The stimuli used in this study were nonlinear frequency glides. The common stimulus parameters are shown in Table I. Duration (t) refers to the duration of the gliding portion of the stimulus. The center frequency (fc) is defined as the frequency midway between start frequency (sf) and the end

TABLE I. Common stimulus parameters of center frequency *fc* and frequency transition span *ft* expressed in ERB and hertz for nine conditions. Four sets of stimuli were synthesized using these nine core values making 36 conditions in all. These were (1) gliding up in 50 ms, (2) gliding down in 50 ms, (3) gliding up in 400 ms, and (4) gliding down in 400 ms. The center frequency *fc* is midway between start and end frequencies at $t/2$.

Condition	*fc* (kHz)	Transition span *(ft)* in ERBs	Transition span *(ft)* in hertz
1	0.5	0.5	39
2	2.0	0.5	120
3	6.0	0.5	336
4	0.5	1.0	79
5	2.0	1.0	241
6	6.0	1.0	672
7	0.5	2.0	157
8	2.0	2.0	481
9	6.0	2.0	1345

frequency (*ef*) at $t/2$. All stimuli had 25-ms onsets and offsets during which frequency remained steady at either *sf* or *ef*. Stimulus amplitudes rose from zero to maximum values in a half hanning function. The gliding portion of the frequency glides in each condition is described by the general form

$$\int \sin[2\pi(aft^3 + bft^2 + cft + d)]dt.$$

The nonlinear glide tracks were generated by evaluating the roots of the equation that defined a frequency glide between *sf* and *ef* which at, $t/2$, passed through either $fc+S$ Hz for the upper stimulus (shown by the dashed line in Fig. 1) or $fc-S$ Hz for the lower stimulus (shown by the thin solid line in Fig. 1). Coefficients were then calculated that defined frequency tracks that passed through center frequencies at intervals of $S=fc\pm1\%$ of the frequency span. When

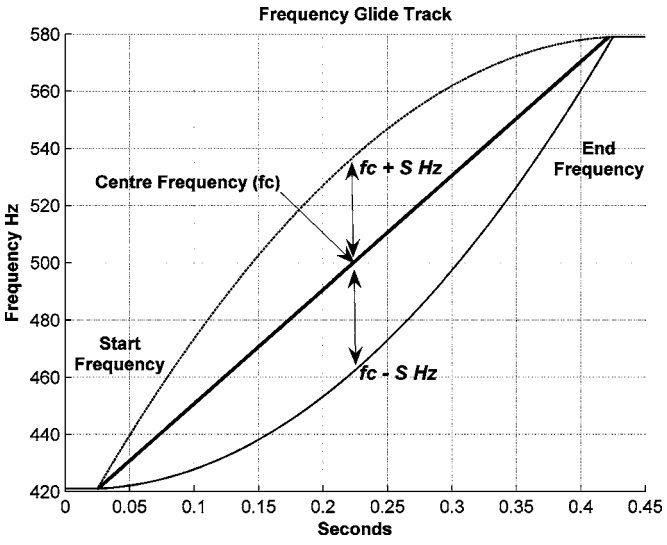


FIG. 1. Schematic view of frequency changes associated with the nonlinear glide stimuli: *fc* is the stimulus center frequency. The distance in hertz between the midpoint frequencies of stimuli in the standard and signal intervals of a trial is given by $(fc+S)-(fc-S)$ and is the raw value used to express threshold in ERBs.

$S=0$ the function is linear and both upper and lower glides are identical and pass through *fc* as shown by the thick solid line in Fig. 1. Threshold was defined in terms of the smallest discriminable difference between an upper and a lower glide expressed as the difference in their center frequencies. A step size for *S* which was a constant percentage of the transition span was chosen since preliminary data indicated that thresholds were greater for larger transition spans. The index *S* ranged from 0 to 90 % of the transition span.

The complete set of tone glides for each of the 36 conditions is described in Table I. Frequency span ERB values were calculated using the formula $\text{ERB}=24.7(4.37fc+1)$ from Glasberg and Moore (1990). Stimuli were digitally synthesized and stored on disk and output using an AWE 16-bit D/A converter controlled by a proprietary PC computer. Sampling rate was 20 kHz with 16 bit resolution. Participants listened individually to the stimuli over Sennheiser 490D headphones, presented binaurally at 68 dB sound pressure level (SPL).

C. Procedure

Stimuli from each of the 36 conditions were presented by computer in a same-different paradigm. Participants practiced for at least 2 h in all 36 conditions before data collection. The experiment was self-paced with the order of condition presentation sequentially balanced. Participants started each trial by clicking a single "present sounds" button displayed on a computer screen before hearing the first trial's stimulus pair. Following the presentation of each trial, a second screen comprising two buttons marked "same" or "different" was displayed and participants indicated their decision by clicking a mouse pointer on the appropriate button. The computer recorded the response, calculated the next stimulus level, and loaded two stimuli from disk for the next trial. When ready to proceed, participants clicked a further button and the next stimulus pair was presented.

On 50% of trials, the upper frequency glide was presented in the first of two intervals and its lower opposite presented in the second. The reverse was true for the remainder of the trials. Difference limen were expressed in terms of *S* as ERBs and were determined using an up-down-transformed-rule (UDTR) procedure to estimate the 70.7% correct point on the psychometric function (Levitt, 1971). An initial step size of $2S$ Hz was used and the procedure was terminated when a step size of *S* was called for. The geometric mean value of *S* over the last four reversals with a step size of *S* Hz was taken as threshold. Thresholds were calculated from the average of two repetitions of this procedure. The interstimulus interval was 500 ms. Duration, center frequency, transition span, and direction were fixed within a block of trials.

III. RESULTS

A. Glide discrimination

Figure 2 shows mean thresholds, expressed in ERBs, with bars representing transition span plotted as a function of center frequency. In general, thresholds increased as the frequency transition span increased from 0.5 to 2 ERBs. This

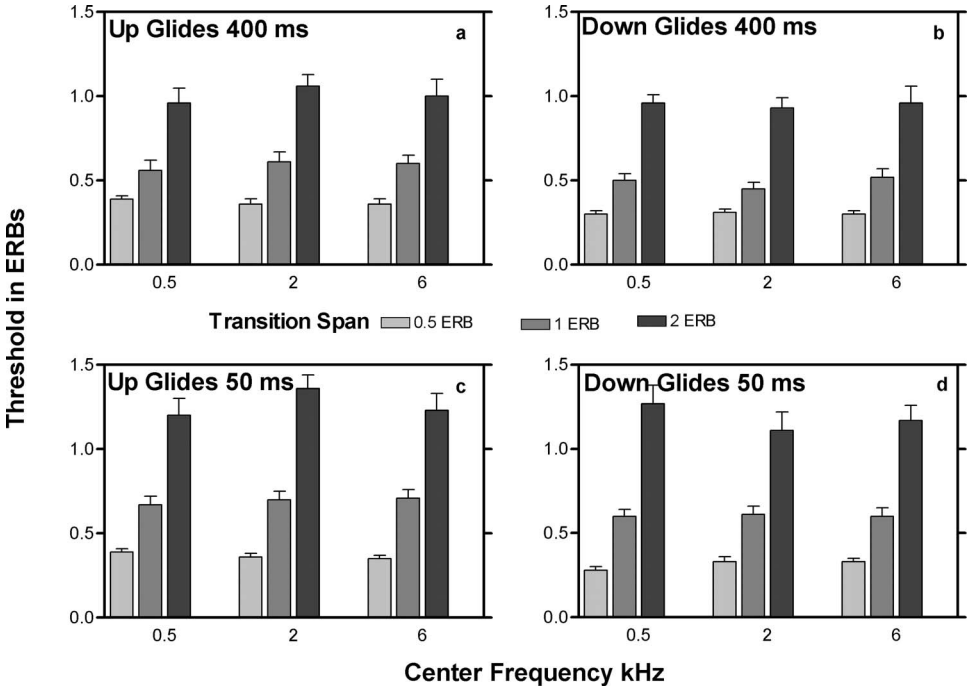


FIG. 2. Mean thresholds for discriminating nonlinear frequency glides expressed in ERBs plotted as a function of center frequency for three transition spans: 0.5, 1.0, and 2.0 ERBs. Error bars are ±2 standard errors of the mean.

pattern was present irrespective of center frequency but was more pronounced for shorter duration glides and for up glides. Figure 3 shows mean thresholds expressed in ERBs, plotted as a function of transition span with duration and direction as parameters, respectively. For 50-ms glides, thresholds became poorer in comparison to those for 400-ms glides as frequency span increased from 0.5 to 2 ERBS irrespective of direction, and thresholds for up glides were marginally better than those for down glides. To statistically test these observations a repeated measures analysis of variance (ANOVA) was performed with two levels of direction and duration and three levels of center frequency and frequency transition span as factors. Thresholds are expressed as a proportion of the equivalent rectangular bandwidth (ERB). As initial inspection suggested, the results revealed significant main effects of transition span, glide direction, and glide duration. Overall mean thresholds increased significantly as the transition span increased ($F_{2,17}=168.39$, $p<0.001$), thresholds were significantly better for down glides than up glides ($F_{1,18}=5.45$, $p<0.05$) and were better by 0.12 ERBs for the 400-ms glides compared to the 50-ms glides ($F_{1,18}=6.36$, $p<0.05$). Since the variation in thresholds was independent of center frequency, the data was collapsed across center frequency for further analysis.

There was a significant interaction between frequency

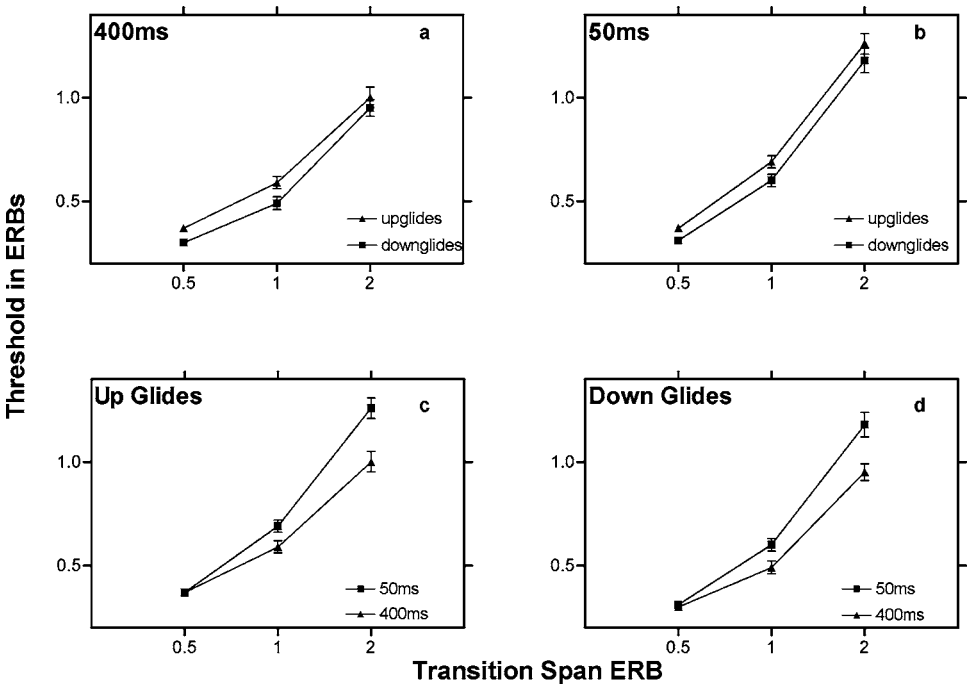


FIG. 3. Panels (a) and (b) results comparing mean thresholds expressed in ERBs for 400- and 50-ms glides as a function of transition span. Panels (c) and (d) results comparing mean thresholds expressed in ERBs for down glides and up glides as a function of transition span. Data is collapsed across center frequency and error bars are ±2 standard errors of the mean.

transition span and glide duration ($F_{2,17}=5.67$, $p<0.05$). To determine the source of the simple effects, pairwise comparisons of the duration marginal mean thresholds for each transition span were made using a bonferroni protected alpha level of $p=0.05$. Inspection of Fig. 3 [panels (a) and (b)] shows mean threshold differences between 0.5, 1, and 2 ERB frequency transition spans that were all significantly different from each other in both the 400-ms glides ($F_{2,17}=71.45$, $p<0.001$) and the 50-ms glides ($F_{2,17}=111.81$, $p<0.001$).

Figure 3 [panels (c) and (d)] shows that for down glides, 400-ms thresholds were significantly better for 1 ERB ($F_{1,18}=4.50$, $p<0.05$) and 2 ERB ($F_{1,18}=4.79$, $p<0.05$) spans than 50-ms glide thresholds (mean difference 0.12 and 0.23 ERB, respectively). For up glides, 400-ms glide thresholds were again better at 1 and 2 ERB spans in comparison to 50-ms glide thresholds, but were only significantly so at the 2 ERB transitions ($F_{1,18}=9.23$, $p<0.01$: mean difference 0.262 ERB). For all transition spans, thresholds for down glides were on average, worse by a factor of 0.07 ERBs. These differences were only significant for the 400-ms 0.5 ERB span ($F_{1,18}=17.51$, $p<0.001$: mean difference=0.07 ERBs) and the 50-ms 1 and 2 ERB spans ($F_{1,18}=8.81$, $p<0.05$, mean difference=0.059 ERBs and $F_{1,18}=4.86$, $p<0.05$, mean difference=0.09 ERBs, respectively).

B. Excitation level

Time-frequency modeling (O'Donovan and Furlong, 2005; Patterson and Allerhand, 2005) indicated a clear difference in the level of excitation elicited by each of the upper and lower glides when stimulus differences were at threshold values. This is most clearly seen in the analysis of 2 kHz glides with a 2 ERB transition shown in Fig. 4. Panels (a) and (b) indicate the patterns evoked by the upper and lower glides, respectively, presented in the standard and signal intervals of a threshold trial. Maximal differences between the excitation patterns produced by the standard and signal stimuli at threshold and for all conditions, were calculated using the method described in Moore and Glasberg (1987), Glasberg and Moore (1990), and Moore, Glasberg, and Baer (1997). The maximum differences in decibel excitation level on the low and high sides of the excitation patterns were not significantly different across center frequency for low frequency and high frequency sides at 400 ms ($F_{1,12}=3.59$, $p>0.05$; $F_{1,2}=0.03$, $p>0.05$, respectively) or for low frequency and high frequency sides at 50 ms ($F_{1,2}=3.63$, $p>0.05$; $F_{1,2}=1.16$, $p>0.05$, respectively). Therefore the data was collapsed across frequency and the maximum difference level on each side of the excitation pattern calculated for the three transition spans, two durations, and directions. These are shown in Fig. 5. The data show differences that are predominantly larger on the low frequency side and increase as transition spans increase.

IV. DISCUSSION

A. General discussion

This study showed that glide discrimination is possible when differences in end-point frequency are unavailable as physical cues. Performance was relatively unaffected by glide center frequency and seemed to depend on the frequency transition span with thresholds becoming poorer as frequency excursion increased. Thresholds for 2 ERB transitions were almost twice as large as those for 1 ERB transitions and three times as large as those for 0.5 ERB transitions. This trend was reduced in the longer duration signals and seemed to be more marked for upward glides than downward glides.

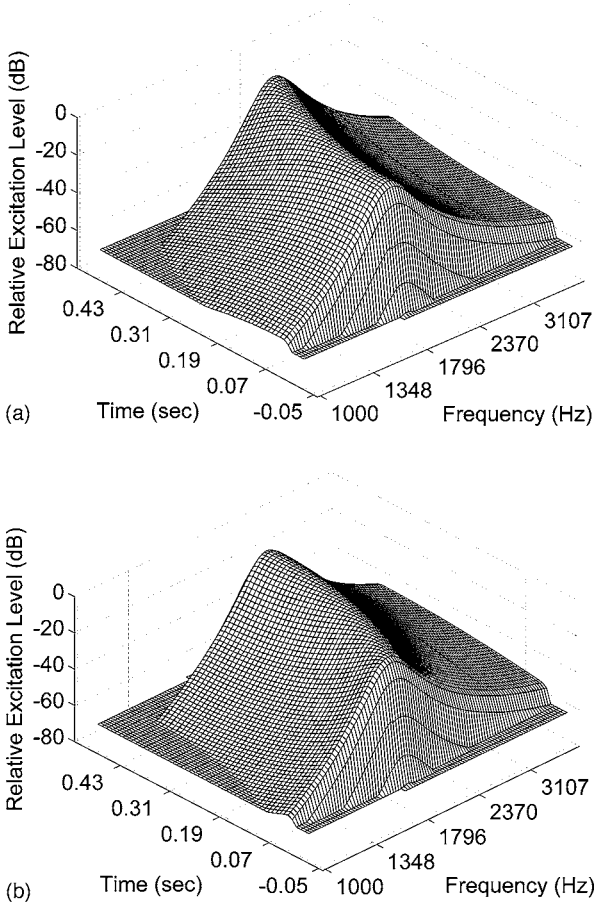


FIG. 4. A model of the excitation pattern elicited by an upper glide (panel a) and a lower glide (panel b) which glide upward with *fc* of 2 kHz and a transition span of 2 ERBs. The stimuli represent a pair of glides at threshold values. The time frequency analysis method used is described by O'Donovan and Furlong (2005).

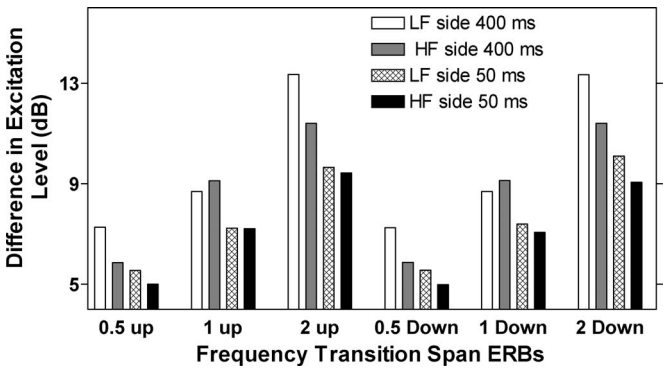


FIG. 5. Maximum difference in level (decibels) between excitation patterns elicited by the upper and lower nonlinear frequency glides that correspond to threshold. Data is given for the low and high frequency sides of the excitation pattern for 50- and 400-ms glides.

In the present study, listeners had no alternative but to use glide information whereas in previous studies (e.g., Moore and Sek, 1998; Madden and Fire, 1996, 1997; Dooley and Moore, 1988a, b) thresholds may have been determined by differences in end-point frequency as well as rate of change information. Limited contrasts with previous work can be drawn. Moore and Sek (1998) used a roving technique to limit listener's use of end-point information in glide discrimination. Their threshold data in comparable conditions were generally of the same order as those reported in this study, indicating that their strategy was successful. The Madden and Fire (1997) threshold data from comparable conditions were also generally similar. But, their absolute thresholds were substantially better than the current data and to a lesser extent the Moore and Sek (1998) data. Madden and Fire roved their stimuli over a much smaller range than Moore and Sek which may have afforded their listeners a greater opportunity to supplement their performance by using end-point cues. One possibility in the present study is that listeners were able to focus on center frequency differences at the midpoint of the stimuli and make a simple frequency discrimination. If this strategy was adopted we would expect to see a pattern of results in the data that reflect frequency DLs that typically vary with center frequency (Moore and Glasberg, 1989; Palmer, 1995) and this was not the case.

B. Modeling perception of nonlinear glides

Evidently listeners can discriminate nonlinear glides using rate of frequency change information, but how do they do this? Place models assume that glide or FM discrimination is performed by monitoring one or both sides of the excitation patterns evoked by the glide stimuli. The task of discriminating an upward linear glide in frequency for example, could be achieved by selecting the signal interval that has the larger decrease in the time-averaged excitation level (Moore and Sek, 1998). The time-averaged frequency content of the upper nonlinear glide stimulus is identical to the lower nonlinear glide stimulus. These stimuli, however, evoke a difference in the rate at which the excitation develops along the basilar membrane. Figure 4 shows that at any instant during upward glide stimulation the developing excitation pattern of the upper glide will always be advanced in terms of the place of maximum excitation on the basilar membrane in comparison with that of the lower glide. This leads to a greater decrease in excitation level on the low frequency side and a greater increase in excitation level on the high frequency side. Since thresholds reported in the present study did not vary significantly with center frequency and are therefore not consistent with a temporal model and we can interpret the data as evidence for a pattern excitation model for glide discrimination. This is consistent with the findings of Madden and Fire (1996, 1997) and of Moore and Sek (1998) suggesting that performance may primarily be based on place information derived from level changes in the excitation pattern.

However, for an upward glide over 0.5 ERBs a difference of at least 7.3 dB in the low frequency side of the excitation pattern is needed to detect a difference between the upper and lower glides. In contrast, for stimuli of the same duration and center frequency which glide over 2 ERBs, 13.3 dB is needed in the low frequency sides of the excitation pattern at threshold. If one only needed to monitor differences in the excitation pattern elicited by each stimulus pair then it would be expected that thresholds for discriminating glides would be similar irrespective of transition span. Similarly thresholds for larger spans would be expected to be better by at least a factor of two than the data indicate since detecting changes of at least 1 dB in the excitation pattern is usually possible (Zwicker, 1970). Differences of less than 13 dB should have been easily detected in 2 ERB transitions suggesting that monitoring excitation patterns might not have been the optimum strategy. It might be that information about differences in the developing excitation pattern is unavailable when the glide rate is too fast. With constant duration, larger transition spans mean a faster rate of change of frequency (and place related excitation level) and if these cues are involved in glide discrimination this might explain increases in threshold with increased transition span and therefore rate, found in the present study and others (Madden and Fire, 1997; Moore and Sek, 1998).

Part of the function of a glide discrimination mechanism must be to integrate activity over time. If we consider our nonlinear glides to be equivalent to the rising (or falling) portion of a frequency modulated tone with a carrier frequency of fc, length 1/4 of the modulating period, and with a modulation depth equivalent to the frequency transition span, then a comparison of rates of change can be made. The 400-ms glides in all conditions are roughly equivalent to the rate of change in frequency that would be produced by a modulation rate of about 0.63 Hz and the 50-ms glides to a modulation rate of about 5 Hz. It has been argued that for low frequency modulation rates (below 2 Hz) a mechanism based on phase locking operates but above modulation rates of 10 Hz a place mechanism operates (Moore and Sek, 1995a, 1996). It is therefore possible that, like low rates of FM, the temporal information contained in the low center frequency 400-ms glides with small transition spans could facilitate monitoring of changes in the instantaneous neural phase-locking pattern which signals frequency change. At modulation rates above 5 Hz the temporal mechanism is less responsive and at 10 Hz a place mechanism takes over since the rate of frequency change is too quick for a temporal mechanism to monitor. It could be argued that if a temporal mechanism based on phase locking is involved in glide discrimination, then at fast rates of change equivalent to 5 Hz modulation it is starting to flag but place information is starting to become available and supplements performance. We could further speculate that when the glide rate becomes even faster, comparison of excitation pattern level becomes difficult and it is this that is reflected in the 1 and 2 ERB threshold data. There is some support for this notion in the present data in that mean thresholds for the 50-ms glides were poorer than those for the 400-ms glides and for the larger transitions.

Our data indicates that glide discrimination is more difficult with increasing rates of spectral change. Thresholds for

all center frequencies are generally greater when the rate of change is fastest. Plack and Carlyon (1994) found that discrimination of modulation depth of 5 Hz FM became poorer as depth and consequently rate of change increased. In this study duration was kept constant within trial blocks such that any resultant changes in rate between the 50-ms and 400-ms conditions were unrelated to overall stimulus duration or frequency span on a given discrimination trial. If an underlying mechanism was based on rate of frequency change it should adhere to Weber's law as suggested by Dooley and Moore (1988a, 1988b) and the Weber fraction for rate[1] should be constant. In our study, while performance was poorer for the 50 ms compared to the 400-ms glide for the longer transition spans of 1 and 2 ERBs, the Weber fraction was fairly constant across the ERB transitions ranging from 0.6–0.7 for 50-ms glides and 0.7 for the 400-ms glides. These findings could be interpreted as being consistent with a mechanism that monitors the rate of spectral change.

Finally, thresholds for rising glides in the present study are generally better than for falling glides particularly for the 50-ms 1 and 2 ERB transitions, although the maximum differences in excitation patterns calculated at threshold for these are of a similar order (between 3 and 12 dB and increase with transition span). Madden and Fire (1997) found higher thresholds for 50-ms down glides (but only for glide detection not for glide discrimination) and Moore and Sek (1998) found no clear evidence of direction effects. The excitation patterns calculated from rising glides and their mirror image falling glides are more or less identical providing few place cues when of the same duration, center frequency, or transition span. An explanation for this finding might be that in gliding from high to low frequencies the traveling wave on the basilar membrane masks decreases in excitation level whereas for glides from low to high frequencies it does not.

V. CONCLUSIONS

The study presented here investigated glide discrimination where duration and end-point cues were unavailable. Our results have shown that nonlinear pairs of glides with identical end-point frequencies and duration are discriminable by their frequency change characteristics. We have shown that thresholds expressed as a proportion of ERB do not alter as a function of center frequency which is commensurate with a place model of processing rather than one related to simple frequency discrimination. We found that thresholds increased with change in glide rate, whether engineered by the manipulation of duration or by frequency transition span, although the increase was mainly associated with transition span. Threshold increase was most marked for the 2 ERB transition but Weber fractions for all transitions were fairly similar, being between 0.6 and 0.7. In consideration of previous studies' findings (Madden and Fire, 1997; Moore and Sek, 1998; Sek and Moore, 1999) our results could be interpreted as a reflection of a glide discrimination mechanism that monitors differences in the integrated excitation pattern level where these differences become less available as glide rate increases and the ability to successfully integrate excitation level information related to frequency change becomes limited.

[1]ERB/(Transition $\times$ Sec^{-1}).

Carlyon, R. P. (**2000**). "Detecting coherent and incoherent frequency modulation," Hear. Res. **140**(1–2), 173–188.

Dooley, G. J., and Moore, B. C. J. (**1988a**). "Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change," J. Acoust. Soc. Am. **84**(4), 1332–1337.

Dooley, G. J., and Moore, B. C. J. (**1988b**). "Detection of linear frequency glides as a function of frequency and duration," J. Acoust. Soc. Am. **84**(6), 2045–2057.

Gardener, R. B., and Wilson, J. P. (**1979**). "Evidence of direction specific channels in the processing of frequency modulation," J. Acoust. Soc. Am. **66**, 704–709.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103–138.

Gockel, H., Moore, B. C. J., and Carlyon, R. P. (**2001**). "Influence of rate of change of frequency on the overall pitch of frequency-modulated tones," J. Acoust. Soc. Am. **109**(2), 701–712.

Kay, R. H., and Matthews, D. R. (**1972**). "On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones," J. Physiol. (London) **225**, 657–677.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **44**, 467–477.

Madden, J. P., and Fire, K. M. (**1996**). "Detection and discrimination of gliding tones as a function of frequency transition and center frequency," J. Acoust. Soc. Am. **100**, 3754–3760.

Madden, J. P., and Fire, K. M. (**1997**). "Detection and discrimination of frequency glides as a function of direction, duration, frequency span, and center frequency," J. Acoust. Soc. Am. **102**(5, Pt 1), 2920–2924.

Moore, B. C. J., and Glasberg, B. R. (**1987**). "Formulae describing frequency selectivity as a function of frequency and level and their use in calculating excitation patterns," Hear. Res. **28**, 209–225.

Moore, B. C. J., and Glasberg, B. R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency," J. Acoust. Soc. Am. **86**, 1722–1732.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness," J. Audio Eng. Soc. **45**(4), 224–240.

Moore, B. C. J., and Sek, A. (**1994**). "Effects of carrier frequency and background noise on the detection of mixed modulation," J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., and Sek, A. (**1995a**). "Effects of carrier frequency, modulation rate and modulation waveform on the detection of modulation and the discrimination of modulation type (AM vs. FM)," J. Acoust. Soc. Am. **97**, 2468–2478.

Moore, B. C. J., and Sek, A. (**1995b**). "Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude versus frequency modulation)," J. Acoust. Soc. Am. **97**(4), 2468–2478.

Moore, B. C. J., and Sek, A. (**1996**). "Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking," J. Acoust. Soc. Am. **100**, 2320–2331.

Moore, B. C. J., and Sek, A. (**1998**). "Discrimination of frequency glides with superimposed random glides in level," J. Acoust. Soc. Am. **104**(1), 411–421.

Nabelek, I., and Hirsch, I. J. (**1969**). "On the discrimination of frequency transitions," J. Acoust. Soc. Am. **45**(6), 1510–1519.

O'Donovan, J. J., and Furlong, D. J. (**2005**). "Perceptually motivated time-frequency analysis," J. Acoust. Soc. Am. **117**(1), 250–262.

Palmer, A. R. (**1995**). "Neural signal processing," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).

Patterson, R. D., and Allerhand, M. H. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**(4), 1890–1894.

Plack, C. J., and Carlyon, R. P. (**1994**). "Detection of differences in the depth of frequency modulation," J. Acoust. Soc. Am. **96**(1), 115–125.

Plack, C. J., and Carlyon, R. P. (**1995**). "The detection of 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.
Pollack, I. (**1968**). "Detection of rate of change of auditory frequency," J. Exp. Pathol. **77**, 535–541.
Sek, A., and Moore, B. C. J. (**1999**). "Discrimination of frequency steps linked by glides of various durations," J. Acoust. Soc. Am. **106**(1), 351–359.
Sek, A., and Moore, B. C. J. (**2000**). "Detection of quasitrapezoidal frequency and amplitude modulation," J. Acoust. Soc. Am. **107**(3), 1598–1604.
Tansley, B. W., and Suffield, J. B. (**1983**). "Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation," J. Acoust. Soc. Am. **74**(3), 765–774.
Tsumura, T., and Nimura, T. (**1972**). "Auditory detection of frequency transition," J. Acoust. Soc. Am. **53**(1), 17–25.
Zwicker, E. (**1970**). "Masking and psychological excitation as a consequence of the ears frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg, Sijthoff, Leiden, Netherlands.

Effect of modulation maskers on the detection of second-order amplitude modulation with and without notched noise

Rosalie M. Uchanski[a)]
Department of Otolaryngology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, Missouri 63110

Brian C. J. Moore and Brian R. Glasberg
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 22 March 2005; revised 15 February 2006; accepted 28 February 2006)

The mechanisms underlying the detection of second-order amplitude modulation (AM) were explored. The detectability of second-order AM (fixed depth for each subject) was measured for first- and second-order modulation rates of 16 and 2 Hz, respectively (slow-rate pair), and 50 and 10 Hz, respectively (fast-rate pair), with no masker, a low-band modulation masker (centered at 2 or 10 Hz), and a high-band modulation masker (centered at 16 or 50 Hz). This was done in the absence and presence of an audio-frequency notched noise centered at the carrier frequency of 4000 Hz. Both modulation maskers were "low-noise" noises, to prevent overmodulation. In the absence of notched noise, both modulation maskers impaired performance for the slow-rate pair, but only the low-band masker impaired performance for the fast-rate pair. When notched noise was present, the low-band masker had no significant effect for either rate pair and the high-band masker had an effect only for the slow-rate pair. These results suggest that second-order AM detection is mediated both by an envelope distortion component at the second-order rate and by slow fluctuations in the output of a modulation filter tuned to the first-order rate. When notched noise is present, the distortion component plays little role. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2188375]

PACS number(s): 43.66.Mk, 43.66.Lj, 43.66.Dc [JHG] Pages: 2937–2946

I. INTRODUCTION

The perception of amplitude modulation (AM) in sounds has often been modeled using an array of bandpass filters simulating the auditory filters, or a single broad filter, each followed by a nonlinearity and a "smoothing" device implemented as a low-pass filter (Viemeister, 1979) or a sliding temporal integrator (Festen amd Plomp, 1981; Moore *et al.*, 1988; Plack and Moore, 1990). An alternative approach is based on the idea that the envelopes of the outputs of the auditory filters are fed to a second array of overlapping bandpass filters tuned to different envelope modulation rates (Kay, 1982; Dau *et al.*, 1997a, 1997b; Ewert and Dau, 2000; Ewert *et al.*, 2002; Verhey *et al.*, 2003). This set of filters is usually called a "modulation filter bank" (MFB). The concept of the MFB implies that the auditory system performs a spectral analysis of the envelope at the output of each auditory filter. Evidence broadly consistent with such a view has come from experiments on "modulation masking," in which listeners are required to detect "probe" modulation in the presence of masker modulation; such experiments show selectivity in the modulation domain analogous to that observed for audio frequencies, but with relatively poor resolution or selectivity (Bacon and Grantham, 1989; Houtgast, 1989; Dau *et al.*, 1997a, 1997b, 1999; Ewert and Dau, 2000; Lorenzi *et al.*, 2001a; Ewert *et al.*, 2002; Sek and Moore, 2003).

A challenge for the MFB approach comes from studies using "second-order" AM (Lorenzi *et al.*, 2001a, 2001b; Verhey *et al.*, 2003). To produce second-order AM, the audio-frequency carrier is first sinusoidally amplitude modulated with a first-order modulator of rate f_m. Then, the depth of modulation of the first-order modulator is itself sinusoidally modulated with a second-order modulator of rate f'_m. The "envelope of the envelope," which corresponds to the second-order modulator, has been called the "venelope" (Ewert *et al.*, 2002; Verhey *et al.*, 2003). Usually, the task of the subject is to detect the second-order modulation, i.e., to distinguish a stimulus with second-order modulation from a stimulus with first-order modulation only. The modulation spectrum of a stimulus with second-order AM does not contain a component at f'_m. For example, if f_m is 50 Hz and f'_m is 10 Hz, the modulation spectrum contains components at 40, 50, and 60 Hz, but not at 10 Hz. However, subjects are able to detect second-order AM (Lorenzi *et al.*, 2001a, 2001b). Furthermore, when second-order AM is presented, subjects report hearing a slow fluctuation at a rate which corresponds to f'_m. The problem for the MFB approach is to explain how second-order AM is detected, since, for a linear MFB, the modulation filter tuned to f'_m would often have a negligible output. Two mechanisms have been suggested. The first is that nonlinearities in the auditory system distort the internal

a)Author to whom correspondence should be addressed. Electronic mail: ruchanski@wustl.edu

representation of the envelope, introducing a component with rate f'_m; detection of the second-order AM might then be mediated by detection of this distortion component (e.g., Moore *et al.*, 1999; Lorenzi *et al.*, 2001a). Various nonlinearities may contribute to the generation of an envelope distortion component, such as basilar-membrane compression (Moore *et al.*, 1999), "thresholding" (Verhey *et al.*, 2003), and adaptation (Shofner *et al.*, 1996). The second possible mechanism is based on the fact that the output of the modulation filter tuned to f_m would show fluctuations at rate f'_m. Subjects may be sensitive to these fluctuations. Thus, relatively slow-rate second-order AM may be detected via temporal fluctuations at the output of a modulation filter tuned to a faster rate (Ewert *et al.*, 2002; Millman *et al.*, 2003; Verhey *et al.*, 2003).

There is reasonably strong evidence supporting the idea that the auditory system introduces a distortion component in the "internal" modulation spectrum at the venelope rate in response to sounds with complex envelopes, and this evidence suggests that the magnitude of the distortion component is small. For example, Moore *et al.* (1999) and Sek and Moore (2004) estimated that, for moderately large modulation depths of the primary components (modulation depth, $m \approx 0.3$ per component), the magnitude of the distortion component was equivalent to that produced by a first-order modulator with $m \approx 0.03$ ($20 \log m = -30$). This corresponds to a barely detectable modulation depth for a sinusoidal carrier (Riesz, 1928; Zwicker, 1952; Sek and Moore, 1994; Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). The magnitude of the venelope component implied by the results of other studies (Ewert *et al.*, 2002; Verhey *et al.*, 2003; Füllgrabe *et al.*, 2005) is about a factor of two to four greater, but still relatively small. It is not clear whether the magnitude of the distortion component is sufficient to account for the detection of second-order AM, especially when the second-order modulation depth is relatively small. However, the distortion component is usually well separated in frequency from the primary modulation components, so it would be masked little, if at all, by the primary components.

To test the role of envelope distortion in the perception of second-order AM, Lorenzi *et al.* (2001b) measured second-order temporal modulation transfer functions (TMTFs) using a 5000-Hz sinusoidal audio-frequency carrier and a 2-Hz-wide narrow-band noise carrier centered at 5000 Hz. The envelope of the noise carrier had inherent slow fluctuations, which were intended to mask envelope distortion components at slow second-order modulation rates. For the sinusoidal carrier, the second-order TMTFs were nearly low pass in shape and the overall sensitivity and cut-off frequency increased when f_m increased from 16 to 256 Hz. For the narrow-band-noise carrier, the second-order TMTFs were nearly flat in shape for $f_m = 16$ and 64 Hz, and they showed a high-pass segment for $f_m = 256$ Hz. For slow second-order modulation rates (i.e., low values of f'_m), thresholds were much poorer (by about 6–8 dB) for the noise carrier than for the sinusoidal carrier. Lorenzi *et al.* (2001b) concluded that a distortion component at the second-order rate probably contributed to the detection of second-order modulation.

As a further test of this idea, Millman *et al.* (2003) measured thresholds for detecting second-order AM of a 5000-Hz sinusoidal carrier in the presence of two types of modulation masker, each of which was a band of noise applied to the same carrier. One noise masker was centered on the second-order modulation rate ($f'_m = 2$ Hz) and the other was centered on the first-order modulation rate ($f_m = 16$ Hz). Increasing the depth of the masker centered on 2 Hz had little effect on detection thresholds for second-order AM (thresholds increased by about 3 dB), but increased detection thresholds for 2-Hz first-order AM by a factor of six (thresholds increased by about 15 dB). Increasing the depth of the masker centered on 16 Hz increased detection thresholds for both first- and second-order AM by a factor of three (about 10 dB). Millman *et al.* concluded that the detection of second-order AM, when f'_m is 2 Hz, is not dependent on the detection of a distortion component at f'_m but is dependent on the detection of fluctuations at the output of modulation filters centered close to f_m. This conclusion is not entirely consistent with that of Lorenzi *et al.* (2001b), who state that detection of spectral energy at the second-order rate contributes to the detection of second-order modulation. However, Millman *et al.* pointed out that the interpretation made by Lorenzi *et al.* was based on the assumption that the inherent amplitude fluctuations in the narrow-band noise carrier would only mask AM at the second-order rate, and not at the first-order rate. This assumption may not have been correct.

One limitation of the study of Millman *et al.* (2003) is that they used relatively small modulation depths. The first-order modulation depth was 0.25 and the second-order depth at threshold ranged from about 0.06 to 0.14. These small modulation depths were used to avoid overmodulation when the noise modulation maskers were added. The magnitude of a putative envelope distortion component at the second-order rate would increase with increasing first- and second-order modulation depths (Moore and Sek, 2000; Sek and Moore, 2003). Thus, for larger first- and second-order modulation depths, a distortion component at the second-order rate might play a greater role in the detection of second-order AM. In the present study, we tested this idea by performing an experiment similar to that of Millman *et al.* (2003), but using larger modulation depths for the first- and second-order modulators. To avoid the problem of overmodulation when the noise modulation maskers were added, we used "low-noise noise" maskers with minimal amplitude fluctuations (Pumplin, 1985; Kohlrausch *et al.*, 1997).

A second limitation of the study of Millman *et al.* (2003) is that the first- and second-order frequencies were both relatively low, 16 and 2 Hz, respectively. It may be that part of the nonlinearity that introduces an envelope distortion component at the second-order rate is time varying. If so, the effective magnitude of the distortion component may vary with first- and second-order modulation rate. In the present study, we extended the work of Millman *et al.* by using first- and second-order modulation rates of 50 and 10 Hz, respectively, in addition to the modulation rates of 16 and 2 Hz used by them.

A factor not considered in most studies of the perception of second-order AM is that the detection of AM of a sinu-

soidal carrier probably depends strongly on the use of information from the high-frequency side of the excitation pattern evoked by the carrier (Zwicker, 1956; Florentine and Buus, 1981; Moore and Sek, 1994; Kohlrausch *et al.*, 2000). This part of the excitation pattern appears to be processed almost linearly on the basilar membrane, at least for medium to high frequencies (Rhode and Robles, 1974; Sellick *et al.*, 1982; Ruggero, 1992; Nelson and Schroder, 1997). If listeners use information mainly from the high-frequency side of the excitation pattern ("off-frequency listening") when detecting second-order modulation, it seems unlikely that envelope distortion produced by basilar membrane compression would play any role. However, the amount of envelope distortion produced at some later stage in the auditory system might be influenced by the amount of amplitude fluctuation at its input. The amount of fluctuation would be greater for an audio-frequency channel centered above the carrier frequency than for a channel centered close to the carrier frequency. Hence, the amount of envelope distortion generated at the later stage might be greater for an "off-frequency" channel than for an on-frequency channel. We return to this point later.

Füllgrabe *et al.* (2005) explored the influence of off-frequency listening by examining the effect of a notched noise centered at the audio-carrier frequency. They measured the detectability of 5-Hz probe modulation of a 5000-Hz carrier in the presence of 5-Hz second-order masker modulation, as a function of the relative phase $\Delta\varphi$ between the probe modulation and the venelope of the masker. In the absence of the notched noise, the value of $\Delta\varphi$ leading to poorest detectability of 5-Hz second-order modulation varied with the first-order modulation rate. If the results are interpreted in terms of an envelope distortion component at the second-order modulation rate, this pattern of results suggests one or more of the following: (1) several nonlinearities contribute to the generation of the distortion component, and the relative contribution of those nonlinearities varies with first-order rate; (2) at least one component of the nonlinearity that generates the distortion component is time varying; (3) some other factor (e.g., the extent of off-frequency listening) is affected by the first-order rate. In the presence of the notched noise, the value of $\Delta\varphi$ giving poorest detectability hardly varied with first-order modulation rate, but it did vary across subjects, from about 45° to 135°. These results suggest that more than one nonlinear mechanism may contribute to the generation of an envelope distortion component. In the present study, the effect of modulation maskers on the detectability of second-order AM was examined both in the presence and absence of a notched noise centered at the audio-carrier frequency; the noise was intended to restrict off-frequency listening. The comparison between the conditions with and without notched noise was intended to allow us to assess the contribution, if any, of basilar-membrane mechanisms to the generation of a distortion component in the modulation domain.

In summary, to explore further the possibility of second-order modulations being detected via a distortion component in the modulation domain, we examined the effect of modulation maskers (low-noise noise maskers), centered at different modulation frequencies, on second-order modulation detection. Also, to assess the contribution of basilar-membrane mechanisms to the generation of a distortion component, the second-order modulation detection experiment was performed both with and without notched noise in the audio-frequency domain. In addition to the experiment on second-order modulation detection, first- and second-order modulation detection thresholds were measured for the same subjects, both with and without notched noise.

II. GENERAL METHOD

A. Subjects

Six subjects with normal hearing were tested. Five of the subjects were recruited from the local community and were paid for their services. One subject was one of the authors, and was unpaid. The ages of the subjects ranged from 20 to 45 years. All had audiometric thresholds less than 20 dB HL (ANSI, 2004) at frequencies between 250 and 8000 Hz. All subjects were similarly inexperienced with modulation detection tasks. Hence, they were given practice runs, for both the first- and second-order modulation detection tasks, until their thresholds were stable, i.e., the standard deviation of the last three thresholds was less than 2.5 dB. This took roughly 9–12 runs per condition, yielding about 9 h total of practice per subject for the two detection tasks.

B. Signal generation

The modulated carrier was digitally generated using a Tucker-Davis Technologies (TDT) System II, controlled by a PC. The sampling rate was 51 200 Hz. The stimuli were presented monaurally over Sennheiser HD580 headphones, at a level of 75 dB SPL (this is the estimated level at the eardrum, taking into account the diffuse-field response of the headphones; the equivalent diffuse-field level was 66 dB SPL). For all modulated stimuli, the carrier was a sinusoid with frequency f_c equal to 4000 Hz.

A noise with a spectral notch at the carrier frequency was used to restrict off-frequency listening and to prevent the detection of distortion components in the audio-frequency domain. This noise, whose characteristics were based on calculation of excitation patterns as described by Glasberg and Moore (1990), consisted of three bands: Band 1 extended from 50 to 1500 Hz and had a spectrum level of 10 dB (re 20 μPa); Band 2 extended from 1600 to 2400 Hz and had a spectrum level of 18 dB; Band 3 extended from 6000 to 13 000 Hz and had a spectrum level of 18 dB (see also, Alcántara *et al.*, 2003). All spectrum levels here are equivalent diffuse-field levels; levels at the eardrum would have been higher, especially for frequencies in the range 1500 to 6000 Hz. Together, these bands produced: (a) negligible excitation at the carrier frequency; (b) relatively high excitation for frequencies near the carrier frequency ($f_c \pm 0.3 f_c$) so as to restrict off-frequency listening; and (c) moderate excitation at frequencies well below the carrier frequency (to mask audio-frequency distortion components at the first- or second-order modulation rate). The noise was digitally generated, by summing sinusoidal components spaced at 0.1-Hz intervals, and the 10-s noise sample was repeated seamlessly and recorded onto a compact disc. For the notched-noise

conditions, this noise was played continuously from a compact-disc player and was added to the modulated signals via a TDT SM3 mixer.

III. EXPERIMENT 1. MODULATION DETECTION WITH AND WITHOUT NOTCHED NOISE

A. Conditions and stimuli

To allow comparisons across first- and second-order modulation conditions, as well as comparisons with data from previous studies, first-order modulation detection thresholds were measured for four first-order modulation rates: f_m=2, 10, 16, and 50 Hz. For these conditions, the subject's task was to detect the presence of sinusoidal amplitude modulation (SAM) applied to the 4000-Hz carrier (the "target"). The "standard" was the unmodulated carrier. Modulation detection thresholds were determined for the four modulation rates in both the absence and presence of notched noise.

To examine second-order modulation perception at both slow and fast rates, second-order modulation detection thresholds were measured using two different pairs of modulation rates: "slow" with f_m=16 Hz and f'_m=2 Hz; "fast" with f_m=50 Hz and f'_m=10 Hz. For these conditions, the subject's task was to detect the presence of sinusoidal modulation applied to the modulation depth of the first-order SAM (the "target"). The "standard," in this case, was the carrier modulated at the first-order modulation rate f_m with a fixed modulation depth m. Second-order modulation detection thresholds were determined in both the absence and presence of the same notched noise as before.

Equations for the first- and second-order modulation "targets" $T(t)$ and $T'(t)$, respectively, are given below

$$T(t) = c\{1 + m\sin(2\pi f_m t + \varphi_m)\}\sin(2\pi f_c t + \varphi_c) \quad (1)$$

$$T'(t) = c\{1 + m[1 + m'_B\sin(2\pi f'_m t + \varphi'_m)] \times \sin(2\pi f_m t + \varphi_m)\}\sin(2\pi f_c t + \varphi_c), \quad (2)$$

where f_m and m are the modulation rate and depth, respectively, of the first-order modulation, and f'_m and m'_B are the second-order modulation rate and depth, respectively, of the second-order modulation. The multiplicative scalar c was used to adjust levels such that all stimuli had a constant overall power. Starting phases, φ_c, φ_m, and φ'_m, were chosen randomly for each trial. For detection of second-order AM, the first-order modulation depth m was 0.5 for both the "target" and the "standard" stimuli. To avoid overmodulation in the second-order modulation conditions, the second-order modulation depth m'_B was constrained to be $\leqslant(1/m)-1$. Stimuli were 1 s in duration with 20-ms raised-cosine on/off ramps.

In recent publications, two different equations have been used to describe second-order modulation signals. Hence, comparisons of second-order threshold values across studies must be made carefully, as the definition of second-order modulation depth differs depending on the equation used. For example, Lorenzi *et al.* (2001a; b) and Füllgrabe *et al.* (2005) describe the second-order "target" signal by the following equation (ignoring initial phases):

$$T'(t) = c\{1 + [m + m'\sin(2\pi f'_m t)] \times \sin(2\pi f_m t)\}\sin(2\pi f_c t), \quad (3)$$

while in this study the second-order modulation "target" signal is described by Eq. (2) above and in Millman *et al.* (2003) an equation of the same form as Eq. (2) is used. In Eq. (2), the second-order modulation depth is defined as m'_B while in Eq. (3), the second-order modulation depth is defined as m' (the first-order modulation depth is m in both equations). Mathematically, the second-order modulation depth m'_B in Eq. (2) is equivalent to m'/m of Lorenzi *et al.* (2001a; b) and Füllgrabe *et al.* (2005).

B. Procedure

First- and second-order modulation detection thresholds were measured using an adaptive three-interval, three-alternative forced-choice procedure, using a three-down one-up stepping rule that estimates the 79.4% correct point on the psychometric function (Levitt, 1971). The "target" signal was presented in one of the three intervals, selected at random, and the stimuli were separated by 400-ms interstimulus intervals. Each stimulus was accompanied by a separate light on the response box; the response buttons were positioned below the lights. Visual feedback, using the lights, was provided after each trial.

For the measurement of first-order AM detection thresholds, the modulation depth m was initially set to 0.4 (−8 dB), while for the measurement of second-order AM detection thresholds, the modulation depth m'_B was initially set to 0.5(−6 dB). For both first- and second-order modulation detection, the initial step size was 5 dB in units of 20 log m or 20 log m'_B. After four reversals the step size was reduced to 2 dB. Twelve reversals were obtained and threshold, for each adaptive run, was defined as the mean of the values of 20 log m or 20 log m'_B at the last eight reversals. For each condition and each subject, modulation detection threshold was estimated from the average of three adaptive runs.

C. Results and discussion

Individual-subject data and group-mean data are shown in Fig. 1 for both first- and second-order modulation detection. The lowest thresholds were obtained for first-order modulation detection in the absence of notched noise (open triangles). There is some variability across subjects and across the range of modulation rates examined (2 to 50 Hz), with first-order modulation thresholds ranging from −19 to −28 dB. These thresholds are higher than typically reported by others for similar carrier frequencies and modulation rates. For example, Lorenzi *et al.* (2001b) reported thresholds of about −33 dB, and Kohlrausch *et al.* (2000) reported thresholds of about −28 dB. Millman *et al.* (2003) reported first-order modulation thresholds of roughly −37 dB for f_m =2 Hz, and roughly −27 dB for f_m=16 Hz, using a 5000-Hz carrier. However, in all these studies the adaptive procedure converged on the modulation depth required for 70.7% correct detection, using a two-interval forced-choice task, which corresponds to a sensitivity index d' of about 0.76. This contrasts with the procedure used in this study which tracks the

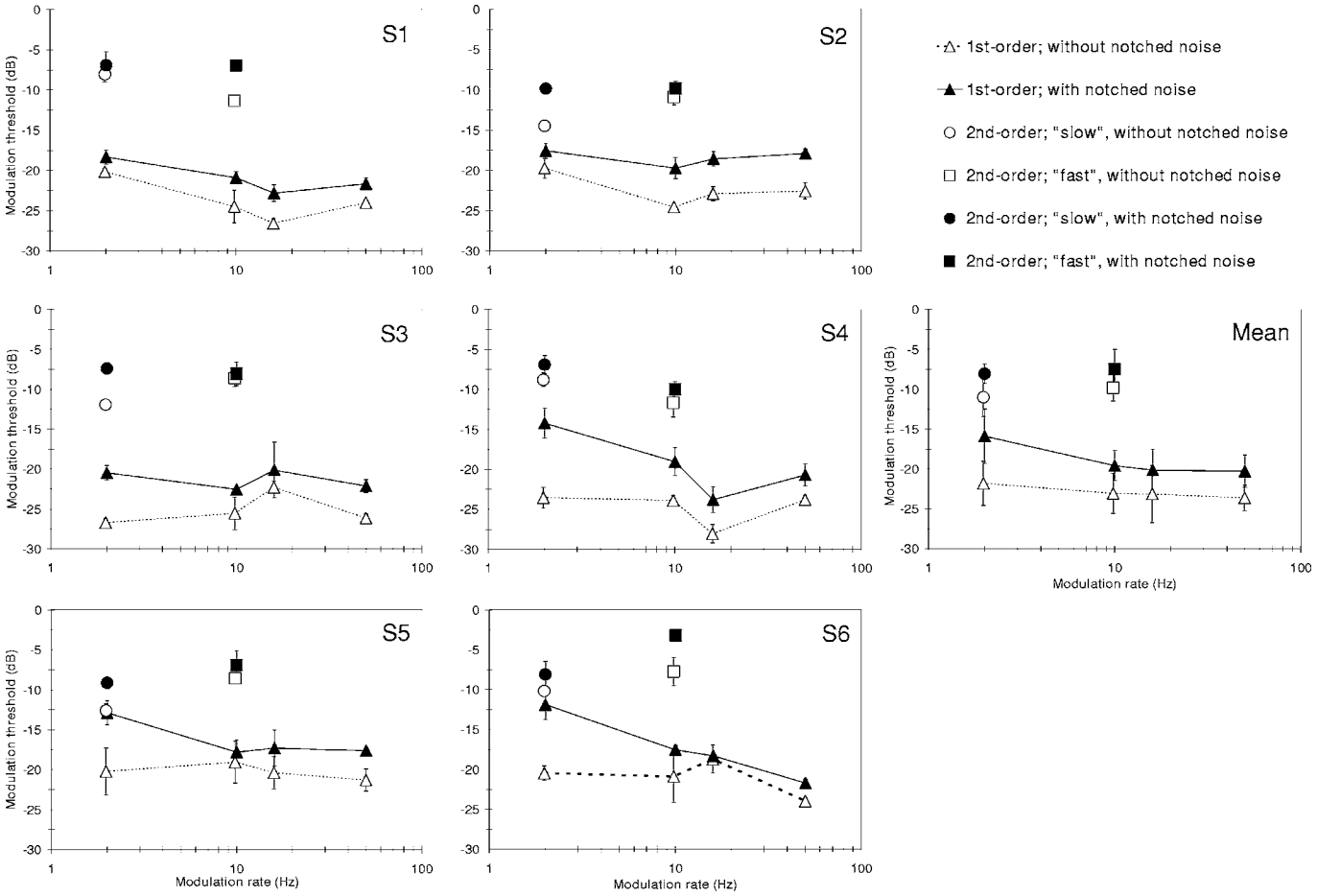


FIG. 1. Individual and mean first- and second-order modulation detection thresholds. Each panel shows the modulation depth at threshold, in decibels, on the ordinate: $20 \log m$ for first-order detection and $20 \log m'_B$ for second-order detection. The abscissa represents the modulation rate: f_m for first-order detection and f'_m for second-order detection. Data for first-order modulation detection are represented by triangles, and are drawn with lines connecting the symbols. Data for second-order modulation detection are represented by circles for the slow-rate pair and by squares for the fast-rate pair. Open and filled symbols represent data from conditions without and with notched noise, respectively. First-order modulation depth m was fixed at 0.5 for all second-order modulation detection conditions. The symbols (triangles, circles, and squares) represent the mean and error bars represent ±1 standard deviation from the mean (note: data from the conditions without notched noise are offset slightly along the abscissa for readability).

79.4% correct point using a three-interval forced-choice task; this corresponds to a d' value of about 1.62. Moore and Glasberg (2001) used the same procedure as here and reported similar thresholds for a 5000-Hz carrier. In addition, the stimuli in the studies of Lorenzi *et al.* (2001b) and Millman *et al.* (2003) were twice as long as the stimuli in this study (2 s versus 1 s), which may have made modulation detection somewhat easier in their studies, especially for the lowest modulation rates.

The presence of the notched noise increased thresholds for first-order modulation detection by about 3–6 dB (compare the filled to the open triangles). Again, there is some variability across subjects and across modulation rates. Threshold increases produced by the notched noise ranged from 2.9 to 5.4 dB across the six subjects, and ranged from 0.4 to 9.3 dB across subjects and modulation rate. Overall, the effect of the notched noise is similar in magnitude to that reported by Strickland and Viemeister (1997) for slow modulation rates (2–50 Hz), although Strickland and Viemeister (1997) used an 800-Hz-wide noise carrier rather than a sinusoidal carrier. Consistent with the results of others (Zwicker, 1956; Füllgrabe *et al.*, 2005), this result indicates that off-frequency listening contributes to first-order modulation detection, for these moderately slow modulation rates.

Individual and mean second-order modulation detection thresholds are also shown in Fig. 1 (see open and filled squares and circles). In quiet, second-order thresholds for the slow-rate pair varied across listeners from −8 to −15 dB (open circles). For the fast-rate pair, thresholds varied from about −8 to −12 dB (open squares). These thresholds are substantially higher than those reported by Lorenzi *et al.* (2001b) and Millman *et al.* (2003) using a carrier frequency of 5000 Hz. For the rates similar to those for our slow-rate pair, Lorenzi *et al.* (2001b), using a first-order depth of 0.5, reported thresholds that correspond to $m'_B = 0.08$ (or −22 dB; Note: $m' = 0.04 \rightarrow m'_B = m'/m = 0.04/0.5 = 0.08$), as compared to the average m'_B value here of about 0.28 (−11 dB). Millman *et al.* also reported m'_B values of about −22 dB, even though the first-order modulation depth was much smaller ($m = 0.25$). Again, the higher thresholds obtained in our study are probably due to the higher d' value tracked by our procedure, and by the shorter stimulus duration used here.

The addition of notched noise elevated thresholds for detecting second-order modulation (filled circles and

squares). On average, thresholds were elevated by 3 dB for the slow-rate pair and by 2.3 dB for the fast-rate pair. Again, this effect varied somewhat across subjects and across rates. For some subjects (S2, S3, S5), the effect of the notched noise was greater for the slow-rate pair than for the fast-rate pair, while for others (S1, S6) the opposite occurred. Overall, the effect of the notched noise suggests that second-order modulation detection also depends to some extent on off-frequency listening in the audio-frequency domain, regardless of the subsequent mechanism for detecting the second-order modulation (generation and detection of a distortion component in the modulation-frequency domain, or detection of temporal fluctuations at the outputs of the modulation filters tuned to the first-order rate).

IV. EXPERIMENT 2. DETECTABILITY OF SECOND-ORDER AM IN THE PRESENCE OF MODULATION MASKERS

In this experiment, the effects of different noise modulation maskers on the ability to detect second-order modulation were measured. A noise modulation masker was centered either at the first- or second-order modulation rate. If the detection of second-order modulation depends on the detection of an envelope distortion component generated by the auditory system at the second-order rate, then a narrow-band noise modulation masker should have a more deleterious effect on second-order modulation detection when it is centered at the second-order rate than when it is centered at the first-order rate. If, on the other hand, second-order modulation detection depends on the detection of fluctuations at the second-order rate in the output of the modulation filter tuned at, or close to, the first-order rate, then a modulation masker centered at the first-order rate should have the more deleterious effect, since this modulation masker would interfere directly with such fluctuations and since the modulation masker centered at the second-order rate would have little effect on such fluctuations (i.e., it would not be passed, or would be significantly attenuated, by the modulation filter tuned to the first-order rate).

A. Conditions and stimuli

For each rate pair, second-order AM-detection performance was measured for three masker conditions: (i) no-masker—no modulation masker; (ii) low-band—modulation masker centered at the second-order modulation rate (f'_m); and (iii) high-band—modulation masker centered at the first-order modulation rate (f_m). Each condition was tested in the absence and presence of the notched noise described in Sec. II B.

As in Experiment 1, the carrier frequency was 4000 Hz for all modulated stimuli. For the no-masker conditions, the "target" stimulus is described by Eq. (2) of Sec. III A. For the low-band and high-band conditions, the equation for the second-order modulated "target" stimulus, $T'_N(t)$, is given below:

$$T'_N(t) = c\{1 + an(t) + m[1 + m'_B \sin(2\pi f'_m t + \varphi'_m)] \times \sin(2\pi f_m t + \varphi_m)\}\sin(2\pi f_c t + \varphi_c), \quad (4)$$

where all variables are the same as in Eq. (2), but with the addition of a modulation masker $n(t)$, which is scaled by the coefficient a. When a was equal to 1, the root-mean-square amplitude of the modulation masker was equal to 1. For the slow-rate pair, the low-band masker was a 2-Hz-wide noise centered at 2 Hz (f'_m) while the high-band masker was an 8-Hz-wide noise centered at 16 Hz (f_m). For the fast-rate pair, the low-band masker was a 5-Hz-wide noise centered at 10 Hz (f'_m) while the high-band masker was a 25-Hz-wide noise centered at 50 Hz (f_m). The bandwidths of the low- and high-band maskers for the slow-rate pair were chosen to match those used by Millman *et al.* (2003). The bandwidths for the fast-rate pair were chosen to be comparable to or less than the bandwidths of the modulation filters at their respective center frequencies; the bandwidth of each modulation filter has been estimated to be approximately equal to its center frequency (Dau *et al.*, 1997a, 1997b; Sek and Moore, 2003). Long samples of these low-band and high-band maskers were generated in advance using an algorithm that generates noise with a very low peak factor, called low-noise noise (Pumplin, 1985; Kohlrausch *et al.*, 1997). The algorithm used here was the one described by Kohlrausch *et al.* Selections from the long samples were chosen randomly for each interval. In all conditions, the first-order modulation depth was fixed at $m=0.5$. The second-order modulation depth m'_B and the masker coefficient a were chosen separately for each subject (see below). As in Experiment 1, stimuli were 1 s in duration with 20-ms raised-cosine on/off ramps.

B. Procedure

The subject's task, as before, was to detect the presence of sinusoidal modulation applied to the modulation depth of the 4000-Hz carrier—but, in this case, in the presence of an additional modulation masker $n(t)$ in the low-band and high-band conditions. The "standard" signal was the 4000-Hz carrier modulated at rate f_m, with a fixed modulation depth ($m=0.5$). For the low-band and high-band conditions, the "standard" signal also had the appropriate modulation masker.

The effects of masker modulation on second-order modulation detection were evaluated using the method of constant stimuli, and a three-interval, three-alternative forced-choice procedure. The method of constant stimuli was chosen because it is easier to prevent overmodulation with this method than with an adaptive procedure. For each subject, the second-order modulation depth was chosen (on the basis of data from a preliminary experiment) so as to give roughly 80%–85% correct in the absence of a modulation masker. Since the second-order modulation depths at threshold found in Experiment 1 were dependent on the rate pair and on the presence or absence of notched noise, the supra threshold values of m'_B used in this experiment also depended on the rate pair and on whether notched noise was present or absent. Table I lists the second-order modulation depth m'_B

TABLE I. Values of second-order modulation depth m'_B and of modulation masker coefficient a used in Experiment 2, for both the low-band and high-band modulation masker conditions. The first-order modulation depth was 0.5 for all conditions and all subjects.

	Slow (f_m=16 Hz, f'_m=2 Hz)				Fast (f_m=50 Hz, f'_m=10 Hz)			
	Without notched noise		With notched noise		Without notched noise		With notched noise	
Subject	m'_B	a	m'_B	a	m'_B	a	m'_B	a
S1	0.36	0.31	0.48	0.26	0.31	0.34	0.47	0.26
S2	0.22	0.35	0.40	0.25	0.26	0.35	0.33	0.33
S3	0.27	0.35	0.45	0.27	0.35	0.30	0.43	0.28
S4	0.40	0.30	0.39	0.30	0.26	0.36	0.37	0.31
S5	0.27	0.36	0.35	0.32	0.37	0.31	0.45	0.27
S6	0.32	0.34	0.38	0.31	0.41	0.29	0.54	0.23
Mean	0.31	0.34	0.41	0.29	0.33	0.33	0.43	0.28

and the modulation masker coefficient a used for each subject for each rate-pair and notched-noise condition. To avoid overmodulation, the masker coefficient a was not allowed to exceed $1-m(m'_B+1)$, but otherwise was as large as possible. The order of conditions was randomized. For each subject and each condition, six blocks of 50 trials each were run, and average percent-correct detection scores across the six blocks are reported.

C. Results and discussion

Individual and group mean results are shown in Fig. 2. Each panel shows percent correct detection scores for one of the four rate-pair × notched-noise conditions. Within each panel detection performance is shown, for each subject and for the mean across subjects, by bars from left to right representing the no-masker, low-band masker, and high-band masker conditions, respectively. The upper-left panel shows scores for the slow-rate pair for the condition without notched noise. The upper-right panel shows scores for the slow-rate pair for the condition with notched noise. The bottom two panels show corresponding data for the fast-rate pair.

Four separate within-subject analyses of variance were performed on the arcsine-transformed data, with factor modulation masker condition (no-masker, low-band, or high-band). For both the slow- and fast-rate pairs, and both with

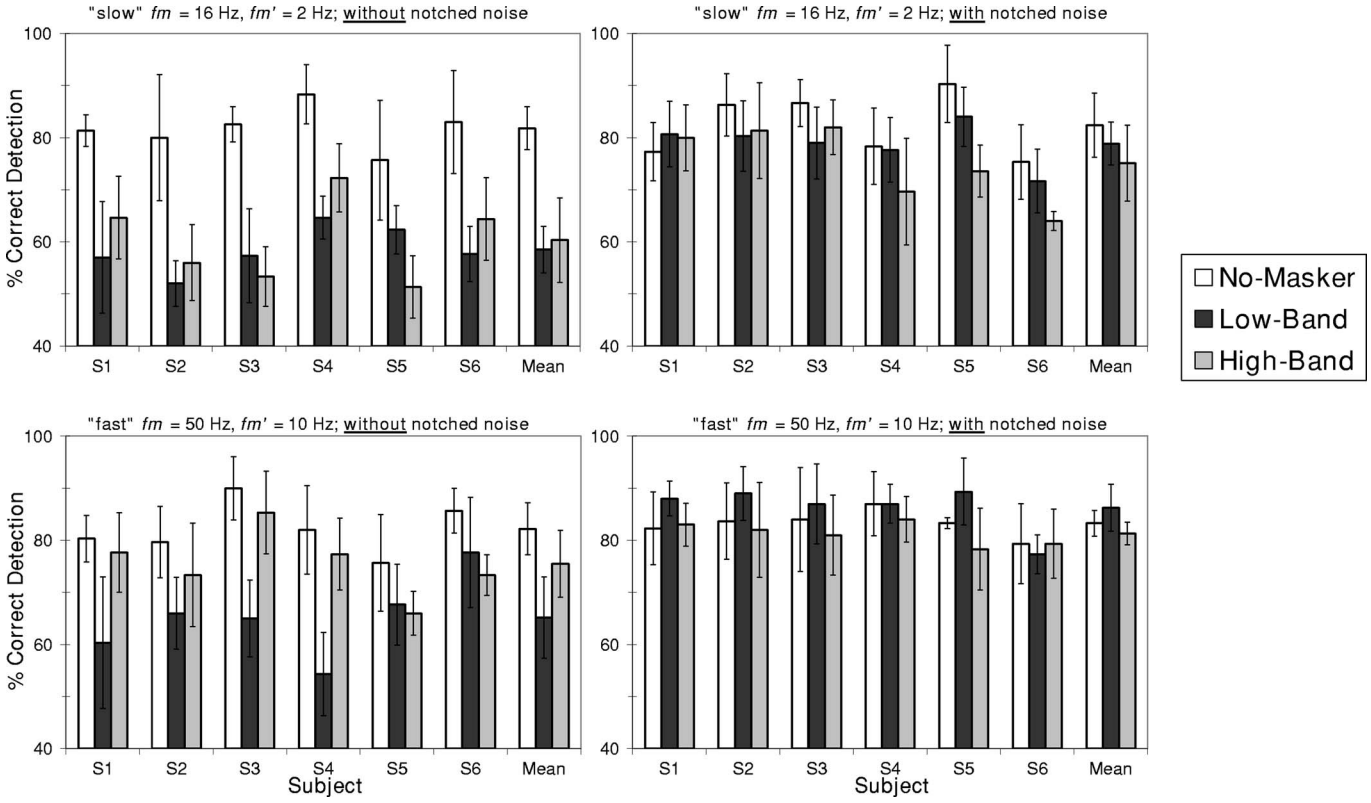


FIG. 2. Individual and mean scores for detection of second-order modulation with and without a notched-noise masker. The top and bottom panels represent data for the slow- and fast-rate pairs, respectively. Left and right panels show data from the conditions without and with notched noise, respectively. Within each panel, the data for each subject are shown as the mean ±1 standard deviation for the conditions with no-masker (unfilled bars), the low-band masker (dark gray filled bars), and the high-band masker (light gray filled bars), from left to right. First-order modulation depth was 0.5 for all subjects and conditions. Other parameters are specified in Table I.

and without notched noise, modulation masker condition had a significant effect on second-order modulation detection performance [$F(2,10) > 5.33, p \leq 0.027$]. The significance of the differences between specific pairs of maskers was determined using Fisher's least-significant differences (LSD) test.

For the slow-rate pair without notched noise (upper-left panel), there was a significant difference between the no-masker condition and the low-band ($p < 0.001$) and high-band ($p < 0.001$) conditions. Thus, both the low- and high-band maskers degraded performance, and by a similar amount (there was no significant difference between scores for the low- and high-band conditions). For the slow-rate pair with notched noise (upper-right panel), scores for the no-masker and low-band conditions did not differ significantly and scores for the low-band and high-band conditions did not differ significantly. However, the mean score was significantly lower for the high-band condition than for the no-masker condition ($p < 0.01$). Thus, the high-band masker degraded performance.

For the fast-rate pair without notched noise (bottom-left panel), there was no significant difference between the no-masker condition and the high-band condition. However, there was a significant difference between the no-masker condition and the low-band condition ($p < 0.01$) and between the low-band and high-band conditions ($p < 0.02$). Thus, the low-band masker significantly degraded performance, while the high-band masker did not. For the fast-rate pair with notched noise (bottom-right panel), there was no significant difference between the no-masker condition and the low-band condition and no significant difference between the no-masker condition and the high-band condition. However, there was a significant difference between the low-band and high-band conditions ($p < 0.01$).

In summary, in the absence of notched noise, both modulation maskers tended to degrade performance relative to the no-masker condition, although the effect of the high-band masker was not significant for the fast-rate pair. The low-band masker had a significant deleterious effect for both rate pairs. In the presence of notched noise, the effects of the modulation maskers differed. For the slow-rate pair, only the high-band masker had a significant deleterious effect on performance, while for the fast-rate pair, neither the low- nor high-band masker significantly affected performance compared to that for the no-masker condition. However, there was some individual variability in the effects obtained when notched-noise was present. For example, for the slow-rate pair, three subjects (S4, S5, S6) showed a larger negative effect for the high- than for the low-band masker, while the other three subjects (S1, S2, S3) showed similar effects for the two maskers.

Our results obtained in the absence of notched noise differ from those reported by Millman *et al.* (2003) for similar stimuli. Using modulation rates the same as for our slow-rate pair, they found that a modulation masker centered at 2 Hz had only a small negative effect on the detectability of second-order AM, while a modulation masker centered at 16 Hz had an effect that was 2–3 times greater for the larger masker modulation depths. For this slow-rate pair, we found a substantial effect of both the low-band and the high-band modulation maskers. The difference may have occurred because of the small modulation depths used by Millman *et al.* It is likely that, when the first- and second-order modulation depths are small, the magnitude of the distortion component at the second-order rate is very small, and therefore that component hardly contributes to the detection of second-order modulation. Our results suggest that, when the first- and second-order modulation depths are larger, and in the absence of a notched noise that restricts off-frequency listening, the detection of second-order modulation is partly mediated by a distortion component at the second-order rate in the internal representation of the envelope. It is not known why the effect of the high-band masker differs for the fast-rate pair (no significant effect) and for the slow-rate pair (significant deleterious effect).

As described in the Introduction, the effective magnitude of the distortion component at the second-order rate is likely to be small, even when the modulation depths of the first- and second-order modulators are large. As a result, even a moderate level of a modulation masker centered at the second-order rate should be sufficient to mask completely the distortion component. The fact that the low-band modulation masker did not abolish the ability to detect second-order modulation, but only partially impaired that ability, suggests that second-order modulation detection can be mediated via a mechanism other than the detection of a distortion component at the second-order rate. An alternative, as mentioned earlier and as proposed by Ewert *et al.* (2002), is that second-order modulation can be detected using fluctuations at the second-order rate at the output of the modulation filter(s) tuned at, or close to, the first-order rate. This alternative was supported by a recent experiment of Füllgrabe and Lorenzi (2005), which involved matching the perceived "beat" rate in a complex modulator with inharmonic components.

Generally, the effects of the modulation maskers were smaller when notched noise was present than when it was absent. This may be partly attributed to the fact that the magnitude of the modulation noise, determined by coefficient a, was somewhat smaller when the notched noise was present than when it was absent (see Table I). For example, for the slow-rate pair, the mean value of a was 0.34 when the notched noise was absent, and 0.29 when the notched noise was present. Also, the depth of the second-order modulator m'_B needed to achieve about 85% correct in the no-masker condition was usually somewhat greater when the notched noise was present (mean m'_B of 0.41) than when it was absent (mean m'_B of 0.31). However, not all of the individual results can be explained in this way. For example, for the slow-rate pair, the values of m'_B and a for subject S4 were almost the same for the conditions with and without notched noise, but for S4 the low-band modulation masker had a large effect on the detectability of second-order AM when the notched noise was absent, and had no effect when the notched noise was present.

In the absence of notched noise, subjects could use information from the high-frequency side of the excitation pattern of the carrier, where the effects of basilar-membrane compression are small or absent, and where the effective

amount of modulation is greater than around the central part of the pattern (Zwicker, 1956; Florentine and Buus, 1981; Moore and Sek, 1994; Kohlrausch *et al.*, 2000). As described in the Introduction, there is evidence that such off-frequency listening is beneficial for the detection of first-order modulation, and the results of Experiment 1 indicate that it is also beneficial for the detection of second-order modulation, in agreement with the results of Füllgrabe *et al.* (2005). Moore *et al.* (1999) suggested that basilar-membrane compression could be partly responsible for the introduction of a distortion component at the second-order rate. However, later work (Verhey *et al.*, 2003; Sek and Moore, 2004; Füllgrabe *et al.*, 2005) showed that the phase of the distortion component was not what would be predicted from basilar-membrane compression. In addition, recent results indicate that listeners who presumably experience little or no basilar membrane compression, such as those with impaired hearing or cochlear implants, have second-order AM-detection thresholds that are similar to or only slightly higher than those for normal-hearing listeners (Tandetnik *et al.*, 2001; Füllgrabe *et al.*, 2003; Lorenzi *et al.*, 2004). These results suggest that there is a nonlinearity other than basilar-membrane compression which is the primary generator of the distortion component at the second-order rate. Whatever this nonlinearity is, it seems plausible that: (a) it occurs later in the auditory system than basilar-membrane compression, and (b) the effective magnitude of the distortion component increases as the amount of amplitude fluctuation on the basilar membrane increases.

This line of reasoning leads to the conclusion that the effective magnitude of the distortion component at the second-order rate may be markedly reduced when off-frequency listening is restricted by notched noise, simply because of the reduction in the effective amount of modulation at the input to the nonlinearity that generates the distortion component. This could account for our finding that the low-band modulation masker had no significant effect when notched noise was present. However, for some subjects, the notched noise had only a small effect on second-order modulation detection thresholds (e.g., for S1, S4, and S6, the presence of notched noise resulted in threshold increases for the slow-rate pair of 1.1, 1.9, and 2.1 dB, respectively; see Fig. 1). For some of these subjects, the lack of an effect due to the low-band modulation masker in the presence of notched noise may be a result of the different values of second-order modulation depth m_B' and modulation masker scalar a in the absence and presence of notched noise.

For the slow-rate pair, the high-band modulation masker impaired the detection of second-order modulation when notched noise was present, but the effect was smaller than when the notched noise was absent. The smaller effect in the presence of notched noise can probably be explained in terms of the relative magnitude of m_B' and a. When the notched noise was absent, m_B' and a were typically about equal in magnitude, although the relative values differed from one subject to another. When the notched noise was present, m_B' was typically larger than a, so one would expect the effect of the high-band modulation masker to be smaller.

According to the argument given above, second-order detection by normally hearing subjects may be improved by off-frequency listening (when no notched noise is present) because the lack of basilar-membrane compression on the high-frequency side of the excitation pattern leads to a greater amount of fluctuation at the input to the nonlinearity generating the envelope distortion product, and hence to a larger magnitude of the distortion product. If this argument is correct, then one might expect that subjects with reduced basilar-membrane compression (those with cochlear hearing impairment or cochlear implants) would show comparable second-order AM detection thresholds to those for normally hearing subjects. As noted earlier, this has been found to be the case (Tandetnik *et al.*, 2001; Füllgrabe *et al.*, 2003; Lorenzi *et al.*, 2004).

V. CONCLUSIONS

(1) Detection thresholds for both first-order and second-order amplitude modulation were increased by the addition of noise with a spectral notch at the carrier frequency, to restrict off-frequency listening. This effect may occur because, in the absence of notched noise, subjects use information from the high-frequency side of the excitation pattern, where the effective amount of modulation is greater because of more linear (less compressive) basilar-membrane processing.

(2) In the absence of notched noise, the detectability of second-order modulation was reduced both by the addition of a modulation masker centered at the first-order rate (high-band masker), and by the addition of a modulation masker centered at the second-order rate (low-band masker), although the deleterious effect of the high-band masker for the fast-rate pair was not significant; for the fast-rate pair (f_m=50 Hz, f_m'=10 Hz), the effect of the low-band masker was greater than the effect of the high-band masker, although performance remained above chance in the presence of the low-band masker. These results suggest that the detectability of second-order modulation when off-frequency listening is not restricted can be mediated by two mechanisms: (a) detection of a distortion component at the second-order rate in the internal representation of the envelope and (b) detection of fluctuations at the second-order rate at the output of modulation filters tuned at, or close to, the first-order rate.

(3) In the presence of notched noise, the detectability of second-order modulation was not significantly affected by the addition of the low-band modulation masker. This suggests that, when off-frequency listening is restricted, detection of second-order modulation is not mediated by detection of a distortion component at the second-order rate, possibly because the amplitude of this component is too small.

(4) In the presence of notched noise, the detectability of second-order modulation was impaired by the addition of the high-band modulation masker for the slow-rate pair but not for the fast-rate pair. At least for slower modulation rates, this result is consistent with the idea that detection of second-order modulation can be medi-

ated via detection of fluctuations at the second-order rate at the output of modulation filters tuned at, or close to, the first-order rate.

ACKNOWLEDGMENTS

The authors thank our subjects for participating in these experiments. The first author thanks all members of the Auditory Perception Group at the University of Cambridge for the intellectual and incredibly congenial work environment. The authors also thank Christian Füllgrabe and an anonymous reviewer for helpful comments on this manuscript. This work was funded by the MRC (UK).

Alcántara, J. I., Moore, B. C. J., Glasberg, B. R., Wilkinson, A. J. K., and Jorasz, U. (**2003**). "Phase effects in masking: Within- versus across-channel processes," J. Acoust. Soc. Am. **114**, 2158–2166.

American National Standards Institute (**2004**). "Specification for Audiometers (ANSI S3.6-2004)," New York: ANSI.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," J. Acoust. Soc. Am. **102**, 2906–2919.

Dau, T., Verhey, J. L., and Kohlrausch, A. (**1999**). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," J. Acoust. Soc. Am. **106**, 2752–2760.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Ewert, S. D., Verhey, J. L., and Dau, T. (**2002**). "Spectro-temporal processing in the envelope-frequency domain," J. Acoust. Soc. Am. **112**, 2921–2931.

Festen, J. M., and Plomp, R. (**1981**). "Relations between auditory functions in normal hearing," J. Acoust. Soc. Am. **70**, 356–369.

Florentine, M., and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. **70**, 1646–1654.

Füllgrabe, C., and Lorenzi, C. (**2005**). "Perception of the envelope beat-frequency of inharmonic complex temporal envelopes," J. Acoust. Soc. Am. **118**, 3757–3765.

Füllgrabe, C., Meyer, B., and Lorenzi, C. (**2003**). "Effect of cochlear damage on the detection of complex temporal envelopes," Hear. Res. **178**, 35–43.

Füllgrabe, C., Moore, B. C. J., Demany, L., Ewert, S. D., Sheft, S., and Lorenzi, C. (**2005**). "Modulation masking produced by second-order modulators," J. Acoust. Soc. Am. **117**, 2158–2168.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**, 1676–1680.

Kay, R. H. (**1982**). "Hearing of modulation in sounds," Physiol. Rev. **62**, 894–975.

Kohlrausch, A., Fassel, R., and Dau, T. (**2000**). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," J. Acoust. Soc. Am. **108**, 723–734.

Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., and Püschel, D. (**1997**). "Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations," Acta Acust. **83**, 659–669.

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

Lorenzi, C., Soares, C., and Vonner, T. (**2001a**). "Second-order temporal modulation transfer functions," J. Acoust. Soc. Am. **110**, 1030–1038.

Lorenzi, C., Sibellas, J., Füllgrabe, C., Gallégo, S., Fugain, C., and Meyer, B. (**2004**). "Effects of amplitude compression on first- and second-order modulation detection threshold in cochlear implant listeners," Int. J. Audiol. **43**, 264–270.

Lorenzi, C., Simpson, M. I., Millman, R. E., Griffiths, T. D., Woods, W. P., Rees, A., and Green, G. G. R. (**2001b**). "Second-order modulation detection thresholds for pure-tone and narrow-band noise carriers," J. Acoust. Soc. Am. **110**, 2470–2478.

Millman, R. E., Green, G. G. R., Lorenzi, C., and Rees, A. (**2003**). "Effect of a noise modulation masker on the detection of second-order amplitude modulation," Hear. Res. **178**, 1–11.

Moore, B. C. J., and Glasberg, B. R. (**2001**). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **110**, 1067–1073.

Moore, B. C. J., and Sek, A. (**1994**). "Effects of carrier frequency and background noise on the detection of mixed modulation," J. Acoust. Soc. Am. **96**, 741–751.

Moore, B. C. J., and Sek, A. (**2000**). "Effects of relative phase and frequency spacing on the detection of three-component amplitude modulation," J. Acoust. Soc. Am. **108**, 2337–2344.

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.

Moore, B. C. J., Sek, A., and Glasberg, B. R. (**1999**). "Modulation masking produced by beating modulators," J. Acoust. Soc. Am. **106**, 908–918.

Nelson, D. A., and Schroder, A. C. (**1997**). "Linearized response growth inferred from growth-of-masking slopes in ears with cochlear hearing loss," J. Acoust. Soc. Am. **101**, 2186–2201.

Plack, C. J., and Moore, B. C. J. (**1990**). "Temporal window shape as a function of frequency and level," J. Acoust. Soc. Am. **87**, 2178–2187.

Pumplin, J. (**1985**). "Low-noise noise," J. Acoust. Soc. Am. **78**, 100–104.

Rhode, W. S., and Robles, L. (**1974**). "Evidence from Mössbauer experiments for nonlinear vibration in the cochlea," J. Acoust. Soc. Am. **55**, 588–596.

Riesz, R. R. (**1928**). "Differential intensity sensitivity of the ear for pure tones," Phys. Rev. **31**, 867–875.

Ruggero, M. A. (**1992**). "Responses to sound of the basilar membrane of the mammalian cochlea," Curr. Opin. Neurobiol. **2**, 449–456.

Sek, A., and Moore, B. C. J. (**1994**). "The critical modulation frequency and its relationship to auditory filtering at low frequencies," J. Acoust. Soc. Am. **95**, 2606–2615.

Sek, A., and Moore, B. C. J. (**2003**). "Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators," J. Acoust. Soc. Am. **113**, 2801–2811.

Sek, A., and Moore, B. C. J. (**2004**). "Estimation of the level and phase of the simple distortion tone in the modulation domain," J. Acoust. Soc. Am. **116**, 3031–3037.

Sellick, P. M., Patuzzi, R., and Johnstone, B. M. (**1982**). "Measurement of basilar membrane motion in the guinea pig using the Mössbauer technique," J. Acoust. Soc. Am. **72**, 131–141.

Shofner, W. P., Sheft, S., and Guzman, S. J. (**1996**). "Responses of ventral cochlear nucleus units in the chinchilla to amplitude modulation by low-frequency, two-tone complexes," J. Acoust. Soc. Am. **99**, 3592–3605.

Strickland, E. A., and Viemeister, N. F. (**1997**). "The effects of frequency region and bandwidth on the temporal modulation transfer function," J. Acoust. Soc. Am. **102**, 1799–1810.

Tandetnik, S., Garnier, S., and Lorenzi, C. (**2001**). "Measurement of first- and second-order modulation detection thresholds in listeners with cochlear hearing loss," Br. J. Audiol. **35**, 355–364.

Verhey, J. L., Ewert, S. D., and Dau, T. (**2003**). "Modulation masking produced by complex tone modulators," J. Acoust. Soc. Am. **114**, 2135–2146.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based on modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Zwicker, E. (**1952**). "Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones (The limits of audibility of amplitude modulation and frequency modulation of a pure tone)," Acustica **2**, 125–133.

Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs (The foundations for determining the information capacity of the auditory system)," Acustica **6**, 356–381.

Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation[a)]

Ville Pekka Sivonen[b)]
Sound Quality Research Unit (SQRU), Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, 9220 Aalborg East, Denmark and Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, 2850 Nærum, Denmark

Wolfgang Ellermeier
Sound Quality Research Unit (SQRU), Department of Acoustics, Aalborg University, Fredrik Bajers Vej 7 B5, 9220 Aalborg East, Denmark

(Received 25 July 2005; revised 7 February 2006; accepted 13 February 2006)

The effect of sound incidence angle on loudness was investigated using real sound sources positioned in an anechoic chamber. Eight normal-hearing listeners produced loudness matches between a frontal reference location and seven sources placed at other directions, both in the horizontal and median planes. Matches were obtained via a two-interval, adaptive forced-choice (2AFC) procedure for three center frequencies (0.4, 1, and 5 kHz) and two overall levels (45 and 65 dB SPL). The results showed that loudness is not constant over sound incidence angles, with directional sensitivity varying over a range of up to 10 dB, exhibiting considerable frequency dependence, but only minor effects of overall level. The pattern of results varied substantially between subjects, but was largely accounted for by variations in individual head-related transfer functions. Modeling of binaural loudness based on the at-ear signals favored a sound-power summation model, according to which the maximum binaural gain is only 3 dB, over competing models based on larger gains, or on the summation of monaural loudness indices. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184268]

PACS number(s): 43.66.Pn, 43.66.Cb [AK] Pages: 2965–2980

I. INTRODUCTION

There is growing awareness in psychoacoustics that, for a thorough understanding of loudness perception, its binaural nature has to be taken into account. That is true for basic research, namely the construction of general loudness models (e.g., Moore *et al.*, 1997), as well as for applications to audio reproduction systems (Zacharov *et al.*, 2001) or to perceived sound quality (Bodden, 1997). Especially for instrumental loudness predictions based on Zwicker's modeling, as standardized in ISO 532 (1975), the fact that it is essentially monophonic has been regarded as a major drawback. Nevertheless, the adjustments recently made to loudness modeling rest on a fairly narrow empirical data base, which the present study hopes to extend.

To clarify the issues, it may be helpful to distinguish two stages of processing involved when the loudness of a real sound source in space is perceived: (1) the physical transformation of the "distal" stimulus emitted by the source to "proximal" stimuli arriving at the listener's ears, and (2) the neural, psychological, and cognitive process of integrating the two at-ear stimuli into a single percept.

A. Physical (HRTF) filtering

The first stage can be described in purely acoustical terms, namely by applying head-related transfer functions (HRTFs, Shaw, 1974; Wightman and Kistler, 1989a; Møller *et al.*, 1995; Blauert, 1997, Chap. 5). These account for the filtering of the source due to the physical effects of the human torso, head, and pinnae, depending on the incidence angle of the sound. Further along, through the ear canal, the physical sound transmission has been shown to be independent of the direction of the sound source (see, e.g., Hammershøi and Møller, 1996). Thus, the direction-dependent part of an HRTF can be measured at the entrance to the blocked ear canal, and described by (adopted from Møller *et al.*, 1995)

$$\mathrm{HRTF}_{\text{dir-dep}}(\phi,\theta)=\frac{P_2}{P_1}(\phi,\theta), \tag{1}$$

where ϕ is azimuth, θ is elevation, P_1 is sound pressure at the center position of head, and P_2 is sound pressure at the entrance to the blocked ear canal.

In the median plane, the HRTFs of the two ears are fairly similar due to the physical symmetry of the human body in this plane. However, level differences between HRTFs for different directions can approach 10 dB or more over a fairly wide frequency range. By contrast, large interaural time and level differences (ITDs and ILDs, respectively) between the two ears emerge in the horizontal plane, where the ILDs can reach up to 30 dB at high frequencies. HRTFs have been a major research topic during the past

[a)]Portions of the data were presented at the 147th Meeting of the Acoustical Society of America, New York, May, 2004, and at Internoise, Prague, August, 2004.
[b)]Electronic mail: vps@acoustics.aau.dk

15 years, but the focus of this research has been on adequate "auralization" or sound localization (Wightman and Kistler, 1989a,b; Bronkhorst, 1995; Møller *et al.*, 1996), not on loudness.

B. Binaural loudness summation

The second stage of processing has been termed binaural loudness summation. It describes how the acoustic inputs to the left and right ear are integrated to yield a single, binaural loudness. Starting from the observation that a sound appears louder when listened to with both ears (i.e., binaurally) than with only one (i.e., monaurally), a number of investigators conducted experiments using headphones, through which different combinations of left- and right-ear sound-pressure levels were presented in order to quantify this effect. The results are often summarized as providing evidence for a binaural-to-monaural loudness ratio of 2:1, or perfect loudness summation, corresponding to a binaural gain of approximately 10 decibels (e.g., Levelt *et al.*, 1972; Marks, 1978; Schneider and Cohen, 1997), in accordance with the sone scale of loudness. The evidence is far from unequivocal, however, with many studies finding less-than-perfect summation (e.g., loudness ratios of approximately 1.5:1; Zwicker and Zwicker, 1991), and a level dependence of the binaural gain, which appears to increase from approximately 3 dB near threshold to 6–10 dB at high sound-pressure levels (Shaw *et al.*, 1947; Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963).

Interestingly, binaural loudness summation, as conceptualized in this paradigm, has not been investigated with sounds that are likely to reach the eardrums when emitted from a real source in space, i.e., with products of the first (HRTF) filtering stage. Rather, artificial sounds such as tones, or broadband noise, lacking all spatial or directional information have been used, often at interaural level differences (e.g., in monotic-to-diotic comparisons) far exceeding what would naturally occur. Such conditions of stimulation do not yield an externalized sound image, but rather more or less lateralized inside-the-head percepts. Generally, it appears that the considerable literature on binaural loudness summation has contributed more to the development of scaling methodologies than to the auditory issues involved.

C. Loudness of free and diffuse sound fields

For practical purposes, in an attempt to relate the monophonic measurement of a sound field to perceived loudness, two specific types of sound fields have been considered: The free field, where the sound incidence angle is frontal to the listener, and the diffuse field, where the sound is reaching the listener's ears with equal intensity from all directions.

In order to account for the fundamental difference in sound incidence, the standardized loudness model (ISO 532, 1975) has different computation procedures for the two sound fields. The two procedures are based on both objective and subjective data (Kuhl and Westphal, 1959; Robinson *et al.*, 1961; ISO 389-7, 1996): The objective data represent the differences in the at-ear sound pressures between the two sound fields, i.e., investigating only the effect of the first (HRTF) filtering stage; the subjective data represent the differences in perceived loudness, including effects of both the first and the second stage. Even though the agreement between the objective and subjective data is fair, these investigations do not specify how the two signals at the ears of a listener are summed into a single loudness percept, due to the fact that the stimulation of the auditory system in both sound fields is essentially diotic.

The increasing use of dummy heads for acoustical recordings and measurements, often resulting in dichotic at-ear signals, has led to growing interest in how dichotic at-ear signals should be summed to correspond to the diotic stimulation of the conventional free- and diffuse-field loudness paradigms.

D. Directional loudness

Thus, while studies of HRTF filtering have not explicitly been concerned with the loudness of dichotic sounds, the work on binaural loudness summation appears to lack ecological validity to predict the perception of real sources positioned in space. What remains, then, is less than a handful of studies that have actually investigated directional loudness of real sources in space, taking into account both stages delineated: the physical filtering due to HRTFs, and the ensuing "psychological" summation.

Sivian and White (1933) investigated the effect of direction on hearing thresholds, reporting that at absolute threshold, the binaural minimum audible field is not significantly different from the monaural one. This implies no or a very small binaural advantage, the ear receiving the higher sound pressure determining the binaural hearing threshold. While the directional HRTF effects are the same at higher sound-pressure levels, extrapolating from a detection task to suprathreshold binaural loudness and to its summation across the two ears may be unjustified.

By far, the most pertinent and complete study investigating directional loudness was published by Robinson and Whittle (1960) more than 45 years ago. The authors used a circular array of 12 equally spaced loudspeakers positioned around the listener seated in an anechoic room to obtain loudness matches between a reference and each test position. Using narrow-band sounds having six center frequencies between 1.6 and 10 kHz, and rotating their apparatus when required, they investigated the horizontal, median, and frontal planes in a sample of 16 to 20 listeners. Using probe-tube microphones they also measured sound-pressure levels at the ears of their subjects, as produced by the same stimuli, thus obtaining crude magnitudes of "HRTFs" for the six test frequencies.

As expected, the average loudness matches showed a strong frequency dependence, with the greatest directional effects (of up to 15 dB; see their Fig. 2) observed at higher frequencies (4–10 kHz). Relating the mean loudness matches to the average at-ear sound-pressure measurements, Robinson and Whittle (1960; see their Fig. 5) found the former to be reasonably well predicted by a "6-dB summation rule,"

$$L_{\text{mon}} = 6 \times \log_2(2^{L_{\text{left}}/6} + 2^{L_{\text{right}}/6}), \quad (2)$$

where L_{mon} is the equivalent sound pressure needed for monotic stimulation to match any binaural (diotic: $L_{\text{left}} = L_{\text{right}}$, or dichotic: $L_{\text{left}} \neq L_{\text{right}}$) combination of left-ear (L_{left}) and right-ear (L_{right}) input levels. If, for example, both ears are exposed to 70 dB SPL, the equivalent monotic SPL turns out to be 76 dB SPL (i.e., a 6-dB binaural gain). Note, however, that there is not sufficient information on the fit of this heuristic (other than what can be judged from visual inspection of their Fig. 5) in Robinson and Whittle's report, or on its feasibility to predict individual subjects' data.

Two more recent studies picked up the issue, though using considerably fewer experimental conditions and subjects, and not including HRTF measurements. Both studies (Remmers and Prante, 1991; Jørgensen, 2002) used wideband noise to obtain loudness matches, thus potentially obscuring a frequency dependence, and obtained much smaller effects than Robinson and Whittle (1960), with directional loudness matches varying by less than 3 dB across incidence angles.

E. Rationale for the present study

It thus appears worthwhile to take up the issue of directional loudness once more. This will be done paying special attention to five methodological issues, which are elaborated in turn:

(1) Well-defined narrow-band stimuli are needed to investigate the effects of HRTFs and binaural loudness summation. Note that, in Robinson and Whittle's (1960) report, the sounds used were not sufficiently specified beyond stating that they were "below a critical band" (p. 75), and the later studies used wideband noise which might wash out some of the directional effects.
(2) Given the evidence from earlier headphone experiments showing the binaural gain to increase from approximately 3 dB near threshold to up to 10 dB at high sound-pressure levels, level effects will be taken into account by making measurements at two overall sound-pressure levels.
(3) With the exception of Jørgensen's (2002) study, classical "method(s) of adjustment" have been used to collect the subjective data. By their transparency, and the explicit control they give listeners over the outcome, these methods are prone to subject-induced biases, such as "correcting" an adjustment due to some expectation. Forced-choice procedures (Levitt, 1971; Jesteadt, 1980), especially when interleaving adaptive tracks for different experimental conditions, are much less susceptible to such biases.
(4) Advances in the methodology to HRTFs will be brought to the study of directional loudness. Note that Robinson and Whittle's (1960) pioneering study was done before the term HRTF was coined, and that their at-ear measurements of the stimuli actually used merely provide six points along the frequency scale, and thus do not constitute HRTFs as we conceive of them today.
(5) Since HRTF filtering is known to be highly idiosyncratic, it is likely that with averaged data frequency-dependent directional effects might partially cancel each other, thus underestimating the true effect size. Therefore, a greater emphasis than in earlier studies will be on individual results and analyses.

To sum up, the present investigation will be conducted by having subjects assess loudness in a directional sound field in an anechoic room, and by relating the listening test data both to the distal stimulus given by the sound-pressure level emitted by the active loudspeaker, and to the proximal stimuli given by the participants' at-ear exposure levels as obtained via state-of-the-art HRTF measurements.

II. METHOD

A. Subjects

Eight normal-hearing listeners (between the age of 22 and 46 years; five male, three female), including the second author, participated in the experiment. The subjects' hearing thresholds were determined using standard pure-tone audiometry in the frequency range between 0.25 and 8 kHz with the requirement that none of the thresholds exceed 15 dB hearing level *re*: ISO 389-1 (1998). The five subjects who were not staff members were paid an hourly wage for their participation.

B. Apparatus

1. Loudspeaker setup in the anechoic chamber

The experiment was carried out in an anechoic chamber, which is anechoic above approximately 200 Hz, and has background noise at an inaudible level.

The loudspeaker setup for the experiment consisted of eight identical speakers (Vifa M10MD-39) mounted in hard plastic balls with a diameter of 15.5 cm. A typical frequency response of the loudspeaker can be found in Møller *et al.* (1995).

The loudspeakers were positioned both in the horizontal and median planes. In the horizontal plane, the incidence angles were 30°, 60°, 90°, and 135° of azimuth, and in the median plane the angles were 45° and 90° of elevation. Loudspeakers were also placed ahead and behind the listening position (at 0° and 180° of azimuth with 0° of elevation), where the horizontal and the median planes coincide. Due to assumed symmetry, the loudspeakers were placed only on the left-hand side in the horizontal plane. The distances from the diaphragms of the loudspeakers to the listening position at the center of the setup were 206±4 cm.

The subjects were seated in a chair, the height of which could be adjusted. The chair had a small headrest to restrict head movements of the subjects during the experiment. The subjects' heads and ears were carefully aligned with the center position of the setup by making adjustments to chair height and headrest position using a laser and two video cameras. A photograph of the setup in the anechoic chamber is shown in Fig. 1. The loudspeakers ahead, at 30° and 60° in the horizontal plane, and at 45° and 90° in the median plane are visible in the photograph. The structure suspending the

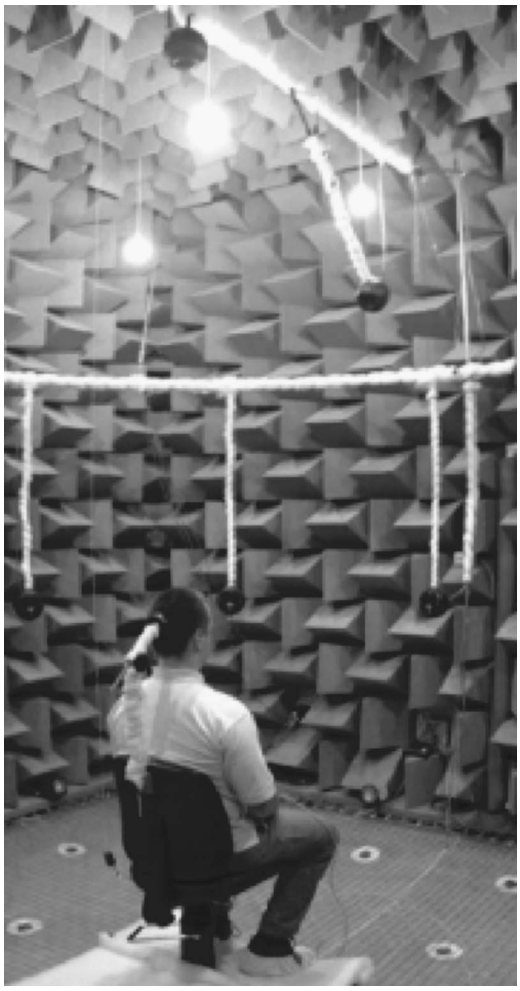


FIG. 1. The experimental setup in the anechoic chamber.

loudspeakers and the platform (an open metal grid) under the chair were covered with sound-absorbing material.

The subjective responses were collected with a two-button response box. The response box had small lights above the buttons to indicate observation intervals. An enlarged copy of the indicator lights was placed behind and slightly above the frontal loudspeaker to avoid subjects tilting their heads downwards to the response box in their hands.

2. Signal generation and control

All other equipment was placed in a control room next to the anechoic chamber. A personal computer (PC) was used for controlling the experiment and carrying out objective measurements. The PC was equipped with a digital sound card (RME DIGI96/8 PST) with eight audio channels, connected to an external AD/DA-converter (RME ADI-DS8). A custom-made eight-channel attenuator with a 128-dB dynamic range and 0.5-dB step size was used to individually control the level of the eight loudspeakers. The signals from the attenuator were amplified by power amplifiers (Rotel RB-976 Mark II), and then fed to the loudspeakers in the anechoic chamber.

The experiment was run using a program developed in LABVIEW. The program took care of reading session files, playing back appropriate stimuli, collecting subjects' responses, adapting the attenuator gains according to the responses, and writing the data into result files.

C. Measurements

Acoustical measurements were carried out using the maximum-length-sequence (MLS) system as specified by Olesen *et al.* (2000), with an MLS order of 12, preaveraging of 16, and a sampling rate of 48 kHz. The length of the acquired impulse responses was 4095 samples, which, due to the scarcity of reflections inside the anechoic chamber, was long enough to avoid time aliasing. The measurements were carried out at a level of approximately 70 dB SPL (at 1 kHz), measured in the absence of a listener at the center position of the setup.

First, responses of each loudspeaker [P_1 pressures, see Eq. (1)] were measured at the center position using a 1/4-in. pressure field microphone (Brüel & Kjær type 4136) with 90° incidence to the loudspeaker under measurement. Then, responses of each loudspeaker at each listener's ears [individual P_2 pressures, see Eq. (1)] were measured at the blocked entrance to the ear canal using two miniature microphones (Sennheiser KE 4-211-2), one microphone specifically for each ear. The miniature microphones were fitted inside foam earplugs (E·A·R Classic, halved in length), and mounted flush with the ear-canal entrance. All microphone signals were bandpass filtered between 22.5 Hz and 22.5 kHz by the measurement amplifier used (Brüel & Kjær type 2607 or type 2690 Nexus).

The above measurements were carried out three times: in the beginning, halfway through, and at the end of the experiment. The loudspeaker responses were used to equalize the stimuli for the listening experiment and to obtain reference pressures (P_1) for the HRTF calculations. The responses at each listener's ears were used to obtain individual HRTFs. The HRTF measurement procedure was as described by Møller (1995) with the following exceptions: The subjects were sitting in a chair instead of standing, the anechoic chamber was smaller, and the MLS measurement system was different.

Computation of the HRTFs involved 1024 samples from P_1 and P_2 pressures. First, individual head-related impulse responses (HRIRs) were calculated from P_1 and P_2 including a correction for the differences in the frequency responses of the two types of microphones used in the measurements. These HRIRs included reflections from the loudspeaker setup; therefore, only 140 samples from the HRIRs were used for calculating the final HRTFs. The resulting samples included all reflections from the subjects themselves (and from the chair), but excluded reflections from the other loudspeakers, the loudspeaker suspension, and any other objects inside the anechoic chamber. Note, however, that the excluded reflections were very small compared to the magnitude of the pure HRTFs.

D. Stimuli

The stimuli used for the listening experiment were third-octave noise bands centered at 0.4, 1, and 5 kHz. The length of each stimulus was 1 s.

For generating the stimuli, a 1-s white-noise signal was created, and subsequently filtered using third-octave-band filters at each center frequency. The relative differences in the frequency responses of the loudspeakers were equalized by applying minimum-phase inverse filters based on the direct sound coming from the loudspeakers. Each narrow-band signal was convolved with each of the inverse filters characterizing the individual loudspeakers, resulting in 24 stimuli for each (center frequency $\times$ loudspeaker) combination. Finally, raised-cosine rise and decay ramps of 20-ms duration were applied. The sound files thus corrected were played back at a sampling rate of 44.1 kHz, and with 16-bit resolution in the experiment proper.

The third-octave-band levels of the stimuli were aligned to 64.7 ± 0.1 dB SPL at 0.4 kHz, 64.7 ± 0.2 dB SPL at 1 kHz, and 63.9 ± 0.1 dB SPL at 5 kHz. At the highest possible playback level (75 dB SPL) the levels of the second- and third-order harmonics were more than 37 and 43 dB below the level of the center frequencies of the narrow-band noises, respectively. The distortion was measured to be highest at the lowest center frequency, but it was inaudible for all stimuli. Furthermore, the spectral envelope of the equalized stimuli was verified to be very similar between different loudspeakers.

In the experiment proper, the stimuli were played back at two overall levels as measured at the listener's position; a "low" overall level of around 45 dB SPL and a "high" overall level of around 65 dB SPL. Even though the actual measured sound-pressure levels deviated slightly from these values, note that the misalignment between the loudspeakers was less than ± 0.2 dB at each center frequency.

E. Procedure

The aim of the experiment was to determine how loudness is affected by the sound incidence angle at three center frequencies and two overall levels. This was accomplished by matching the loudness of test sounds emanating from each of the loudspeakers in the setup to the loudness of the same sound coming from the reference loudspeaker positioned in front of the subject at 0° of azimuth and elevation.

1. Adaptive matching procedure

Matches were obtained using a two-interval, adaptive forced-choice (2AFC) procedure (Levitt, 1971) converging on the point of subjective equality (PSE) by following a simple 1-up, 1-down rule. On each trial, the (variable) test sound, and the (fixed) frontal reference were presented in random order, with a 500-ms pause in between. Synchronized with the sounds, two light-emitting diodes were successively lit both on the hand-held response box, and on its larger model in order to mark the observation intervals to be compared. The subject's task was to judge which of the two noises sounded louder by pressing one of the two buttons aligned with the observation-interval lights. The participants were instructed to judge the loudness of the sounds only, and to disregard any other differences (due to direction, or timbre, for example) they might perceive.

For each adaptive track, the overall level of the frontal reference was fixed to either 45 or 65 dB SPL, as was the center frequency of the sounds to be played, and the test loudspeaker to be matched. The level of the test loudspeaker, however, was controlled by the adaptive procedure: Whenever the subject judged the test sound to be louder than the (frontal) reference, its sound-pressure level was lowered by a given amount; whenever the subject judged the reference to be louder, the level of the test loudspeaker was increased by that same amount. The initial step size was 4 dB; after two reversals (i.e., changes in the direction of the adaptive track) it was decreased to 1 dB. A total of eight reversals was collected in each adaptive track; the arithmetic mean of the last six of them was used to estimate the PSE. Two different starting levels were employed for the adaptive tracks, one 10 dB above, one 10 dB below the level of the reference loudspeaker, thus providing clearly discriminable loudness differences at the outset of each track.

2. Experimental design

For a given overall level, the experimental design required loudness matches to be determined in 44 different experimental conditions. These resulted from the factorial combination of 7(test loudspeakers) $\times$ 3(center frequencies) $\times$ 2(adaptive starting levels), and additional two conditions of the reference loudspeaker being matched to itself for the 1-kHz center frequency only (using both starting levels) to obtain a measure of the baseline variability of the matches.

Collection of these data was organized as follows: In order to allow subjects to adapt to a given loudness range, "high-SPL" (65 dB; "*A*"), and "low-SPL" (45 dB; "*B*") measurements were strictly separated in different sessions, which were counterbalanced following a succession of *ABBA* (respectively, *BAAB*) schemes. The order of the 44 experimental conditions to be investigated at each level was randomized, and subsequently divided into blocks of eight (the remaining four being assigned to the next block, i.e., the following replication of the measurements). Thus, within a given block of trials, eight adaptive tracks were randomly interleaved on a trial-by-trial basis, providing some random sampling of loudspeaker locations, center frequencies, and starting levels. Consequently, it was impossible for the subjects to track the immediate "adaptive" consequences of their judgments, and from their perspective the task was just a succession of unrelated paired comparisons with respect to loudness.

Each listening session consisted of four such blocks (containing eight adaptive tracks each). Completing a block of trials took approximately 10 min. While it lasted, the subjects were instructed to sit as still as possible in order to maintain the alignment with the loudspeaker setup. A short break was taken after each first and third block in a session, and participants were allowed to move their heads and upper body during those breaks, but not to leave the chair. After each second block they had a longer break during which they

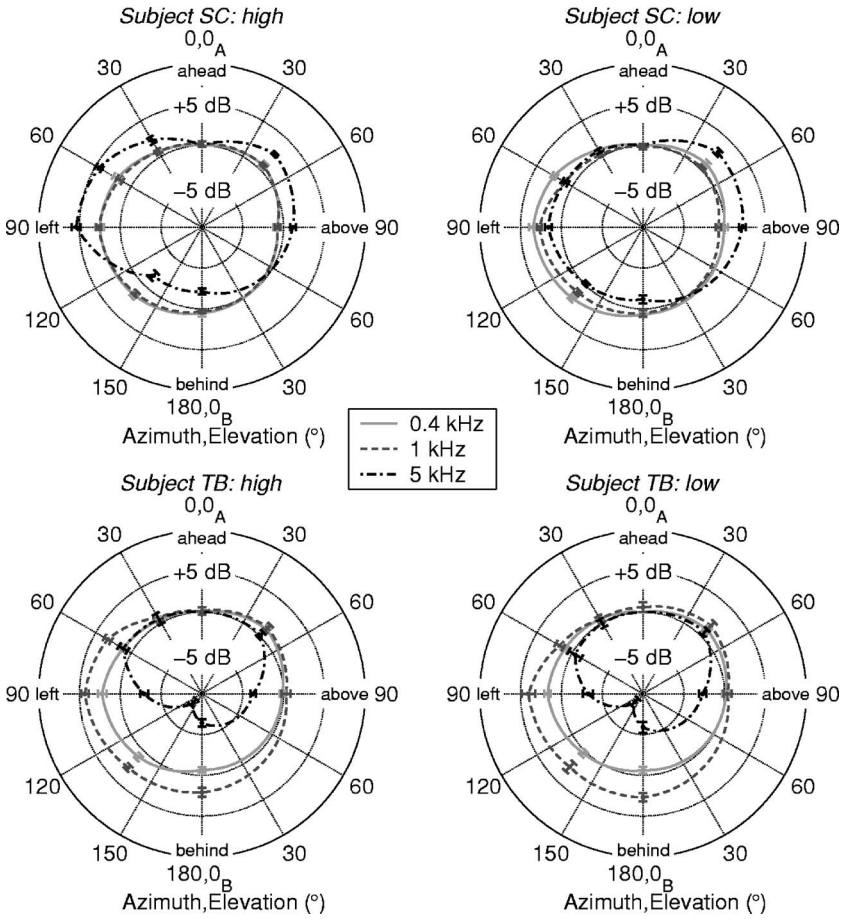


FIG. 2. Directional loudness sensitivities at the two overall levels for subjects SC and TB with 95% confidence intervals of 16 replications. The graphs on the left show the results for the high overall level (65 dB SPL), the graphs on the right for the low overall level (45 dB SPL). Elevations 0°_{A} and 0°_{B} are ahead and behind the listener, respectively.

left the anechoic chamber, thus requiring them to realign the seating position upon return. An entire session lasted approximately 1 h.

Since 16 replications of the matches (eight with each of the two adaptive starting levels) were collected per experimental condition, all subjects had to participate in 22 listening sessions. The participants completed a maximum of two sessions per day with a minimum of 1 h between sessions. With three additional sessions reserved for audiometry, HRTF measurements, and practice (one block in each of the high-SPL and low-SPL conditions), the total number of hours amounted to 25 per subject.

III. RESULTS

A. Directional loudness sensitivities

The adaptive procedure matched the loudness of a sound of a given center frequency coming from one of the loudspeakers in the horizontal or the median plane to the loudness of the same sound with frontal incidence. Thus, the raw data from the experiment were the sound-pressure levels (in dB SPL) the loudspeakers would have to be set to, in order to be perceived equally loud as the frontal reference. These raw data were averaged across the 16 repetitions that each participant accumulated in each condition, and normalized by subtracting the result from the fixed level of the respective frontal reference (65 or 45 dB SPL). That way, (relative) directional loudness sensitivities[1] were obtained, positive values of which indicate loudness enhancement, i.e., a lower sound-pressure level required for that direction to achieve a match with the frontal reference.

1. Individual data

Individual directional loudness-sensitivity curves are depicted for two subjects, SC (upper panels) and TB (lower panels), representing extremes of performance, in Fig. 2. The data are rendered in polar coordinates, though in a particular, asymmetrical way: The left-hand side of each polar graph shows the data for the horizontal plane as the loudspeakers were physically positioned in the setup. On the right-hand side of each polar graph the data are shown for the median plane where the loudspeakers were actually above the subjects. Note that these two planes coincide ahead of and behind the subjects.

For subject SC, loudness matches at 0.4 and 1 kHz vary as a function of sound incidence angle over a range of approximately 3 dB, the subject being most sensitive to loudness for sounds coming from the side, i.e., from 90° to the left in Fig. 2. That holds for both overall levels used. At 5 kHz, by contrast, this pattern is observed at the high overall level only, whereas at the low level the loudness pattern is fairly omnidirectional in the horizontal plane. In the median plane the directional patterns are similar across overall levels.

For subject TB, loudness matches vary over a range of less than 3 dB at 0.4 kHz. At 1 kHz the direction has a larger

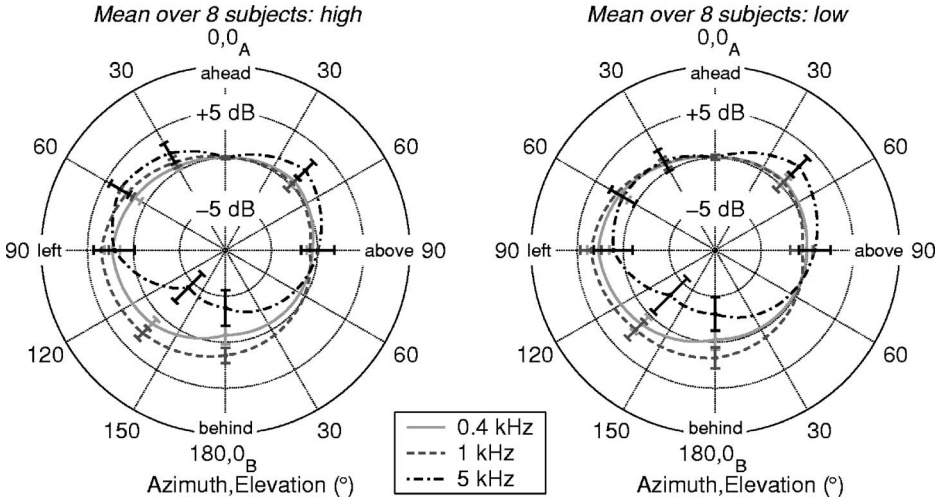


FIG. 3. Directional loudness sensitivities at the two overall levels for means across all eight subjects with 95% confidence intervals of the means. The graph on the left shows the results for the high overall level (65 dB SPL), the graph on the right for the low overall level (45 dB SPL). Elevations 0°_{A} and 0°_{B} are ahead and behind the listener, respectively.

effect on loudness, the sensitivity being up to 4 dB higher on the left-hand side than straight ahead. At 5 kHz the directional effect is even more pronounced, the minimum sensitivity at 135° in the horizontal plane being approximately 8 dB below and the maximum being close to the frontal sensitivity. The directional sensitivity patterns for this subject do not appear to be level dependent.

The confidence intervals for the matches of subjects SC and TB in Fig. 2 are small. Average individual standard deviations of the loudness matches across all subjects were 1.0 and 1.2 dB at the high and low overall levels, respectively. All participants adjusted the (identical) 1-kHz frontal test sound to a sound-pressure level close to the (fixed) reference, indicating that there was no systematic bias in the matches. The standard deviation of the identical-direction matches (0.9 dB) was only marginally lower than that of the across-direction matches, suggesting that these were of no greater difficulty.

2. Group data

Figure 3 shows mean loudness sensitivities when data are aggregated across all of the eight subjects. When the listener-specific idiosyncrasies are thus removed, directional loudness sensitivity still varies over some 3 dB at the two lowest center frequencies, whereas at 5 kHz the directional effect is approximately twice as large. Also, the error bars are larger at the highest center frequency due to a wider spread in the individual data. The overall level does not seem to have a marked effect on the patterns when considering the average data: the left and the right panels of Fig. 3 are hardly distinguishable.

The data and the subsequent analyses show that loudness is not constant over sound incidence angles, and the directional loudness-sensitivity patterns change considerably with center frequency, and to a lesser extent, with overall sound-pressure level.

3. Statistical analysis

The significance of the effects observed in the averaged data was confirmed by a $7\times3\times2$ (directions $\times$ center frequencies $\times$ levels) repeated-measures analysis of variance (ANOVA) on the means obtained from each subject in each of the experimental conditions.

In addition to a significant main effect of direction, $F(6,42)=28.35$, $p<0.001$, indicating that directional loudness-sensitivity differences persist, even when collapsing across levels and frequencies, all its interactions were highly significant:

(1) As expected, the direction $\times$ frequency interaction produced the highest F value, $F(12,84)=31.29$, $p<0.001$, confirming that the way in which directional loudness varies is strongly frequency dependent (see Fig. 3). It should be noted that this interaction is also highly significant for each of the eight subjects when statistical analyses are done individually.
(2) Furthermore, there is a significant direction $\times$ level interaction in the pooled data, $F(6,42)=7.29$, $p<0.001$. Inspecting the average directional loudness sensitivities in Fig. 3, it appears that—ignoring center frequency—the directional effects on loudness are slightly more pronounced at the higher overall level (65 dB SPL).
(3) More importantly, there is a three-way (direction $\times$ frequency $\times$ level) interaction, indicating that the frequency-dependent directional effects show a different pattern for the two overall levels, $F(12,84)=7.42$, $p<0.001$. This appears to be largely due to the 5-kHz data showing a slightly larger gain in sensitivity in front of the listener, and a slightly larger loss behind when comparing the high with the low overall level (see Fig. 3). Again, this interaction is significant for all of the eight subjects, even though the patterns show strong individual differences (see Fig. 2).

B. Head-related transfer functions

Individual head-related transfer functions were measured to investigate how sound is being filtered from a free field to the subjects' ears, depending on the angle of incidence. As an example, the HRTF magnitude spectra for subject IA from all eight directions are plotted in Fig. 4. Each panel depicts curves for the three separate sets of measurements made at different stages of the experiment. These mea-

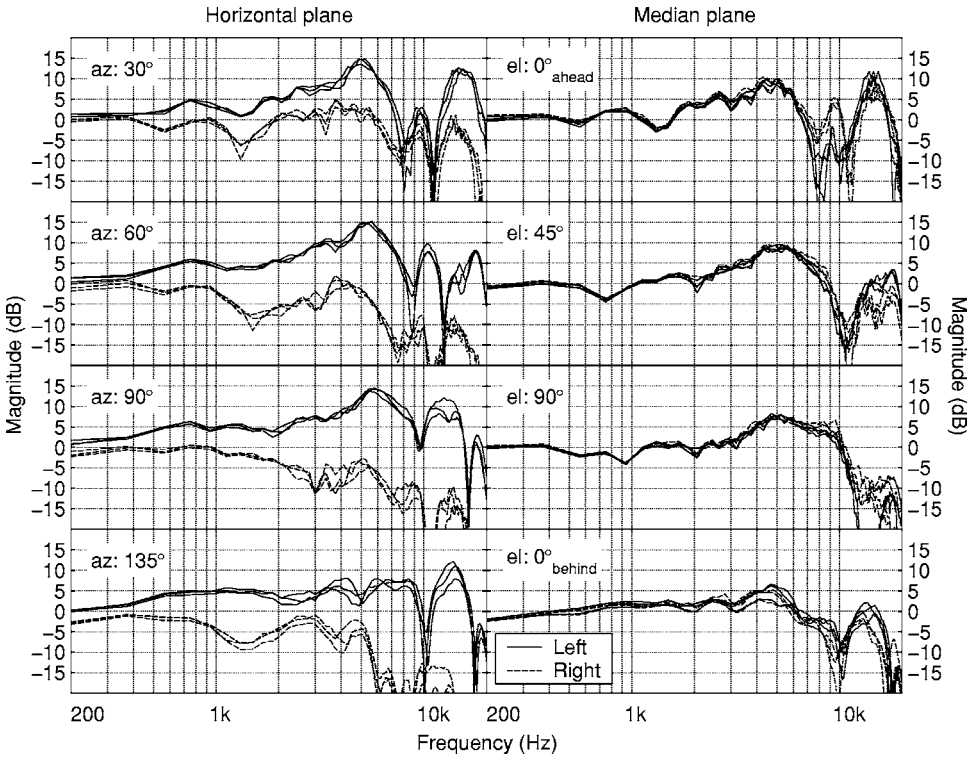


FIG. 4. Three HRTF measurements, performed at different stages of the experiment. The figure shows data for the left and right ears of a single subject (IA), for stimulation from all eight directions. The left panel depicts measurements obtained in the horizontal plane (azimuths of 30°, 60°, 90°, and 135°), the right panel those obtained in the median plane (elevations of 0°_{ahead}, 45°, 90°, and $0^{\circ}_{\text{behind}}$). 0°_{ahead} incidence is the frontal (reference) direction.

surements include individual fitting and positioning of the microphones, aligning of the subjects to the listening position, calibration, and acoustic measurements. As seen in Fig. 4, the measurements are highly repeatable, the variation below 1 kHz on average being within ±0.4 to ±0.6 dB (comparable to e.g., Møller *et al.*, 1995).

Figure 4 also shows that the interaural level differences in the median-plane HRTFs are very small up to around 7 kHz. In the horizontal plane, however, HRTFs of the left and right ears differ considerably due to a pressure buildup at the ipsilateral ear and head shadowing at the contralateral ear, especially at high frequencies. For the fairly representative subject whose HRTFs are depicted in Fig. 4, the maximum magnitudes of the ipsilateral (left) ear in the horizontal plane are around 15 dB for azimuths from 30° to 90° (front-left side), while the magnitudes at the contralateral (right) ear are typically below 0 dB.

C. HRTFs and directional loudness

1. Calculating normalized at-ear exposure

In order to investigate the effects of the physical HRTF filtering on the directional loudness matches on an individual basis, the objective HRTF measurements and the subjective loudness data were combined. This was done in order to obtain the actual frequency-specific at-ear exposure, and to evaluate whether the peculiarities of individual HRTFs might account for some of the interindividual variation seen in the directional loudness matches. Note that this analysis was based on the magnitude spectra of the HRTFs, and that the effect of the interaural time difference was disregarded.

The individual HRTFs were averaged across the three repetitions by calculating the mean of the linear magnitude spectra. These means were then converted to the corresponding third-octave-band levels in decibels. Finally, the left- and right-ear SPLs were normalized, for each incidence angle and at each center frequency, by subtracting the respective frontal left- and right-ear levels from them, since the loudness matches were always made to the frontal reference.

2. Relating loudness matches to HTRFs

For each of the eight participants, the normalized at-ear levels and directional loudness matches are combined in Fig. 5 and Fig. 6.

a. Horizontal plane. The combined data for the horizontal plane are plotted in Fig. 5. As seen in Fig. 5, in this plane the individual ILDs reach a maximum of 5 dB at 0.4 kHz, of 12 dB at 1 kHz, and of up to 30 dB at 5 kHz for the calculated third-octave-band at-ear SPLs.

For all subjects, except for subject IA at 0.4 kHz, the subjective directional loudness sensitivities at the high and low overall levels largely fall between the objective at-ear sound-pressure levels. It thus seems that the agreement between the two types of data is fair: For example, by considering the 5-kHz data for subjects TB and WE in Fig. 5, the idiosyncrasies in their at-ear SPLs are reflected in equally individual directional loudness sensitivities. However, the picture is not as clear when considering the two overall levels ("high" at 65 dB SPL and "low" at 45 dB SPL): Generally, the subjective data at the two overall levels are fairly congruous. In some cases, however, the most extreme case being subject SC at 5 kHz in Fig. 5, a clear overall level dependence can be observed.

If loudness were perceived as being constant over sound incidence angles, the subjective directional sensitivity data would follow the 0-dB horizontal in Fig. 5 or, equivalently, the 0-dB circles in Figs. 2 and 3. That would imply loudness

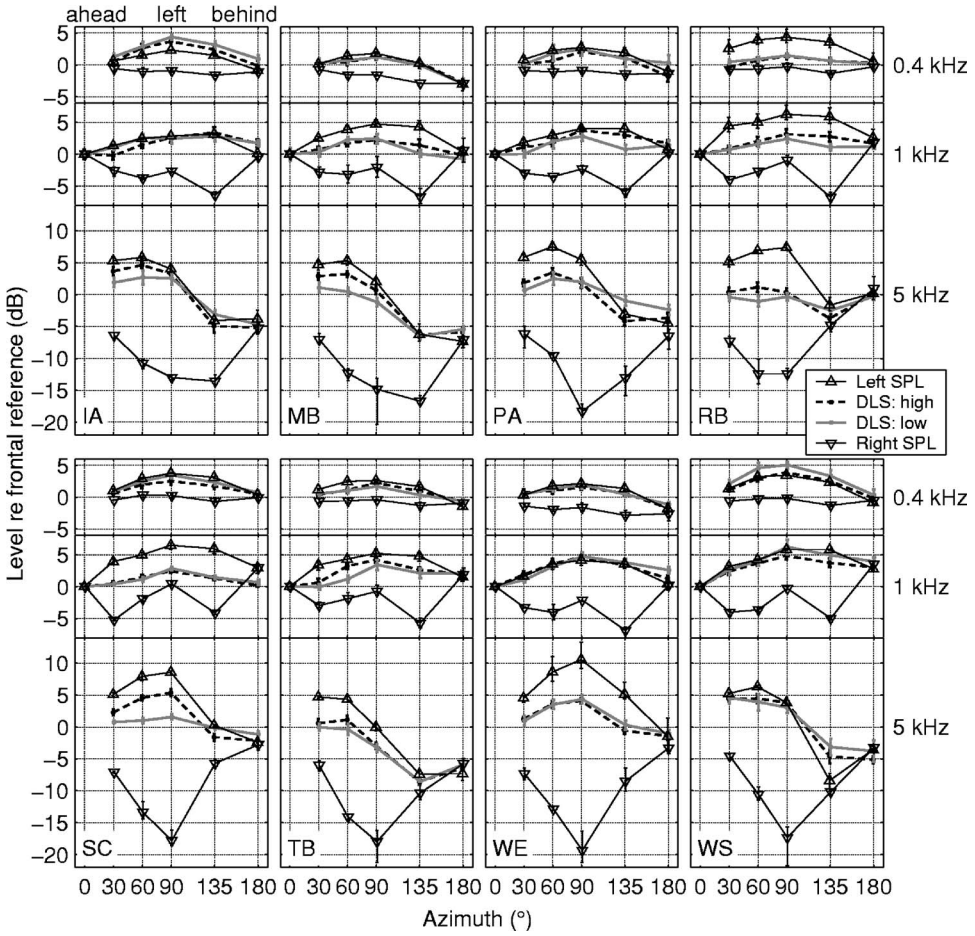


FIG. 5. Horizontal plane: Directional loudness sensitivities (DLS) at the high and low overalls level (DLS high: 65 dB SPL and DLS low: 45 dB SPL) with 95% confidence intervals, along with left- and right-ear sound-pressure levels, plotted relative to the frontal reference (see the text).

to be governed solely by the sound-pressure level of the source measured in the absence of a listener, irrespective of the changes in the at-ear sound-pressure levels as a function of sound incidence angle. This does not seem to be the case for any of the data sets.

If, on the other hand, the subjective loudness data always followed the ear with the higher SPL, this would imply *no* binaural loudness summation, i.e., loudness would be determined by the ear getting the higher input alone. Evidence for this kind of behavior may be seen in the data of IA, WS, and to some extent in those of WE and PA, though not at 5 kHz.

b. Median plane. In the median plane the ILDs are small, and the two ears are getting approximately the same

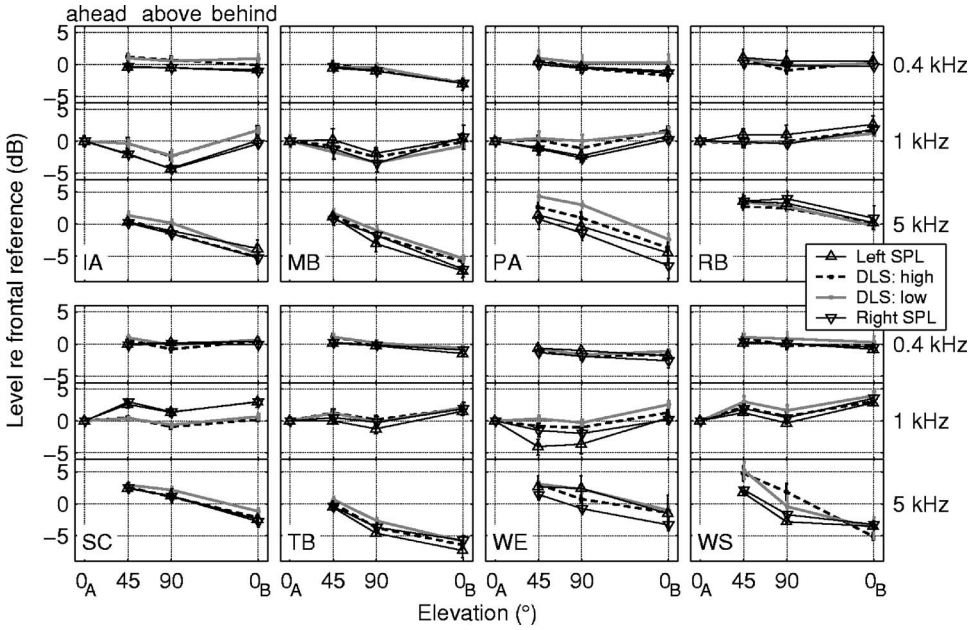


FIG. 6. Median plane: Directional loudness sensitivities (DLS) at the high and low overall levels (DLS high: 65 dB SPL and DLS low: 45 dB SPL) with 95% confidence intervals, along with left- and right-ear sound-pressure levels, plotted relative to the frontal reference (see the text). Elevations 0°_{A} and 0°_{B} are ahead and behind the listener, respectively.

input at all sound incidence angles; see Fig. 6. The differences between the ears are largest for subject WE, producing ILDs as large as 3 dB.

The normalized at-ear levels as a function of direction vary over less than 3 dB at 0.4 kHz, by up to 5 dB at 1 kHz, and over a range of almost 10 dB at 5 kHz. In this plane a change in the at-ear SPLs with incidence angle should presumably be reflected in a similar change in directional loudness sensitivity, which is true for most of the subjects. Occasional exceptions from this rule can be seen, however, for example for subjects SC and IA at 1 kHz, and for subject WS at 5 kHz.

c. Summary. Both in the horizontal and median planes, the patterns of the individual directional loudness sensitivities can largely be explained by directional effects the individual HRTFs have on at-ear sound-pressure levels. The way the subjects combine their left- and right-ear SPLs to a single loudness percept is further explored in the next section concerned with modeling binaural loudness.

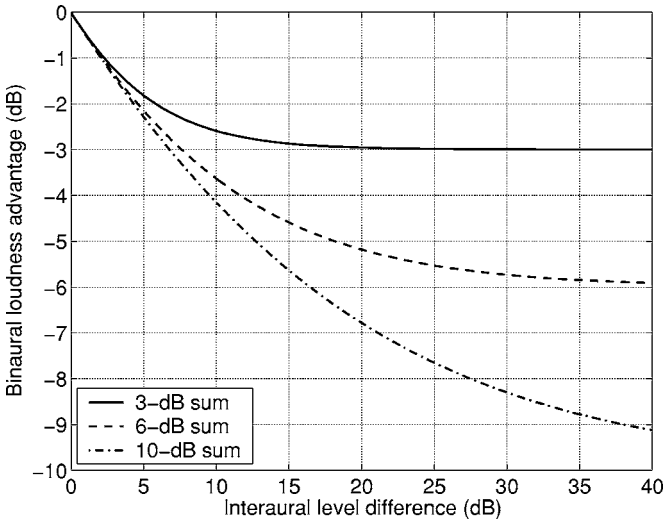


FIG. 7. Binaural loudness advantage as a function of interaural level difference; solid: 3-dB, dashed: 6-dB, and dash-dotted line: 10-dB summation rule.

IV. MODELING OF BINAURAL LOUDNESS

Large interindividual variation was found in subjects' directional loudness sensitivities. As seen in the previous section, these sensitivities exhibit systematic dependencies on the directional variations in individual HRTFs. Thus, a straightforward strategy in modeling binaural loudness is to take the HRTF effects into account, and to relate the physical changes in the at-ear signals—independent of direction—to the corresponding changes in loudness as perceived in a real sound field.

In the median plane, where the loudspeakers were positioned symmetrically with respect to the subjects' left and right ears, the listening situation was close to diotic. In this plane, the sound-pressure levels at the two ears were similar at the elevations under investigation (0°_{A}, 45°, 90°, and 0°_{B}); see Fig. 6. In such a situation, the actual amount of summation across the two ears has no effect on binaural modeling. This is due to the fact that the same binaural listening advantage takes effect both for the reference and the comparison to be matched. Note that the same applies for the traditional free- and diffuse-field loudness paradigms.

Dichotic stimulation, with different at-ear levels, thus constitutes the most interesting case for the modeling of binaural loudness. Dichotic at-ear SPLs were observed for the azimuths of 30°, 60°, 90°, and 135° in the horizontal plane (see Fig. 5). At these azimuths subjects typically had to match a dichotic sound to the diotic frontal reference.

Narrow-band stimuli were used in the listening experiment in order to simplify the modeling of binaural loudness, by being able to ignore spectral summation of loudness across critical bands. Also, assuming that perceived loudness is doubled when the listening is binaural (diotic) instead of monaural, a relationship between the psychophysical dimension of loudness (as measured in sones) and its physical correlate, the sound-pressure level (in dB SPL) can be established. By definition, a loudness of 1 sone is produced by a 40-dB SPL, 1-kHz tone, and doubling or halving loudness (in sones) corresponds to a 10-dB increment or decrement in sound-pressure level, respectively. Due to the shape of the equal-loudness contours (ISO 226, 2003), the increments within the range of sound-pressure levels used in the present experiment are approximately 10.5 and 9.5 dB SPL at 0.4 and 5 kHz, respectively, for a doubling of loudness. Thus, at all three center frequencies (0.4, 1, and 5 kHz) used in the present study, doubling in sones corresponds fairly closely to a 10-dB gain in sound-pressure level.

In order to illustrate how binaural loudness is affected by various interaural level differences, theoretical curves can be obtained utilizing Eq. (2), taken from Robinson and Whittle (1960). It is reasonable to assume that the summation of sound-pressure levels across the two ears is nonlinear, as suggested by Eq. (2): At large ILDs, the ear receiving the lower sound-pressure level presumably has little effect on overall binaural loudness, and the stimulation is effectively monaural. When approaching a diotic situation, however, the signals at the two ears tend to be weighted equally in contributing to overall loudness.

Theoretical curves for three hypothetical binaural loudness-summation rules are plotted as a function of the interaural level difference in Fig. 7. In addition to the 6-dB summation rule adopted from Robinson and Whittle (1960), two other curves were derived by changing the binaural gain factor in Eq. (2): A 3-dB summation rule corresponding to the "power summation" of the linear at-ear magnitude spectra, and a 10-dB summation rule, which for the stimuli used in the present study roughly corresponds to perfect binaural summation in sones.

The different curves in Fig. 7 are normalized so that they all coincide in the origin of the graph: it represents the diotic case with an ILD of zero. As the ILD increases, loudness decreases by different amounts, depending on the summation rule with the "loss" to be read from the ordinate corresponding to the "binaural loudness advantage" achievable by switching from dichotic to diotic stimulation. The 3-dB summation rule fairly quickly converges to the −3-dB level in the graph: when the ILD increases beyond 15 dB, binaural loud-

TABLE I. Least-squares estimates for the amount of binaural loudness summation [x in Eq. (3), in dB], at the three center frequencies (f_c: 0.4, 1, and 5 kHz) at high (65 dB SPL) and low (45 dB SPL) overall levels. The two right-most columns show the best fits when pooling center frequencies, and the best fits across center frequencies when the data are averaged across subjects (bottom row). Extreme values are marked with stars (see the text for details).

	f_c							
	0.4 kHz		1 kHz		5 kHz		Best fit across f_c	
Subject	High	Low	High	Low	High	Low	High	Low
IA	0.1	0.1	0.1	0.1	0.7	1.5	0.1	0.1
MB	0.3	0.4	3.3	3.5	2.2	4.8	2.4	4.6
PA	1.0	0.4	0.7	1.3	3.1	3.0	2.1	2.4
RB	99.9*	99.9*	4.0	9.3	10.0	18.4*	9.1	17.6*
SC	1.6	0.3	13.1	29.7*	3.1	8.2	3.8	8.5
TB	1.1	2.8	2.1	3.0	3.7	5.0	3.3	4.6
WE	0.1	0.1	0.1	0.3	4.9	4.5	2.8	2.6
WS	0.1	0.1	1.7	0.5	0.1	0.1	0.7	0.1
Median	0.7	0.4	1.9	2.2	3.1	4.7	2.6	3.6
Averaged data							2.6	3.9

ness is no longer affected. At these ILDs loudness is determined by the ear with the higher sound-pressure level alone, and dichotic loudness is 3 dB lower than the corresponding diotic one. With the 6- or 10-dB summation rules, much larger ILDs are required until the curve asymptotes at −6 and −10 dB, respectively. For the 10-dB summation rule, at an ILD of 40 dB (far exceeding the ILDs observed in the present study) binaural loudness still continues to decrease.

A. Individual data

The third-octave-band at-ear sound-pressure levels computed from the HRTFs were used in the modeling, in order to find the best-fitting binaural summation rule to predict the directional loudness-sensitivity data. Robinson and Whittle (1960) reported their average data to support a 6-dB loudness summation across their listeners' ears [see Eq. (2)]. This type of modeling was explored for the present data, but on an individual basis. The modeling was carried out by relaxing the factor 6 in Eq. (2).

To that effect, the optimal amount of binaural loudness summation (x)—assumed to be fixed at 6 dB in Eq. (2)—was *estimated* by minimizing the sum-of-squares of the errors (SSE) between the actual directional loudness sensitivity (DLS) and the sensitivity predicted (L_{mon}) from the changes in at-ear sound-pressure levels using Eq. (3). All 16 (j) repetitions of each condition, and the mean at-ear sound-pressure levels for each of the four horizontal-plane angles of incidence (i; 30°, 60°, 90°, and 135°) were included in the modeling, which was performed individually for each subject, and separately for the three center frequencies and the two overall levels.

$$\mathrm{SSE} = \sum_{i=1}^{4} \sum_{j=1}^{16} \{\mathrm{DLS}_{\mathrm{high/low}_{i,j}} - [L_{\mathrm{mon,comp}_i}(x) - L_{\mathrm{mon,ref}}(x)]\}^2, \quad (3)$$

where

$$L_{\mathrm{mon,comp}_i}(x) = x \times \log_2(2^{L_{\mathrm{left,comp}_i}/x} + 2^{L_{\mathrm{right,comp}_i}/x}), \quad (4)$$

and

$$L_{\mathrm{mon,ref}}(x) = x \times \log_2(2^{L_{\mathrm{left,ref}}/x} + 2^{L_{\mathrm{right,ref}}/x}). \quad (5)$$

In these equations, $L_{\mathrm{left,comp}}$ and $L_{\mathrm{right,comp}}$ refer to the third-octave-band levels for the comparison incidence calculated from the individual left- and right-ear HRTFs, respectively. Likewise, $L_{\mathrm{left,ref}}$ and $L_{\mathrm{right,ref}}$ refer to the corresponding levels for the frontal reference at the left and right ears, respectively.

The subjective directional loudness sensitivities had been normalized to the frontal reference (see Figs. 2 and 5). Therefore, the predictions were normalized as well by subtracting Eq. (5) from Eq. (4) in the minimization of the sum of squares of the errors. Due to this normalization, the overall level (65 vs 45 dB SPL) does not have an influence on the predictions. The possible dependence of binaural loudness summation may nevertheless show up in the subjective directional loudness sensitivities at the high and low overall levels, and may thus influence the estimate of the variable x, the binaural gain estimated from the data. Forty-eight such estimates (for eight subjects, three center frequencies, and two overall levels) for the amount of binaural loudness summation are listed in Table I. The minimization algorithm was limited to a summation value between 0.1 and 99.9 dB.

As was already seen in Fig. 5, the amount of binaural loudness summation varies greatly across subjects, and also within subjects across the three center frequencies. The best-fitting binaural gain estimates roughly fall into three categories: The summation is minor (less than 1 dB) for 19, moderate (from 1 to 10 dB) for 24, and extreme (greater than 10 dB) for 5 out of the 48 cases analyzed. There is a tendency for the summation values to increase with center frequency, but due to the fact that the center frequencies are confounded with variations in ILDs, the comparison may not be fair.

The smaller the amount of binaural loudness summation, the more binaural loudness is determined by the ear getting a higher input. By contrast, the higher the summation value, the more influence the ear receiving the lower sound-pressure level has on binaural loudness. Some extreme values marked with stars in Table I, e.g., subject RB at 0.4 kHz, seem to imply the latter behavior. Closer inspection of Fig. 5, however, reveals that for this subject the directional loudness sensitivity remains close to the 0-dB line, even if the at-ear sound-pressure levels vary over a fairly wide range. As Robinson and Whittle (1960) pointed out, the actual value of the summation parameter (at the natural ILDs in question) does not have a great effect on the directional loudness sensitivities predicted from the at-ear SPLs. For these reasons the minimization algorithm can reach very high summation values (up to the limit of 99.9 dB) when searching for the best fit. However, it is unrealistic that the binaural gain (i.e., the loudness match between monotic and diotic stimulation) for a normal-hearing subject is much larger than 10 dB.

To get a more stable estimate, the amount of binaural summation was also determined by pooling across the three center frequencies; see the two right-most columns in Table I. This was achieved by aggregating the data across center frequencies, and finding the best-fitting summation rule to the aggregated data set. The individual differences are still retained, and the summation values again fall into the three categories defined above.

In order to deal with the variance inherent in the subjective data, a partial F-test (Bates and Watts, 1988, Chap. 3) was performed to investigate whether the subjects summed their at-ear levels in significantly different ways. In a "restricted" model *one* least-squares fit of binaural loudness summation [x in Eq. (3)] common to all subjects was estimated, whereas in a "full" model the summation value was relaxed to estimate *different* parameters for the eight subjects. The data were aggregated across incidence angles, overall levels and center frequencies. The partial F-test showed that the error sum of squares between the subjective data and the estimate was significantly larger for the restricted model having a common parameter for all subjects [$F(7,3064)=211.58$; $p<0.001$]. Therefore, the full model allowing for individually different binaural-gain parameters predicted the data better, and hence, the differences in the way the subjects summed the at-ear levels appear to be significant.

B. Group data

The individual third-octave-band HRTFs and directional loudness sensitivities were averaged across subjects, to make an estimate for the mean data thus obtained. Aggregating over center frequencies, as before, the best fits for the averaged data came fairly close to suggesting a 3-dB summation rule both at the high and the low overall level (see the bottom row of Table I).

Thus far the prediction was entirely based on the at-ear sound-pressure levels at the center frequency of the narrow-band noises used. However, by using a loudness model, the possible spread of excitation to neighboring critical bands can be taken into account in the modeling. Furthermore, given a relatively large dynamic range, the shape of the loudness function may be better accounted for when using a loudness model.

Therefore, the most widely accepted loudness model by Moore *et al.* (1997) was tested in predicting the present data. This model facilitates the use of eardrum pressures for loudness computations, i.e., using at-ear signals as a product of the HRTF-filtering stage. The model also predicts monaural loudness, by assuming perfect loudness summation in sones between the two ears, and calculating monaural loudness simply as being one half of the binaural, diotic loudness. Dichotic loudness can then be computed as a sum of the two monaural loudness values in sones.

Since the HRTFs of the present study had been measured at the entrance to the blocked ear canal, a direction-independent transfer from the measurement point to the eardrum (mean P_4/P_2) was adopted from Fig. 13 in Hammershøi and Møller (1996). In contrast to the summation rule explored in the previous section, here absolute binaural loudness values were computed. The effects of the HRTFs were taken into account, as before, but now the entire at-ear spectra were included (instead of only using the level at the center frequency). The input data to the loudness model thus were third-octave-band spectra based on the measured stimulus spectrum in the absence of a listener (P_1), combined with the left- and right-ear HRTFs (P_2/P_1), and corrected by the eardrum-to-the-measurement-point transfer function (P_4/P_2).

Monaural loudness values were computed for (dichotic) left- and right-ear signals, subsequently summed, and compared to the loudness produced by the (close to) diotic frontal reference. First, binaural loudness values for each of the frontal reference stimuli were computed, as described above. Then, sone values for the comparison directions were computed by varying the level of the P_1 pressures, within the range of ±10 dB from the frontal reference level, in steps of 0.5 dB. The P_1 sound-pressure levels yielding the binaural loudness values closest to that of the frontal reference were selected. In this way the loudness model was used to find equal-loudness sound-pressure levels for each incidence angle, including the effects of the HRTFs. The inverses of these sound-pressure levels relative to the frontal reference were taken as the directional loudness sensitivities predicted by the model.[2]

Figure 8 contrasts the predictions made by the loudness model (Moore *et al.*, 1997) with the 3-dB power summation, which fared best in the earlier analysis. Since the effect of overall sound-pressure level on directional loudness was minor for the averaged data (see Fig. 3), only the high-level (65 dB SPL) directional sensitivities are plotted.

For all dichotic situations (horizontal plane, left column in Fig. 8), the 3-dB summation rule predicts the obtained mean loudness-sensitivity data quite well. At each center frequency, the patterns of the 3-dB prediction and the actual matches made are congruous, and only in two instances (at 0.4 kHz, azimuths of 90° and 135° in Fig. 8) do the 95% confidence intervals of the subjective data not include the 3-dB prediction. By contrast, the prediction of the loudness

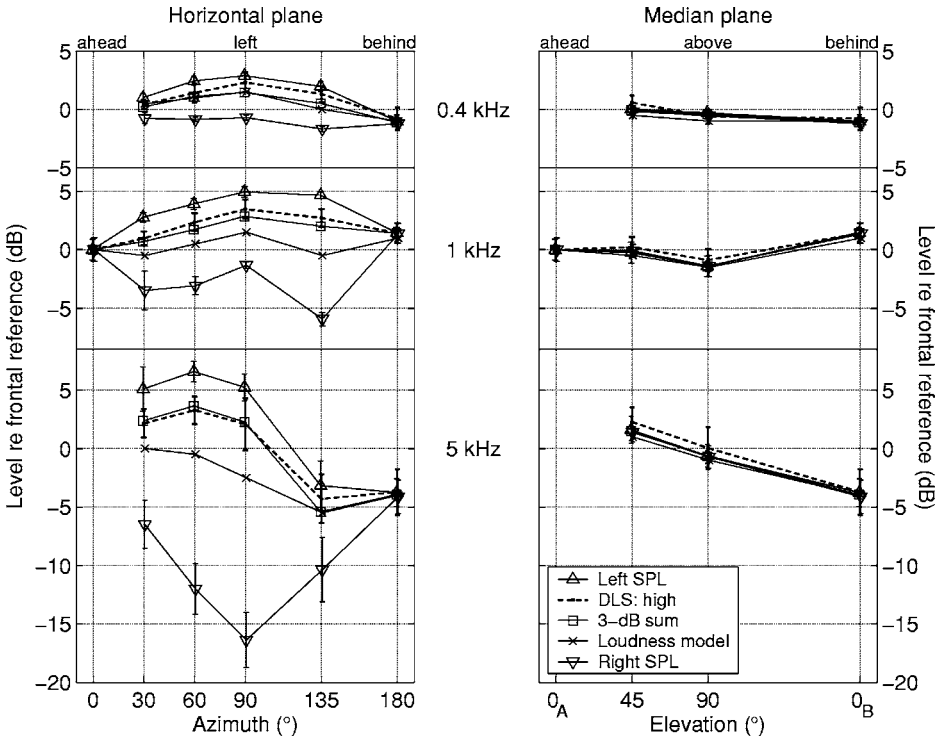


FIG. 8. Average left- and right-ear sound-pressure levels, a 3-dB summation rule, loudness summation in sones, and obtained average DLS at the high overall level (65 dB SPL). The error bars denote the 95% confidence intervals of the means across the eight subjects. The left panel depicts data and predictions for the horizontal plane, the right panel for the median plane. Elevations 0°_{A} and 0°_{B} are ahead and behind the listener, respectively.

model markedly deviates from the obtained directional loudness sensitivities, particularly at the two higher center frequencies (1 and 5 kHz). These are the situations in which the interaural level differences range from 6 to over 20 dB. For these ILDs, the prediction is not bracketed by the confidence intervals of the data for seven out of the eight dichotic conditions, the difference between data and predictions reaching up to 5 dB (5 kHz, azimuth 90° in Fig. 8). It thus seems that the 3-dB summation rule of at-ear sound-pressure levels predicts the directional loudness of dichotic sounds considerably better than the assumption of perfect binaural loudness summation in sones.

In the median plane, all five curves (at-ear levels, directional loudness sensitivities, and model predictions) are nearly indistinguishable; see the right panels in Fig. 8. The 95% confidence intervals of the subjective data include both the physical changes in left- and right-ear sound-pressure levels, and the predicions of 3-dB sum and loudness summation in sones. Obviously, the diotic stimulation condition does not provide a critical test for these models.

V. DISCUSSION

A. Comparison with previous work

When comparing the present results to the work of Robinson and Whittle (1960), it may be observed that the average directional effect sizes they obtained are comparable to those measured in the present study: For the incidence angles presented here, the average directional loudness sensitivities Robinson and Whittle (1960) obtained at center frequencies below 6.4 kHz varied from −6.5 to +5.0 dB (see their Fig. 2) relative to the frontal reference level. The corresponding range for the average data in the present study is −4.3 to +3.5 dB (see Fig. 3), although the actual stimulus center frequencies used differed somewhat between the two investigations.

In both investigations, direction had a smaller effect on loudness at lower center frequencies, and the effect increased with stimulus center frequency. Qualitatively, this can be explained by the fact that with increasing frequency the physical dimensions of a listener start to obstruct the sound field. The obstruction also becomes more direction dependent at higher frequencies (as can be seen in the sample HRTFs plotted in Fig. 4), and this is reflected in its increasing effect on the directional loudness sensitivities.

The present empirical data collection, however, goes beyond previous work by reporting individual analyses. Consequently, and as expected from research on HRTFs, idiosyncratic directional loudness-sensitivity patterns were found. The individual data also showed that all participants were highly consistent in their judgments, even though the loudness of two sounds coming from different directions, and typically having different timbres, had to be compared.

The consistency in the participants' directional loudness matches provided considerable statistical power. On the one hand, that means that the significance of the major frequency-dependent effects of the direction of incidence on perceived loudness may be ascertained with great confidence. On the other hand, that entails that even small effects on the range of 1–2 dB level will emerge as statistically significant, and thus require further interpretation. That is the case for the effects of overall presentation level, [3] and its interaction with the directional and frequency-specific effects.

Comparison of both individual data (e.g., Fig. 2, top row) and of the group averages (Fig. 3) shows a tendency for the frequency-dependent directional effects to become more

pronounced with increasing level. Likewise, small but systematic level effects are found when trying to estimate the amount of binaural gain from the data (Table I). Contrary to what is reported in the literature (Shaw *et al.*, 1947; Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963; Scharf and Fishken, 1970), this gain appears to be smaller at the higher overall level. That may be due to the low-level directional sensitivities being less affected by the ear receiving the greater input than the high-level directional sensitivities (see Fig. 5). Due to the small magnitudes of the overall-level effects, the present authors consider them to be negligible for most practical purposes, at least in the midlevel range investigated here (45–65 dB SPL).

Furthermore, the relatively low binaural-gain parameter derived from the present data is in conflict with the outcome of most of the classical studies (such as Reynolds and Stevens, 1960; Hellman and Zwislocki, 1963; Marks, 1978; Zwicker and Zwicker, 1991, among others) employing headphones, and largely focusing on monotic-to-diotic comparisons. But, note that—apart from other methodological distinctions—a key feature of these earlier studies is that signals may have been generated that would never reach the two ears when being emitted by a real source positioned in space, and fail to produce an externalized auditory event. It is unclear whether the results of the two paradigms (binaural loudness summation versus directional loudness) can be compared directly, since the auditory events produced are so drastically different. The directional loudness paradigm, however, is not only closer to "real-world" stimulation, but also to the application of measuring sound fields using a dummy head, where the signals at the ears of the dummy are due to the physical obstruction in the sound field.

B. Individual differences

Even though tentative general conclusions on the computation of binaural loudness may be drawn from the present data, it is striking how large the interindividual differences in loudness matches (see Figs. 2 and 5), and hence, in directional loudness sensitivity are when comparing the eight listeners participating. The original hope, that *all* of this interindividual variance might be accounted for by the equally large differences in individual HRTFs (e.g., Fig. 4) does not seem to be warranted, as is evident from our analysis of individual "summation rules" displayed in Table I. Obviously, using the actual at-ear sound-pressure levels rather than the levels emitted by the loudspeakers in the analysis still leaves us with considerable residual individual variance.

Several potential reasons for that variance might be explored: An obvious reason may be that the third-octave-band levels derived from the HRTF measurements do not reflect the actual at-ear stimulation well enough. However, the quality of the HRTFs may be examined by contrasting the present results with data obtained in individual binaural synthesis where the directional sound sources are recreated via virtual acoustics, the crucial difference being that the at-ear levels are precisely known in that situation. Performing such an experiment with six listeners from the original sample of eight (Sivonen *et al.*, 2005), we found no appreciable, or statistically significant, differences between the two sets of data (real vs virtual sound field). Rather, the individual differences remained, leading us to look for factors other than differences in the physical shape of pinnae, heads, and torsos.

A more speculative explanation for the individual differences found might be that the participants exhibited different degrees of "loudness constancy" in our experimental setup. The notion of "perceptual constancy" refers to situations in which a percept remains constant despite profound changes in the physical stimulation affecting the sensory receptors (Zahorik and Wightman, 2001). Typically, loudness constancy is observed when the loudness of a source (e.g., a musical instrument, a human voice) is judged to remain constant, even though its distance to the observer is varied. Stretching this notion somewhat, we might also speak of loudness constancy when listeners judge sounds to be equally loud, despite variations in their angle of incidence (which greatly affects the at-ear stimulation). It might be speculated that observers have learned to deconvolve the signals with their HRTFs in order to infer the loudness at the source.

Do the present data show evidence for loudness constancy defined in this way? The answer is clearly negative: Note that perfect constancy would mean that all of the identical-distance, identical-power sources used in the present experiment should be judged to be equally loud, i.e., the matches should fall on the 0-dB (reference) circle in Fig. 2, or on the 0-dB horizontal in Figs. 5 and 6. That, obviously, is not the case. Nevertheless, subjects might have a tendency to preserve constancy to varying degrees, thus producing different amounts of bias towards the zero line. Potentially, they could do so by using the localization and timbre cues available, as well as the fact that the loudspeakers producing the sounds are in plain view.

The constancy problem is related to that of the "listening attitude" a participant might adopt: In a pioneering investigation of loudness constancy (Mohrmann, 1939), this was operationalized as judging hidden sources at various distances while either adopting a sender attitude ("Sendereinstellung;" p. 155), or a receiver attitude ("Empfangseinstellung") which yielded appreciably different results. In modern terminology one would refer to judging the distal stimulus vs the proximal stimulus, and in the present situation that would be equivalent to judging the sound power of the loudspeaker as opposed to judging how it affects the listener. It is unclear, however, whether subjects can make that distinction in an anechoic situation, and the present authors know of no published reports implementing the instructional variations required.

Nevertheless, it may safely be said that a "bias" towards constancy can only play a minor role in accounting for the present data. The fact that knowing the individual HRTFs goes such a long way in accounting for the idiosyncrasies seen in the matches argues against constancy being a major factor in directional loudness perception, at least for the synthetic sounds and the anechoic environment studied here.

VI. CONCLUSIONS

(1) Loudness matches obtained with narrow-band noises in an anechoic environment showed that loudness is not constant over sound incidence angles. Rather, directional loudness sensitivities varied by up to 10 dB in individual, and up to 8 dB in averaged data.
(2) The directional effects on loudness showed considerable dependency on center frequency, with greater directional effects being observed at higher center frequencies, and to some extent on the overall sound-pressure level of the stimuli.
(3) Large, but highly reliable individual differences in directional loudness perception were observed.
(4) The individual patterns of directional loudness could largely be accounted for by the corresponding changes in physical stimulation, as determined by head-related transfer functions (HRTFs).
(5) These transfer functions were utilized for modeling binaural loudness based on the at-ear sound-pressure levels encountered. A 3-dB binaural loudness-summation ("power-summation") rule predicted the obtained mean data best, but sizable interindividual differences remained, even after the effect of individual HRTFs was taken into account.
(6) The results can be used for predicting loudness in any type of sound field (be it free, diffuse, or directional, resulting in diotic or dichotic at-ear signals) using a dummy head.

ACKNOWLEDGMENTS

The authors would like to thank Claus Vestergaard Skipper for his help in setting up the apparatus used in the present experiment. Henrik Møller and Dorte Hammershøi are thanked for their advice on how to carry out HRTF measurements. Pauli Minnaar helped in analyzing the HRTFs and with including them in the modeling, Florian Wickelmaier advised us with the statistical analyses. We would further like to thank Brian C.J. Moore for giving us access to and advice in utilizing his loudness model. Finally, we would like to thank the participants for their patience and devotion donated to extensive and at times exhaustive listening tests. This research was carried out as part of the "Centercontract on Sound Quality" which establishes participation in and funding of the "Sound Quality Research Unit" (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and DELTA Acoustics & Vibration. Further financial support comes from the Ministry for Science, Technology, and Development (VTU), and from the Danish Research Council for Technology and Production (FTP).

[1]Directional loudness sensitivities are thus defined in loose analogy to the directivity characteristics of microphones (Beranek, 1986, Chap. 6). Being the inverse of the relative sound-pressure level required to produce a loudness match, they are—despite the similarity in terms—not related to sensitivity parameters (such as d') as conceptualized in signal detection theory (Green and Swets, 1988).

[2]These predictions were made both for the individual and the mean data, essentially yielding the same conclusions. Thus, only the results for the mean data are presented here.

[3]Even though a 20-dB range may not appear sufficient to investigate the effects of overall level, note that when considering the extra headroom required for the adaptive starting values (±10 dB), and HRTF effects boosting or attenuating levels by approximately the same amount, the effective range listeners were exposed to in the experiment was quite large, covering what can be handled in a loudness-matching experiment without encountering floor ("too soft") or ceiling ("too annoying") problems.

Bates, D. M., and Watts, D. G. (**1988**). *Nonlinear Regression Analysis and Its Applications* (Wiley, New York).
Beranek, L. L. (**1986**). *Acoustics* (American Institute of Physics, New York).
Blauert, J. (**1997**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).
Bodden, M. (**1997**). "Instrumentation for sound quality evaluation," Acust. Acta Acust. 83, 775–783.
Bronkhorst, A. W. (**1995**). "Localization of real and virtual sound sources," J. Acoust. Soc. Am. **98**, 2542–2553.
Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsula, Los Altos, CA).
Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.
Hellman, R. P., and Zwislocki, J. (**1963**). "Monaural loudness function at 1000 cps and interaural summation," J. Acoust. Soc. Am. **35**, 856–865.
ISO 226 (**2003**). "Acoustics—Normal equal-loudness-level contours," (International Organization for Standardization, Geneva, Switzerland).
ISO 389-1 (**1998**). "Acoustics—Reference zero for the calibration of audiometric equipment—Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones," (International Organization for Standardization, Geneva, Switzerland).
ISO 389-7 (**1996**). "Acoustics—Reference zero for the calibration of audiometic equipment—Part 7: Reference threshold of hearing under free-field and diffuse-field listening conditions," (International Organization for Standardization, Geneva, Switzerland).
ISO 532 (**1975**). "Acoustics—Method for calculating loudness level," (International Organization for Standardization, Geneva, Switzerland).
Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.
Jørgensen, S. L. (**2002**). "MATLAB model for loudness perception—directional loudness," Unpublished masters thesis, Technical University of Denmark (DTU), Copenhagen.
Kuhl, W., and Westphal, W. (**1959**). "Unterschiede der Lautstärken in der ebenen Welle und im diffusen Schallfeld" ("Differences in loudness levels between a plane wave and a diffuse sound field,") Acustica **9**, 407–408.
Levelt, W. J. M., Riemersma, J. B., and Bunt, A. A. (**1972**). "Binaural additivity of loudness," Br. J. Math. Stat. Psychol. **25**, 51–68.
Levitt, H. (**1971**). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Marks, L. E. (**1978**). "Binaural summation of the loudness of pure tones," J. Acoust. Soc. Am. **64**, 107–113.
Mohrmann, K. (**1939**). "Lautheitskonstanz im Eutfernungswechsel" ("Loudness constancy with changing distance,") Z. Psychol. **145**, 145–199.
Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). "Head-related transfer functions of human subjects," J. Audio Eng. Soc. **43**, 300–321.
Møller, H., Sørensen, M. F., Jensen, C. B., and Hammershøi, D. (**1996**). "Binaural technique: Do we need individual recordings?," J. Audio Eng. Soc. **44**, 451–469.
Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness, and partial loudness," J. Audio Eng. Soc. **45**, 224–239.
Olesen, S. K., Plogsties, J., Minnaar, P., and Møller, H. (**2000**). "An improved MLS measurement system for acquiring room impulse responses," in Proceedings of NORSIG 2000, IEEE Nordic Signal Processing Symposium, Kolmården, Sweden, 117–120.
Remmers, H., and Prante, H. (**1991**). "Untersuchung zur Richtungsabhängigkeit der Lautstärkeempfindung von breitbandigen Schallen" ("Investigation of the directional dependence of the loudness of broadband sounds"), *Fortschritte der Akustik*-Deutsche Arbeitsgemeinschaft für Akustik (DAGA), 537–540.
Reynolds, G. S., and Stevens, S. S. (**1960**). "Binaural summation of loudness," J. Acoust. Soc. Am. **32**, 1337–1344.
Robinson, D. W., and Whittle, L. S. (**1960**). "The loudness of directional sound fields," Acustica **10**, 74–80.
Robinson, D. W., Whittle, L. S., and Bowsher, J. M. (**1961**). "The loudness

of diffuse sound fields," Acustica **11**, 397–404.
Scharf, B., and Fishken, D. (**1970**). "Binaural summation of loudness: Reconsidered," J. Exp. Psychol. **86**, 374–379.
Schneider, B. A., and Cohen, A. J. (**1997**). "Binaural additivity of loudness in children and adults," Percept. Psychophys. **59**, 655–664.
Shaw, E. A. G. (**1974**). "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," J. Acoust. Soc. Am. **56**, 1848–1861.
Shaw, W. A., Newman, E. B., and Hirsh, I. J. (**1947**). "The difference between monaural and binaural thresholds," J. Exp. Psychol. **37**, 229–242.
Sivian, L. J., and White, S. D. (**1933**). "On minimum audible sound fields," J. Acoust. Soc. Am. **4**, 288–321.
Sivonen, V. P., Minnaar, P., and Ellermeier, W. (**2005**). "Effect of direction on loudness in individual binaural synthesis," in Proceedings of the 118th Convention of the Audio Eng. Soc., Convention Paper No. 6512.
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.
Zacharov, N., Tuomi, O., and Lorho, G. (**2001**). "Auditory periphery, HRTF's and directional loudness perception," in Proceedings of the 110th Convention of the Audio Eng. Soc., Convention Paper No. 5315.
Zahorik, P., and Wightman, F. L. (**2001**). "Loudness constancy with varying sound source distance," Nat. Neurosci. **4**, 78–83.
Zwicker, E., and Zwicker, U. T. (**1991**). "Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution," J. Acoust. Soc. Am. **89**, 756–764.

Binaural detection of 500-Hz tones in broadband and in narrowband masking noise: Effects of signal/masker duration and forward masking fringes

Leslie R. Bernstein and Constantine Trahiotis
Department of Neuroscience and Department of Surgery (Otolaryngology), University of Connecticut Health Center, Farmington, Connecticut 06030

Richard L. Freyman
Department of Communication Disorders, University of Massachusetts, Amherst, Massachusetts 01003

(Received 6 September 2005; revised 20 February 2006; accepted 27 February 2006)

NoSπ thresholds for a 500-Hz tonal signal were measured with broadband and with narrowband maskers using a single-interval adaptive matrix procedure [C. Kaernbach, J Acoust. Soc. Am. **88**, 2645–2655 (1990)]. The purpose of the study was to investigate and to account for the effects on thresholds of varying the durations of the signals and maskers and the durations of forward masking fringes that preceded the occurrence of signal-plus-noise. For detection in both broadband and narrowband noise, the addition of brief forward fringes of masking noise resulted in elevations in threshold for the shortest signal durations. Longer forward fringes led to larger decreases in threshold when the masker was broadband as compared to when the masker was narrowband. The complex patterning of the data was explained by the operation of: (1) "predetection" temporal integration associated with peripheral auditory filtering; (2) duration-dependent, across-frequency influences that differentially affect broadband and narrowband NoSπ thresholds, (3) "post-detection" temporal integration associated with the central binaural mechanism, and (4) consideration of the detection thresholds in terms of changes in interaural correlation rather than in terms of signal level or signal-to-noise ratio, *per se*. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2188373]

PACS number(s): 43.66.Pn, 43.66.Dc, 43.66.Ba [AK] Pages: 2981–2993

I. INTRODUCTION

In 1966, McFadden reported two important findings concerning the detection of a tonal signal masked by broadband noise when the stimuli were presented in the NoSπ configuration. The first finding was that thresholds of detection were elevated when the masking noise was pulsed on and off simultaneously with the signal, as compared to when the masking noise was presented continuously. The second finding was that thresholds lower than those measured when the masking noise was pulsed on and off simultaneously were obtained when the masking noise had a "forward fringe" that began prior to the presentation of the otherwise pulsed signal and masker. When the duration of the forward fringe of masking noise reached 600 ms, thresholds were essentially equal to those measured with continuous masking noise. We refer to the first finding as the "pulsed effect" and the second finding as the "forward fringe effect." McFadden's general results have been replicated and extended in several studies designed to reveal whether and to what degree each of the effects depends upon parameters of the signals and maskers, including their duration, spectrum, and interaural configuration (e.g., Robinson and Trahiotis, 1972; Gilkey *et al.*, 1990; Yost, 1985). The studies cited above also demonstrated that neither the pulsed effect nor the forward fringe effect was observed when both maskers and signals are presented diotically in the NoSo configuration. That is, the two effects were only observed when detection was based on binaural cues.[1]

The beneficial effects of forward fringes of masking noise on binaural signal detection notwithstanding, other studies have indicated that forward and/or backward temporal fringes of masking noise generally hamper performance. Those studies have reported that preceding and/or trailing diotic noise led to *increased* thresholds for discrimination of interaural temporal disparities (ITDs) and interaural intensitive disparities (IIDs) (e.g., Gilkey *et al.*, 1990; Bernstein *et al.*, 2001; Akeroyd and Bernstein, 2001). The results of these studies were explained by assuming that the listener effectively averages binaural information via a temporal window that exceeds the duration of the brief target. Such averaging of the target-based binaural information with the masker-based binaural information would effectively dilute the binaural information conveyed by the target, leading to increases in threshold. Similarly, assuming the operation of such a temporal window, one might expect that adding a forward fringe of diotic noise to a NoSπ stimulus would render the signal *less* detectable.

An experimental condition employed by Yost (1985) is especially pertinent. Yost inserted a temporal gap between the offset of a 500-ms-long forward fringe and the onset of a NoSπ signal plus masker. He found that progressively increasing the temporal gap progressively increased thresholds. This outcome seems *opposite* to the decreases in threshold that would be expected if the listener processed the binaural information through a temporal window. One might expect that increasing the temporal gap would result in there being

included *lesser* amounts of diotic noise within the temporal window and, therefore, *less* dilution of the signal-plus-noise-based binaural cues. This would, in turn, lead to *lower* thresholds of detection.

The purpose of this study was to investigate these issues. To that end, NoSπ thresholds were measured with both broadband and narrowband maskers while varying independently the durations of the signals and maskers and the durations of forward masking fringes. The inclusion of narrowband masking conditions was considered important because two prior "pilot studies" indicated that the forward fringe effect was minimal or absent when the masker was a narrowband of noise centered on the frequency of the signal. The durations of the stimuli were varied in a manner similar to, but with a finer "grain," than that employed by Robinson and Trahiotis (1972). Robinson and Trahiotis (1972) always gated their masking noise and signal such that the masking noise was at steady-state level during the 5-ms rise/decay of the tonal signal. Thus, even in their "no fringe" condition, there actually was 5 ms of masking noise preceding and following the steady-state portion of the signal. It occurred to us that such short "fringes" of diotic masking noise could, via an averaging process, dilute the binaural information conveyed by the NoSπ stimulus. In order to evaluate this hypothesis, it was deemed necessary to include measures of NoSπ detection thresholds in a true "no-fringe" condition, wherein signals and maskers are gated on and off simultaneously with the same rise/decay times.

It will be seen that the complex patterning of the data can be explained by the operation of: (1) a "predetection" temporal integration associated with peripheral auditory filtering; (2) duration-dependent, across-frequency influences that differentially affect broadband and narrowband NoSπ thresholds, and (3) a separate "post-detection" temporal integration associated with the central binaural mechanism.

II. EXPERIMENT 1

A. Procedure

Binaural (NoSπ) detection thresholds were measured for 500-Hz tonal signals presented against a background of broadband noise (bandwidth 100–3000 Hz) or narrowband noise (450–550 Hz). In both cases, the spectrum level of the noise was equivalent to 45 dB SPL. Data were obtained using a single-interval, "yes/no" task. This type of task was chosen for two reasons. First, as shown by McFadden (1966), the pulsed effect is larger when the task is single interval than it is when a two-interval, forced-choice task is employed. Second, studies following McFadden's that investigated the pulsed effect and the forward fringe effect and to which we wished to compare our data (e.g., Robinson and Trahiotis, 1972; Yost, 1985) also employed single-interval procedures.

In order to obtain thresholds more efficiently than could be done using "fixed" conditions (i.e., a method of constant stimuli), data were collected using the single-interval adaptive matrix (SIAM) procedure described and evaluated by Kaernbach (1990). Each trial consisted of a 500-ms warning interval and a single observation interval during which a visual display on a computer monitor marked the potential occurrence of the signal. Noise alone trials and signal-plus-noise trials were presented randomly with equal *a priori* probability. Correct-answer feedback was provided via the computer monitor for 400 ms after the listener responded followed by a 400-ms pause before the beginning of the next trial. The procedure was used to target 75% correct performance. As discussed by Kaernbach (1990, p. 2649), this required the level of the signal to be reduced by one step-size unit after each "hit," increased by one step-size unit after each "miss," increased by two step-size units after each "false alarm," and left unchanged after each "correct rejection." In order to minimize the errors of estimation of threshold, the step size was initially 2 dB and was reduced to 1 dB after the first two "reversals," a run was terminated after 16 reversals, and threshold was defined as the average level of the signal across the last 12 reversals (see Kaernbach 1990, Fig. 6, p. 2651).

Prior to the collection of formal data, detection thresholds yielded by the SIAM procedure were "validated" by comparing them to thresholds obtained with a single-interval fixed procedure in which a maximum probability correct $[p(c)_{\max}]$ of 0.75 was derived from the psychometric function relating 10 log d' to 10 log E/No (see Egan, 1965). Four young adults, two of whom were to participate in the main experiments, served as listeners. Because the two sets of thresholds proved to be, for all practical purposes, equivalent, all of the data to be reported here were collected using the more efficient SIAM procedure.

When the broadband maskers were employed, signals and maskers were each gated on and off using 5-ms $\cos^2$ ramps. The total durations of the tonal signals (including the ramps) were 10, 20, 40, 125, or 500 ms. When the narrowband maskers were employed, signals and maskers were each gated on and off using 10-ms $\cos^2$ ramps and the total durations of the tonal signals were 20, 40, 125, or 500 ms. The on/off ramp durations were increased to 10-ms $\cos^2$ when narrowband maskers were employed in order to effectively confine the spectrum of the stimuli to their nominal bandwidth of 100 Hz and thereby guard against detection of the signal on the basis of spectral "splatter" (e.g., Leshowitz and Wightman, 1971; Wightman, 1971). The spectral content of the forward fringe matched that of the masking noise in both the broadband and the narrowband conditions.

In the pulsed conditions, noise and signals were gated on and off simultaneously. In the "forward fringe" conditions, the onset of the noise preceded the onset of the signal by 2.5, 5, 10, 15, 30, 60, 125, 250, or 500 ms and the offset of the noise always occurred 5 ms after the offset of the signal. That is, in all of the forward-fringe conditions, there was also a 5-ms "backward" fringe. A condition in which the signal was masked by a continuous masker was not included as a reference condition (as has been the case in earlier studies). This decision was made for two reasons. First, the major focus of the study was to reveal and understand forward fringe effects on detection thresholds. Therefore, it was deemed necessary to measure those effects while holding constant and keeping brief, the duration of any "backward" fringe of masking noise. Second, the employment of 500-ms-

long forward fringe conditions virtually assured performance very close or equivalent to that measured in continuous noise (e.g., McFadden, 1966; Robinson and Trahiotis, 1972; Yost, 1985). All stimuli were generated digitally with a sampling rate of 20 kHz via a custom software library (MLSIG) running within MATLAB©, were converted to analog form via a TDT PD1, and were low-passed filtered at 8.5 kHz (TDT FLT1) before being presented via matched TDH-39 earphones.

The noise and signal waveforms were each selected randomly from within their own two-second-long buffers that were generated anew prior to each adaptive run. The relative levels of the stimuli and their rise-decay ramps were controlled via software and the absolute levels of the stimuli were determined by programmable attenuators (TDT PA4) in a manner that maximized the use of the 16-bit range of the digital-to-analog converters.

Four young adults (one male and three female) with audiometrically normal hearing served as listeners. Thresholds were obtained with broadband maskers before they were obtained with the narrowband maskers. Because one of the original female listeners was unavailable to participate in the experiment after the broadband masking thresholds were obtained, she was replaced with another female listener who only participated in the narrowband stimulus conditions. Thresholds were measured using blocks of conditions for which the duration of the tonal signal was held constant. When the broadband masker was employed, the durations were visited in the order of 20, 125, 500, 10, and 40 ms and the durations of the forward fringe employed within each block were chosen pseudorandomly. At the beginning of the testing for each duration, two thresholds were obtained in the pulsed (i.e., no fringe) condition and then one threshold was obtained for each forward fringe condition until all fringe durations were exhausted. Then, the forward fringe conditions were revisited in reverse order. Overall, this yielded two estimates of threshold for the pulsed and for each forward fringe condition. The same sequence of testing was repeated twice more. In this manner, a total of six estimates of threshold was obtained for the pulsed and for each forward fringe condition. The same scheme was used for the collection of thresholds when the narrowband masker was employed. For that set of thresholds the durations were visited in the order 20, 125, 40, and 500 ms.

Thresholds were calculated by first omitting the lowest and highest of the six estimates of threshold for each stimulus condition and then computing the mean of the remaining four. For about 9% of the 360 estimates of threshold (90 stimulus conditions $\times$ four listeners), the standard deviation of the four estimates exceeded 3 dB. In those instances, additional measures of threshold were obtained substituting the new measures one-for-one with the "oldest" measures until the four estimates of threshold yielded a standard deviation of less than 3 dB. This typically required the measurement of two additional thresholds. The data were scrutinized to determine whether "order effects" were present and none was found.

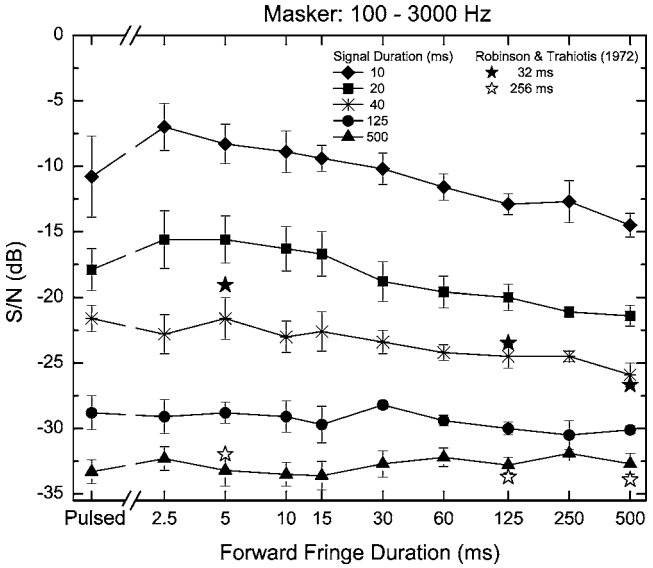


FIG. 1. NoSπ threshold signal-to-noise ratio (S/N) in decibels plotted as a function of forward fringe duration. The data represent the thresholds averaged across four listeners when the masker was broadband. Error bars represent ±1 standard error of the mean. The parameter of the plot is the duration of the signal. Data plotted as stars represent thresholds obtained by Robinson and Trahiotis (1972), after adjustment as described in the text.

B. Results

Figure 1 displays thresholds of NoSπ detection obtained with the broadband masker and averaged across the four listeners. Thresholds are reported in terms of signal-to-noise ratio (S/N) (the ratio of the total power of the signal to the total power of the noise, in decibels), calculated across the 2900-Hz bandwidth of the masker, as a function of the duration of the forward fringe of the masking noise. The error bars represent ± one standard error of the mean. The reader is reminded that all of the forward fringe conditions also had a "backward" temporal fringe of 5 ms and the pulsed condition is the one in which the signal and noise were gated on and off simultaneously. Thresholds obtained in the pulsed condition, in which there was neither a forward fringe nor a backward fringe, are plotted to the left of the break in the abscissa. The parameter of the plot is the duration of the signal.

For purposes of comparison, the thresholds obtained by Robinson and Trahiotis (1972) with signal durations of 32 and 256 ms are plotted, respectively, as closed and open stars. The Robinson and Trahiotis thresholds, which were originally reported in terms of $10\log(E/No)$, were transformed to S/N (dB) and reduced by 1 dB in order to account for the fact that their "threshold" level represented 80% correct while, in the current study, the SIAM procedure targeted 75% correct. The adjustment of 1 dB was based on the fact that the linear portion of the psychometric functions fit by Robinson and Trahiotis, like those commonly found in other studies of binaural detection (e.g., Egan, 1965) had a slope of about 5% per dB (Trahiotis, 2005). Note that the positioning and trends of the thresholds obtained by Robinson and Trahiotis (1972) are consistent with those obtained in the current study. This confirms the validity of the SIAM procedure found in our pilot studies because Robinson and Trahiotis employed "fixed" stimulus conditions.

The new data in Fig. 1 were subjected to a two-factor within-subjects repeated-measures analysis of variance [signal duration × forward fringe duration]. The error terms for the main effects and for the interaction were the interaction of the particular main effect (or the interaction) with the subject "factor" (Keppel, 1973). In addition to testing for significant effects, the proportions of variance accounted for (ω^2) were determined for each significant main effect and interaction (Hays, 1973). The main effect of signal duration was significant [$F(4,12)=707.02$, $p<0.001$] and accounted for 90% of the variance in the data. The main effect of forward fringe duration was also significant [$F(9,27)=7.50$, $p<0.001$] and accounted for 2% of the variance in the data. The interaction of signal duration and forward fringe duration was significant [$F(36,108)=4.53$, $p<0.001$] and accounted for 2% of the variance in the data. Overall, then, the physical parameters of the stimuli and their interaction accounted for 94% of the variance in the data.

Visual comparisons among the data reveal three important trends. First, when the signal duration was 10 or 20 ms, adding a short forward fringe increased thresholds by as much as 3–4 dB, as compared to those measured in the pulsed condition. This is the type of outcome that was hypothesized to occur if the listeners were monitoring the stimuli via a short temporal window that served to dilute the binaural cues produced by the addition of the Sπ signal to the masker.

Second, for signal durations below 125 ms, thresholds declined as the duration of the forward fringe was increased from 2.5 to 500 ms. In addition, the magnitudes of the declines in threshold were inversely proportional to signal duration. These trends in the data, taken together, are what led to the significance of the main effect of forward fringe duration and the significant interaction between signal duration and forward fringe duration. For the four shortest signal durations, the longest forward fringes yielded thresholds that were lower than those obtained in their corresponding pulsed conditions, thus replicating the fringe effects originally reported by McFadden (1966) and Robinson and Trahiotis (1972). An appreciation for the potency of the forward fringe in reducing masking can be gained by focusing on the data obtained when the duration of the signal was either 10 or 20 ms. For those short signals, extending the duration of the forward fringe beyond 30 ms resulted in *increases* in sensitivity that more than compensated for *decreases* in sensitivity attributable to the operation of a temporal window.

The third important trend in the data is the well-ordered, vertical separation of thresholds between differing signal durations. This is the significant main effect that accounted for 90% of the variability in the data. Further discussion and interpretation of this important aspect of the data is postponed until the presentation of the data obtained with narrowband maskers.

Figure 2 displays thresholds of NoSπ detection obtained with the narrowband masker, plotted in terms of the S/N (in decibels) calculated across the 100-Hz bandwidth of the masker. These thresholds were subjected to a two-factor within-subjects repeated-measures analysis of variance [signal duration × forward fringe duration]. Once again, the

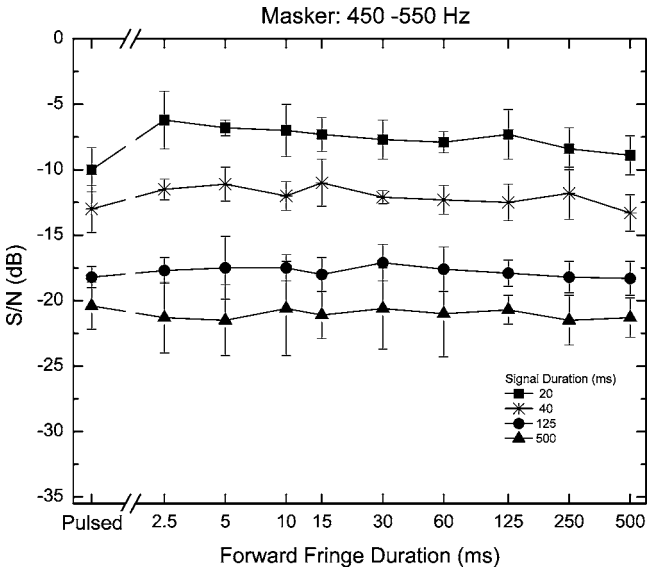


FIG. 2. Similar to Fig. 1, but for NoSπ thresholds obtained with the narrowband masker.

error terms for the main effects and for the interaction were the interaction of the particular main effect (or the interaction) with the subject "factor" (Keppel, 1973). The main effect of signal duration was significant [$F(3,9)=122.44$, $p<0.001$] and accounted for 90% of the variance in the data. The main effect of forward fringe duration was *not significant* [$F(9,27)=2.06$, $p=0.07$] and accounted for less than 1% of the variance in the data. The interaction of signal duration and forward fringe approached but did not reach significance [$F(27,81)=1.59$, $p=0.06$] and accounted for about 1% of the variance in the data. It was necessary to evaluate whether this lack of significance of the main effect of fringe duration and the lack of significance of the interaction of signal duration with fringe duration resulted from the omission of the 10-ms signal duration in the narrowband conditions. In order to do so, we reran the analysis of variance (ANOVA) on the data obtained with broadband maskers (Fig. 1) while omitting the 10-ms signal duration. Importantly, as with the original analysis, both main effects and their interaction reached significance ($p<0.001$) and, overall, about 90% of the variance in the data was attributable to these effects. Thus, the differences in the statistical outcomes between broadband and narrowband conditions appear to reflect actual differences in the patterning of the data.

Visual comparisons among the data in Fig. 2 and between the data in Figs. 2 and 1 parallel the statistical findings. First, for the 20-ms signal masked by narrowband noise, adding a short forward fringe increased thresholds by about 3 dB as compared to those measured in the pulsed condition, very much like what was found when the maskers were broadband. This is consistent with the listeners' monitoring of the stimuli via a temporal window. Second, in contrast to what was found with the broadband maskers, thresholds with narrowband maskers decreased much less as the duration of the forward fringe was increased from 2.5 to 500 ms. Thus, based on visual inspection and statistical analyses, the large forward fringe effects found here and in the literature with broadband maskers appear to be attenu-

ated or absent when the masker is narrowband (i.e.,100-Hz wide). On the other hand, the data in Fig. 2, are like those in Fig. 1 in that there is a well-ordered vertical separation of thresholds between differing signal durations.

C. Discussion and interpretations of the data

As discussed above, there are three major trends in the data: (1) effects logically attributable to the operation of a temporal window; (2) effects stemming from the presence of a forward fringe of masking noise; and (3) effects resulting from variations of signal duration. Each of these will be discussed in turn.

1. A temporal window

The data obtained with both broadband and narrowband maskers reveal that the addition of a short forward fringe of noise increases thresholds compared to those measured in the pulsed condition by amounts that are inversely proportional to the duration of the signal. Specifically, the largest increases occurred for 10-ms-long signals, a slightly smaller increase occurred for the 20-ms signals, and no increase was observed for signal durations of 40 ms and longer. Such an outcome is consistent with the assumption that the listeners process the stimuli via a short temporal window that dilutes the binaural cues produced by the addition of the $S\pi$ signal to the masker.

An attempt was made to formalize this notion in a quantitative, parsimonious, and theoretically "neutral" manner. This was done by investigating whether the relevant increases in threshold could be accounted for by considering any "temporal integration" that would occur as a consequence of the stimuli being processed via peripheral auditory filtering. That is, the temporal integration was, parsimoniously, assumed to be "predetection" and to occur within each monaural channel as a necessary consequence of peripheral filtering. A similar approach was used by Hartung and Trahiotis (2001).

The temporal integration was accomplished by passing the stimuli through a gammatone filter centered at 500 Hz and having a 3-dB down bandwidth of about 75 Hz (for details concerning the validity and implementation of such a filter, see, e.g., Glasberg and Moore, 1990; Slaney, 1993; Patterson *et al.*, 1995; Hartung and Trahiotis, 2001). The impulse response of the filter and its envelope are depicted in Fig. 3. Note that the maximum of the response is reached at about 6 ms and that the response falls to essentially zero at 25–30 ms.

In order to form a decision variable that would yield predictions of threshold, functions were generated that related input S/N to output $\Delta\rho$ (i.e., the change in interaural correlation of the waveforms produced by the addition of the $S\pi$ signal as measured at the output of the gammatone filter). This was done by averaging 100 estimates at each of several behaviorally relevant values of S/N for each of the 50 broadband stimulus conditions (five signal durations $\times$ 10 forward fringe durations) and each of the 40 narrowband stimulus conditions (four signal durations $\times$ 10 forward fringe durations). Each of the measures was made by evaluating the interaural correlation of the signal-plus-noise waveform utilizing the 30-ms epoch of the gammatone response that produced the maximal decrease in interaural correlation (from the "masker" interaural correlation of 1.0). This enabled the predictions to include any advantages in detection that would be gained by "off-time" listening.

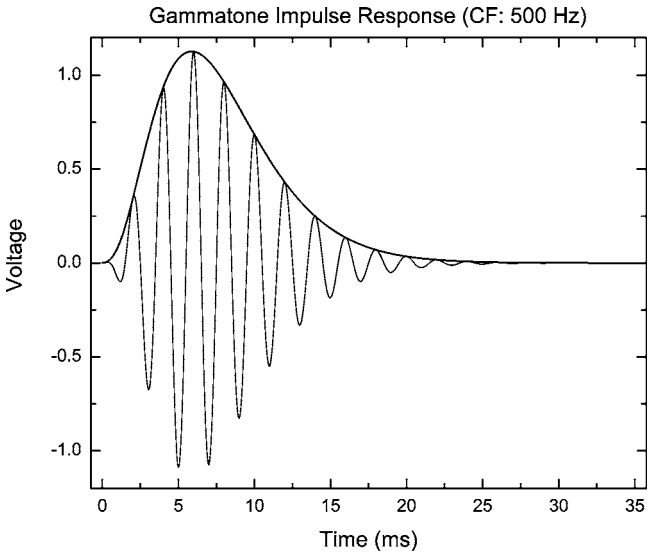


FIG. 3. The impulse response of the gammatone filter centered at 500 Hz. The solid line atop the peaks in the fine structure is the envelope of the response.

For each stimulus configuration, the paired values of S/N and $\Delta\rho$ were fit by either polynomial (broadband stimuli) or rational (narrowband stimuli) functions. This enabled accurate interpolations to be made. For each signal duration, the value of $\Delta\rho$ corresponding to the value of S/N at threshold in the pulsed condition was determined via interpolation using separate polynomial functions relating S/N and $\Delta\rho$. That value of $\Delta\rho$ was taken to be the criterion for threshold and its corresponding S/N was used as the predicted threshold for each of the forward fringe conditions, respectively.

Figure 4 contains the predictions (dotted lines) and data when the masker was broadband.[2] The predictions capture whether and to what degree increases in threshold, over those obtained in the pulsed conditions, occurred when short forward fringes of either 2.5 or 5 ms were added along with a 5-ms backward fringe of masker. The reader is reminded that the largest increase in predicted and obtained thresholds occurred when the duration of the signal was 10 ms; a somewhat smaller increase occurred when the duration of the signal was 20 ms. No increases in threshold occurred for durations of the signal equal to or greater than 40 ms. This is understandable in that signal durations of 40 ms and longer exceed the 30 ms or so duration of the impulse response of the gammatone filter centered at 500 Hz. Thus, for the longer signal durations, adding a forward and/or backward fringe of noise to the pulsed signal-plus-masker would not be expected to dilute the binaural cues and thereby affect thresholds. By the same token, the predictions, by their very nature, asymptote to the threshold predicted at the point at which the temporal window is "filled" by the signal-plus-masker-plus-fringe waveforms. Therefore, the *decreases* in

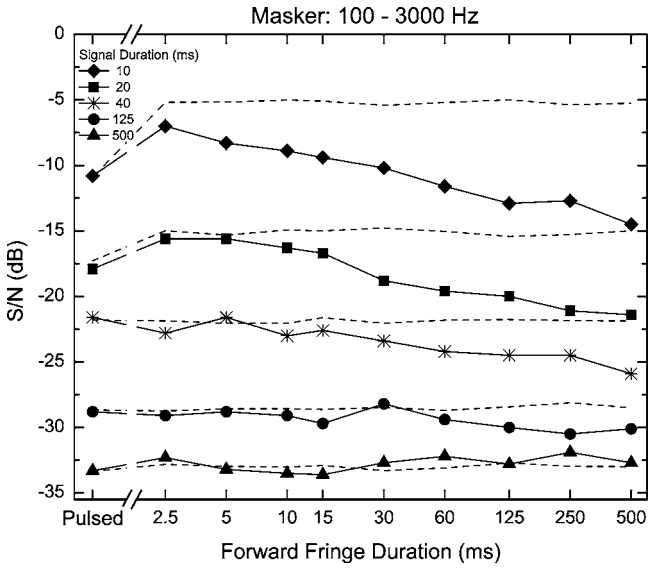


FIG. 4. Thresholds obtained with the broadband masker replotted from Fig. 1. The dashed lines represent predicted thresholds based on temporal integration arising from the peripheral auditory filter centered on 500 Hz (see the text).

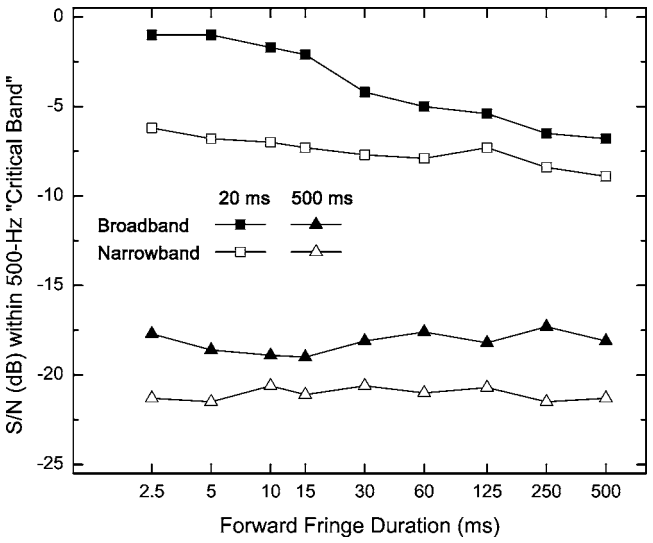


FIG. 6. Thresholds obtained with signals having durations of either 20-ms (squares) or 500-ms (triangles) in both broadband (closed symbols) and narrowband (open symbols) masking conditions plotted in terms of the S/N within the critical band. The dashed line represents predicted thresholds for the 20-ms signal masked by broadband noise. The predictions are based on a model that incorporates changes in the variation in the activity across auditory filters produced by variations in the duration of the forward fringe (see the text).

threshold that define McFadden's (1966) forward fringe effect appear to be an independent phenomenon that coexists with the increases in thresholds that arise from peripheral temporal integration. Thus, for broadband stimuli, the decreases in $N_0S\pi$ threshold produced by the addition of forward fringes of masking noise do not preclude or contradict the notion that listeners integrate binaural information via a temporal window.

The predictions and data for the narrowband conditions are shown in Fig. 5. Once again, the predictions capture quite well whether and to what degree the addition of forward and backward fringes produce increases in threshold over and above those obtained in the pulsed conditions. Note that a slight increase in threshold was observed and predicted for the 40-ms signal when the masker was narrowband. Recall that such an increase was neither observed nor predicted for this signal duration when the masker was broadband. The difference results from the fact that a rise/decay time of 10 ms was employed with the 100-Hz-wide narrowband masker and a rise/decay time of 5 ms was employed with the 2900-Hz-wide broadband masker. This resulted in the 40-ms pulsed signal-plus masker having a plateau of 20 ms when the masker was narrowband and a plateau of 30 ms when the masker was broadband. Consequently, there was a bit of "room" within the approximately 30-ms-long temporal window for the addition of a masking fringe to dilute the binaural cues. To us, considering the types and amounts of within- and across-listener variability that can occur in psychophysical experiments, it is remarkable that such subtle theoretically-expected effects can be measured empirically.

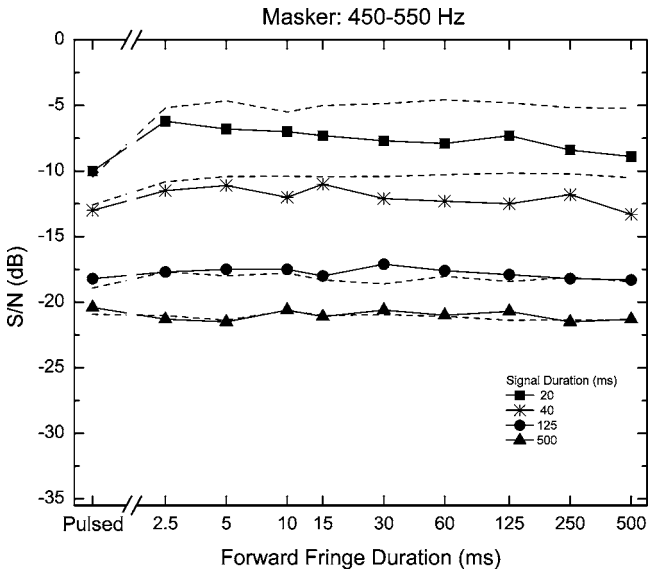


FIG. 5. Similar to Fig. 4, but for thresholds obtained with the narrowband masker.

2. The forward fringe effect

The addition of longer and longer forward fringes produced substantially greater decreases in threshold in broadband conditions, as compared to narrowband conditions. This outcome was expected on the basis of pilot studies and is especially interesting. It suggests that merely having diotic masker stimulation available for comparison to subsequent signal-plus-masker stimulation is not, in and of itself, sufficient to produce the relatively large decreases in threshold that define the forward fringe effect with broadband maskers. Therefore, explanations of the forward fringe effect wherein the fringe provides a stable diotic referent against which the addition of the signal is compared (e.g., McFadden, 1966; Robinson and Trahiotis, 1972; Yost, 1985), cannot, without further qualification, account for the data.

Figure 6 contains a subset of the data presented in a manner that helps foster an understanding of what aspects of the stimuli and their processing by the auditory system result

in forward fringe effects for broadband but not for narrowband masking conditions. The thresholds depicted are those obtained with signals having durations of either 20 ms (squares) or 500 ms (triangles) in both broadband (closed symbols) and narrowband (open symbols) masking conditions and when forward fringes were present.

In order to make direct and meaningful comparisons among the thresholds of detection obtained in those four conditions, the data are plotted in terms of the S/N in decibels within an assumed 100-Hz-wide "critical band" filter centered at 500 Hz. To this end, the threshold S/Ns obtained with the broadband masker were increased by 14.6 dB, the value representing the ratio between the 2900-Hz bandwidth of the broadband masking noise and the 100-Hz-wide critical band. The threshold S/Ns obtained with the narrowband masker required no adjustment because that masker was 100-Hz wide.

Beginning with the data obtained with the 500-ms forward fringe, note that, for both signal durations, higher values of S/N (about 2–3 dB) were required to reach threshold in the broadband conditions. This outcome is consistent with a number of previous studies indicating that detection of a $S\pi$ signal against diotic masking noise is adversely affected by noise that is present in the approximately 400-Hz-wide spectral regions surrounding the (monaural) critical band (e.g., Bourbon, 1966; Hall *et al.*, 1983; Kohlrausch, 1988; van der Heijden and Trahiotis, 1998; van der Par and Kohlrausch, 1999). This finding was, for decades, often interpreted to mean that binaural processing was constrained by there being a "binaural critical band" that was analogous to, but wider than, the monaural critical band. As discussed in the three most recent of the studies cited above, the current view is that binaural processing is constrained by the same peripheral filtering as is monaural processing, but, depending upon the spectral content of the masker, binaural processing is also influenced by a wider operational bandwidth that includes information within auditory filters adjacent to and centered on the frequency of the signal.

Overall, the figure shows that thresholds obtained with the broadband masker are always larger than their narrowband counterparts and, for 20-ms signals, the differences between the thresholds obtained with broadband versus narrowband maskers become greater and *exceed* the 2–3 dB differences attributable to the wider "operational band" as the duration of the forward fringe is reduced. These outcomes may provide important clues concerning the interpretation of the forward fringe effect. They suggest that what requires explanation is the *degradation* of detectability in broadband masking conditions produced by *shortening* the forward fringe, rather than any *enhancement* in detectability produced by *lengthening* the fringe.

Before describing our thoughts concerning what may underlie the forward fringe effect, it is helpful to first summarize van de Par and Kohlrausch's (1999) explanation of why the binaural operational bandwidth appears to be broader than the monaural critical bandwidth (auditory filter). According to van de Par and Kohlrausch (see pp. 1944–1945), in $NoS\pi$ detection conditions employing relatively high-level *narrowband* maskers, listeners maximize performance by optimally combining useful binaural information from both the auditory filter centered on the signal frequency and from adjacent auditory filters near enough in frequency to contain appreciable signal energy. An underlying assumption is that the internal noise within each of the relevant filters is independent, and, therefore, combining the binaural information across filters improves sensitivity relative to that possible using only the binaural information present in the filter centered on the signal.

In the case of a broadband masker, however, little or no useful binaural information would be present in the adjacent filters because the high levels of external diotic masking noise result in vanishingly small S/Ns in those channels. In this case, optimizing performance would entail essentially ignoring the binaural information carried by the adjacent filters while "attending" only to the binaural information carried in the on-frequency filter. Therefore, the type of benefit that stems from combining information across filters in the case of the narrowband masker is lost when the masker is broadband and the signal level required for threshold is increased.

van de Par and Kohlrausch's (1999) ideas appear to provide a framework for understanding why shortening the forward fringe degrades $NoS\pi$ detectability for short signals masked by broadband noise. Perhaps shortening the forward fringe either precludes or hampers listeners' abilities to attend primarily to the auditory filter centered on the frequency of the signal. In order to illustrate how this might come about, Fig. 7 displays the average powers at the outputs of gammatone filters for our broadband masker. The top left panel shows the outputs for each of three independent 20-ms samples of the broadband masker for auditory filters having center frequencies spanning the range 100 to 4000 Hz. The three curves reveal large variations in output both across filters and across 20-ms samples of noise. The large variations in powers at the outputs of the filters are a manifestation of the independent, Rayleigh-distributed, variations in the magnitudes of the input that occur within differing spectral regions for short samples of broadband Gaussian noise. The top right panel shows the same filter outputs, but only for filters spanning the binaural operational bandwidth of 300 to 700 Hz.

The bottom row of Fig. 7 depicts corresponding plots for 125-ms-long samples of a broadband diotic masker. Note that both within-plot and across-plot variability in the activity of the filters is substantially reduced, as compared to when the maskers were only 20-ms long. Thus, the binaural information provided by the 125-ms-long broadband forward fringe could lead to relatively increased detectability because there would be little filter-to-filter variability that would serve to hamper the listener's ability to detect the addition of the $S\pi$ signal.

In order to assess the viability of this line of reasoning, we attempted to relate the expected amounts of duration-dependent, across-frequency variability in the outputs of auditory filters to changes in threshold produced by variations in the duration of forward fringes. Measures of filter-by-filter variability were made in the following manner. The first step was to measure a "reference" output of a bank of gammatone

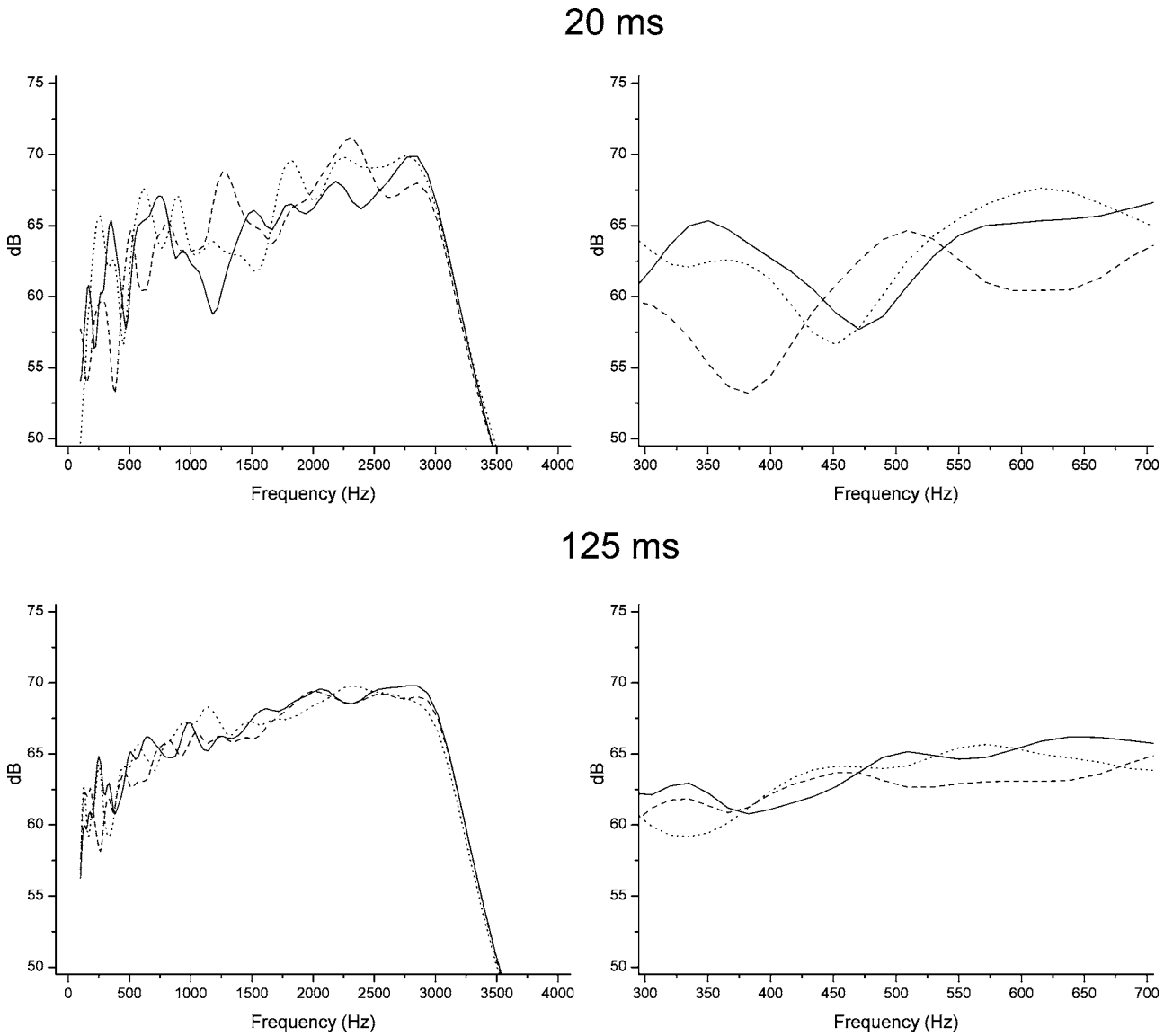


FIG. 7. Average powers at the outputs of gammatone filters for the broadband masker. Top row: outputs for each of three independent 20-ms samples of the broadband masker for filters spanning the range 100 to 4000 Hz (left) or for filters spanning the binaural operational bandwidth of 300 to 700 Hz (right). Bottom row: same as top row but for 125-ms-long tokens of broadband noise.

filters having center frequencies spanning the range 300–700 Hz, the operational binaural bandwidth discussed earlier. By "reference" output, we mean the output that would occur for an essentially steady-state input of our broadband Gaussian noise. This was done by averaging the outputs across 100 independent, 30-s-long samples of the noise. This produced a smoothly rising output of the filters as a function of frequency, commensurate with their increasing values of bandwidth. Then, the outputs of the filters were measured for each of the nine durations of the forward fringe. To each of the nine forward-fringe durations, 15, 25, 45, or 130 ms of the masking noise was appended in order to match the total durations of the maskers employed in the behavioral experiments when the signal was 10, 20, 40, or 125 ms long, respectively (i.e., our simulations included the 5-ms backward fringe that accompanied all of the forward fringe conditions). Two thousand five hundred independent tokens of diotic broadband noise were used as inputs to the filter bank and were constructed anew for each of the 36 different total durations of the masker. For each such token, the decibel difference in the outputs was computed, filter by filter, from the reference output. Then, the variance of the deviations from the reference output was computed. For each duration of the masker, the estimate of the across-filter variance was taken to be the average of the variance computed across the 2500 tokens.

The estimates of the relative variance of the decibel differences measured across filters are plotted as the dashed lines in Fig. 8 and are indexed by separate right-hand axes for each of the four signal durations. The vertical position of each line was adjusted to provide a reasonable visual fit to the detection thresholds obtained at each signal duration. The changes in the behavioral thresholds appear, line by line, to be closely paralleled by the changes in the variance of the outputs of the 300–700 Hz-wide gammatone filter bank.

After extensive deliberation, it occurred to us that several factors could underlie or be related to *how* behavioral thresholds are affected by changes in across-filter variability

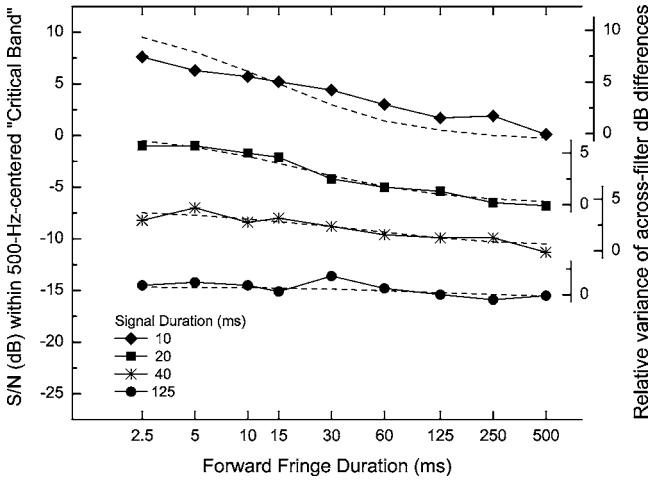


FIG. 8. Thresholds obtained with signals having durations of 10, 20, 40, or 125 ms in broadband masking conditions plotted in terms of the S/N within the critical band. The dashed lines represent estimates of the relative variance of the decibels differences measured across filters and are indexed by separate right-hand axes for each of the four signal durations.

that, in turn, stem from changes in the duration of the forward fringe. For example, we considered the effects that across-filter variations could produce when combined with the presence of a slightly positively interaurally correlated internal noise such as that typically used to explain how NoSπ thresholds are affected by the level of a diotic masker (e.g., Diercks and Jeffress, 1962; McFadden, 1968; Yost, 1988). Those papers illustrate how internal noise serves to decorrelate the perfectly interaurally correlated external masker.

In the face of across-filter variations in power that occur in response to short-duration broadband maskers, the amount of decorrelation produced by internal noise would vary across filters, both within and between brief samples of the masking noise. This is so because the decorrelation within each auditory filter would be inversely proportional to the amount of external masker energy that was momentarily present. As a result, from one brief sample of the masker to the next, the listener might have difficulty deciding which particular auditory filter contained the additional change in interaural correlation produced by the addition of the Sπ signal. Said differently, and a la van de Par and Kohlrausch (1999), performance with short signals masked by short broadband noises (including either brief or no forward fringe) could be hindered by an inability to discern and/or to attend to the auditory filter centered on the signal. Within the same rubric, and consistent with the behavioral data (Fig. 5), one would not expect strong forward fringe effects to be present when the masker is narrowband. This is so because the type of across-filter variability that occurs for broadband maskers would be lessened or would not occur, depending on the actual bandwidth of the masker.

In order for this approach to be plausible, however, psychoacoustically relevant filter-by-filter variations in *internal* interaural correlation would have to occur when short-duration, external broadband diotic noise is combined with the aforementioned internal noise. In order to investigate this issue, computer simulations were run using a model based on interaural correlation that included stages of bandpass filtering, envelope compression, and rectification (for details, see Bernstein *et al.*, 1999). The external noises that served as inputs were generated in the same fashion (and had the same characteristics) as those employed in our experiments when the signal duration was either 10 or 20 ms and the forward fringe duration was 2.5 ms. Those stimulus conditions were two for which there were large forward fringe effects and large across-filter variations in power. The additive "internal" noise had the same spectral shape, level, and interaural correlation as that described by Yost (1988).

Normalized interaural correlations were computed filter by filter assuming an integration window of 100 ms (see Bernstein and Trahiotis, 1999). The magnitudes of the interaural correlations ranged, both across samples and across filters, from about 0.98 to 1.0, with the largest values of decorrelation occurring for auditory filters at the lower portions of the 300–700 Hz operational bandwidth, where the power of the internal noise was greatest. Such values of interaural correlation are not unlike the value of about 0.95, calculated via the same model, that corresponds to thresholds obtained in the pulsed NoSπ configuration when the signals and maskers are of long-duration (i.e, 250–500 ms). It is the case, however, that for signals and maskers of *short duration* (i.e., 10–20 ms), that are of greatest interest here, the interaural correlation corresponding to threshold is much lower, being about 0.0 to 0.4. Based on this analysis, it appears that the across-filter variations in power, when combined with internal noise estimates based on Yost's (1988) analyses cannot in and of themselves be responsible for the forward fringe effect. This outcome does not negate the potential import of the close correspondence between the measures of variations in power across filters and the behavioral thresholds measured as a function of forward fringe duration for signals having durations between 10 and 125 ms (Fig. 8). It remains for future research to determine what factor(s) or mechanism(s) operate in conjunction with the variations in power across filters to produce the changes in behavioral thresholds observed under NoSπ forward fringe conditions with short-duration signals.

The general theoretical framework described here helps to explain many of the outcomes Yost (1985) reported in his comprehensive study of how forward fringes affect the NoSπ detectability of a 500-Hz, 20-ms tone under broadband masking conditions. For example, Yost (1985) found that thresholds increased and rose to the value observed in the pulsed condition when the bandwidth of the 500-ms-long forward fringe (but not that of the broadband masker) was made progressively narrower from the "operational" bandwidth of 400- to a 500-Hz tone. That is, the forward fringe effect was either attenuated greatly or was absent for narrowband forward fringes. According to our schema, the more narrow the spectrum of the forward fringe the less the opportunity for reduction of the across-filter and across-token variability arising from the short-duration broadband masker.

Yost (1985) also found that reductions in the level of the forward fringe led to proportional reductions in the magnitude of the forward fringe effect. Now, assume that the listener integrates information in the forward fringe with infor-

mation provided during the last 20 ms of the masker (when the signal could be added). Reducing the level of the forward fringe would have the effect of producing relative *increases* in the salience of internal noise (that is slightly positively interaurally correlated). Integration of the internal noise with the diotic masker would result in the effective broadband masker having, filter by filter, a "reference" interaural correlation of less than 1.0. That, in turn, would lead to poorer detectability of the Sπ signal (e.g., Robinson and Jeffress, 1963; van der Heijden and Trahiotis, 1998).

Direct support for the notion that forward fringe effects stem from some type of integration of the information in the forward fringe with information provided by the signal-plus-masker waveform is also provided by Yost (1985). Yost separated the offset of a 500-ms forward fringe from the onset of the 20-ms pulsed signal-plus masker. Greater temporal separations led to proportionately greater attenuations of the forward fringe effect. Such temporal separations would reduce the effectiveness of the integration of binaural information in the masking fringe with that provided by the signal-plus noise, thereby leading to greater across-filter variability and poorer detectability of the Sπ signal. Thus, many of the data concerning forward fringe effects are consistent with our hypothesis linking forward fringe effects with across-filter variations in binaural activity.

At this time, details concerning the exact nature of the integration of binaural information in the forward fringe with the information occurring during the time the signal could be present are unknown. For example, one may conceive of the process as being one of "running" integration that begins with the onset of the forward fringe. The addition of the Sπ signal would produce a maximal change in the interaural correlation within the auditory filter centered on the signal during that portion of the running integration that encompasses the signal. Seen in this light, detection of the signal comes about via both across-filter and across-time changes in interaural correlation.

The type of integration described above would be expected to produce the same magnitude of fringe effects independent of whether the fringe preceded or followed the signal. It is the case, however, that backward masking fringes yield enhancements of detectability for short-duration Sπ signals masked by diotic broadband noise that are smaller than those observed for forward masking fringes (e.g., Trahiotis *et al.*, 1972; Zwicker and Zwicker, 1984; Gilkey *et al.*, 1990). Perhaps, the asymmetric fringe effects stem from differential "predetection" effects arising directly from the response of the peripheral auditory filters. In the case of forward fringes, the Sπ signal is added to the noise after the hair cell/auditory filter mechanism has reached a "steady-state" response. In the case of backward fringes, the Sπ signal is added to the noise at the onset of stimulation. The internal representation of the signal-plus masker in the two cases would be expected to differ as a result of well-documented onset-related non-linearities (adaptation) in the neural responses (e.g., Meddis, 1986; Dau *et al.*, 1996a,b; Kohlrausch and Fassel, 1997). Perhaps differential effects of adaptation underlie the differences observed for NoSπ detection with forward and backward masking fringes.

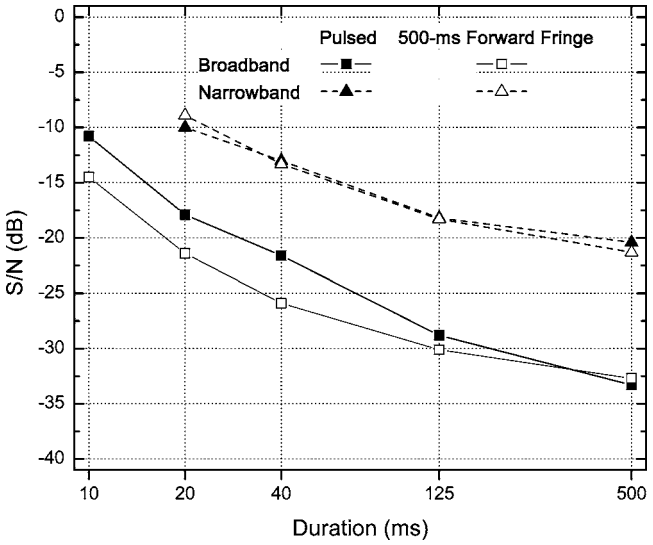


FIG. 9. Threshold signal-to-noise ratio (S/N) in decibels plotted as a function of the duration of the signal for the broadband (squares) and narrowband (triangles) masking conditions. Thresholds obtained with a 500-ms forward fringe are indicated by open symbols; thresholds obtained in the pulsed conditions are indicated by closed symbols.

3. Signal duration

The third important trend in the data that remains to be explained concerns the substantial increases in threshold found with decreasing signal durations in both broadband and narrowband masking conditions (Figs. 1 and 2). We begin with a consideration of the thresholds obtained with a 500-ms forward fringe. Those thresholds are replotted as open symbols in Fig. 9 with the abscissa being signal duration. Note that the thresholds obtained with broadband (squares) and narrowband (triangles) maskers are affected similarly by changes in duration. For both, thresholds decrease at a rate of approximately 4.5 dB per doubling of duration as duration is increased toward 125 ms and decrease at a much slower rate (about 1.5 dB per doubling) as duration is increased to 500 ms. These trends are consistent with several previous studies that are summarized and explained by Breebaart *et al.* (2001). They explain the slope of 4.5 dB per doubling as being a manifestation of two different processes. The first is the integration of binaural information within a fixed-duration "post-detection" temporal window. For durations less than the duration of the temporal window, thresholds would be expected to drop by 3 dB for each doubling of the duration of the signal. This is so because doubling the duration of the signal doubles the *effective* power of the signal as measured over the fixed (longer) duration of the "post-detection" integration. The second process is the relative reduction of an independent "decision noise" that comes about as a result of having the signal available for a longer time. Breebaart *et al.* argue that the second process produces an additional reduction in threshold of 1.5 dB per doubling which, when combined with the 3 dB per doubling reduction in threshold, yields an overall reduction of 4.5 dB per doubling for signal durations less than the duration of the "post-detection" temporal window.

Let us now consider the thresholds obtained with broadband and narrowband maskers when signals and maskers

were pulsed on and off with *no* temporal fringe (the points plotted at the far left of Figs. 1 and 2). Those thresholds are replotted as closed symbols in Fig. 9. Note that, for these "pulsed" conditions, thresholds obtained with broadband (squares) and narrowband (triangles) maskers are not affected similarly by changes in duration. Specifically, the narrowband data exhibit the same rates of change as were found with the 500-ms forward fringe. The broadband data, however, exhibit a steeper slope. That steeper slope is part and parcel of the differential elevations in thresholds at the shortest signal durations that define the forward fringe effect. Those differential elevations were explained as arising from across-filter variability. Taking all of this into account, we believe that the various sets of data collected across studies of the "pulsed" effect in broadband NoSπ masking conditions (e.g., McFadden, 1966; Robinson and Trahiotis, 1972; Yost, 1985) reflect several interacting processes: (1) temporal integration via peripheral auditory filtering; (2) duration and bandwidth-dependent variability in output across auditory filters; and (3) "post-detection" integration of binaural information via a fixed temporal window combined with reduction of noise in the decision process. The degree to which each of these processes affects measures of the pulsed effect in any particular study depends upon rise/decay and timing issues regarding the manner in which the signals and maskers are gated on and off. As an example, consider thresholds measured in studies wherein the broadband masker was gated so that it was at steady-state prior to and just after the occurrence of the Sπ tonal signal (e.g., Robinson and Trahiotis 1972). Those thresholds would reflect dilution of binaural cues via integration of short diotic masking fringes with the signal-plus-noise waveform while their counterparts measured with in true "no-fringe" conditions (such as those employed in this study) would not.

4. The integration of binaural cues

Our explanations of the changes in NoSπ thresholds entail the integration of binaural information across time. Either implicitly or explicitly, it has been assumed that a change in interaural correlation, or some measure isomorphic with it, serves as the quantity that is integrated. Historically, similar types of trading relations between the amount of stimulus power and stimulus duration required for detection have been found for both monaural (NoSo, NmSm) and binaural (NoSπ) thresholds (Blodgett *et al.*, 1958). Following Kohlrausch (1988), we too believe that it is important to make explicit that, in the case of binaural detection, it is not signal power nor energy, *per se*, that underlies performance but, rather, interaural disparity as quantified by interaural correlation. According to this point of view, the pulsed effect found under NoSπ detection of a 500-Hz tonal signal masked by diotic broadband noise should also be found if the "signal" were a narrow (i.e., critical-band-wide) change in interaural correlation. In such an experiment, there would be no added signal power to integrate over time.

We knew of no study in the literature that could provide the relevant data. While Jain *et al.* (1991) did measure correlation discrimination thresholds using a similar (spectrally) "fringed" paradigm, they did not vary duration. Consequently, we collected data in a supplementary experiment with four listeners, utilizing three listeners who had participated in the main experiment. Decreases in interaural correlation of a 100-Hz-wide band of noise centered at 500 Hz were detected against the same diotic masking noise as that employed in Experiment 1. The selection of 100 Hz as the bandwidth over which to decrease the interaural correlation was made in order to provide changes in the interaural cues over approximately one critical bandwidth centered at 500 Hz. A similar size of monaural critical bandwidth was assumed/discussed, for example, by Kohlrausch (1988) and by van der Heijden and Trahiotis (1998). By changing the interaural correlation of a 100-Hz-wide band, the same range of frequencies that putatively provide the interaural cues when a 500-Hz Sπ tone serves as the signal would be operative. All other aspects of the stimuli were the same as those described for the pulsed condition of Experiment 1.

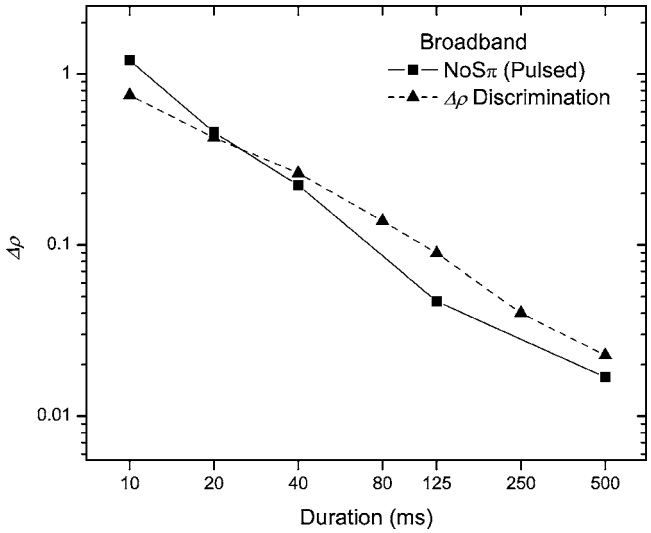


FIG. 10. Threshold change in interaural correlation ($\Delta\rho$) plotted as a function of signal duration. Squares represent pulsed NoSπ thresholds (see Fig. 1) transformed to $\Delta\rho$. Triangles represent analogous thresholds obtained using a "direct" interaural correlation discrimination paradigm.

Data were obtained using a two-cue, two-alternative, forced choice adaptive procedure targeting 71% correct (e.g., see Bernstein and Trahiotis, 1997, for details). This procedure was employed so that direct comparisons could be made in this "fringed" interaural correlation experiment with data obtained in interaural correlation-discrimination experiments conducted earlier (e.g., Bernstein and Trahiotis, 1997). The variable that was adaptively manipulated was the decrease in interaural correlation ($\Delta\rho$) from the reference interaural correlation of 1.0. Following van der Heijden and Trahiotis (1998), values of $\Delta\rho$ were generated by adjusting the ratio of two independent 100-Hz-wide bands of noise, one interaurally phase reversed (Nπ) and one diotic (No). The ratio of Nπ-to-No noise was varied adaptively in decibel steps. This produced essentially equal steps of 20 log ($\Delta\rho$) for behaviorally relevant (small) values of the ratio. The initial step size was 2 dB and it was decreased to 1 dB following two reversals.

The squares in Fig. 10 represent the NoSπ thresholds obtained in the pulsed condition of the main experiment transformed to changes in interaural correlation. The trans-

formation was done in three steps. First, the S/N thresholds obtained with the SIAM procedure were reduced by 2 dB in order to have the thresholds in that single-interval task map to the same value of d' (0.78) that defined threshold in the two-interval correlation-discrimination task. Then, the S/Ns were increased by 14.6 dB in order to reflect the threshold S/N values within a 100-Hz-wide band centered at 500 Hz (rather than the bandwidth of 2900 Hz that defined the external masker). Said differently, the value of the noise power N was taken to be the power of a 100-Hz-wide portion of the 100–3000 Hz masking noise used in Experiment 1. Finally, the adjusted values of S/N were transformed to $\Delta\rho$ via the equation (e.g., Durlach *et al.*, 1986)

$$\Delta\rho = 1 - \left[\frac{(1 - \mathrm{S/N})}{(1 + \mathrm{S/N})}\right]. \quad (1)$$

The triangles in Fig. 10 represent the data collected in the interaural correlation-discrimination task. They are the averages of 16 estimates of threshold (four estimates × four listeners) at each signal duration.

The close correspondence across the two sets of thresholds strongly supports the point of view that $\mathrm{NoS}\pi$ detection thresholds obtained in pulsed masking conditions while varying duration can be considered to be, fundamentally, manifestations of changes in interaural correlation.

D. Summary

$\mathrm{NoS}\pi$ thresholds were measured with either broadband or narrowband maskers while varying independently the durations of the signals and maskers and the durations of forward masking fringes. Comparisons among the data revealed three important trends. First, when the signal duration was 10 or 20 ms, adding a short forward fringe and backward fringe *increased* thresholds by as much as 3–4 dB, as compared to those measured in the pulsed condition. This was explained by assuming that listeners monitor the stimuli via a predetection temporal window arising from peripheral auditory processing.

Second, with a broadband masker, thresholds generally increased as the duration of the forward fringe was decreased from 500 to 2.5 ms. The magnitudes of the increase in threshold were greatest for signal durations of 10, 20, and 40 ms. These findings were accounted for by considering duration-dependent variability in activity or output that occurs across auditory filters that span the binaural operational bandwidth.

The third important trend in the data is the well-ordered, forward-fringe-duration-dependent vertical separation of thresholds between differing signal durations. These were discussed as resulting from "post-detection" integration of binaural information via a fixed temporal window combined with reduction of noise in the decision process. Further, it was argued that the degree to which each of these processes affects measures of the "pulsed effect" originally reported by McFadden (1966) depends upon details concerning exactly how signals and maskers are gated on and off.

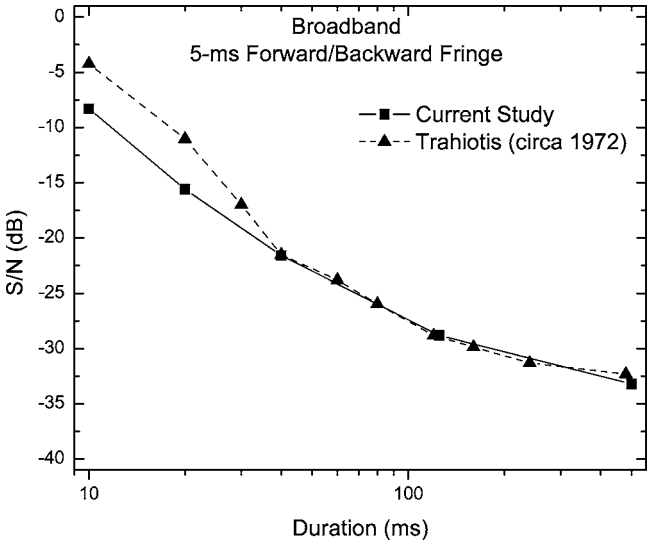


FIG. 11. Threshold signal-to-noise ratio (S/N) in decibels plotted as a function of the duration of the signal. A subset of the thresholds obtained with broadband masking in the main experiment (squares, see Fig. 1) is plotted together with thresholds (triangles) obtained in a pilot study by Trahiotis *et al.* (circa 1972). In both cases, a 5-ms forward and backward fringe of diotic noise surrounded the 500-Hz, $\mathrm{S}\pi$ signal.

ACKNOWLEDGMENTS

The authors thank Dr. Thomas N. Buell for many helpful discussions that served to both differentiate and bring focus upon the many interrelated phenomena and conceptual issues that had to be addressed. The authors also thank Dr. Armin Kohlrausch, Dr. William A. Yost, and two anonymous reviewers for their helpful comments on an earlier version of the manuscript. This research was supported by research Grants No. NIH DC-04073, No. DC-04147, and No. DC-00234 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

APPENDIX

The purpose of this Appendix is to provide further validation of the SIAM procedure employed in the main experiments and to demonstrate that duration-dependent changes in $\mathrm{NoS}\pi$ thresholds are robust, repeatable, and relatively free of interindividual differences. To that end, a subset of the thresholds obtained with broadband masking in the main experiment is plotted together (Fig. 11) with thresholds obtained in a pilot study by Trahiotis *et al.* (circa 1972). In both cases, a 5-ms forward and backward fringe of diotic noise surrounded the 500-Hz, $\mathrm{S}\pi$ signal. The earlier study employed four normal-hearing young adults [by now, probably American Association of Retired Persons (AARP) members]. As was common then, fixed psychometric procedures were employed and 400-trial-per-point psychometric functions were generated in order to estimate the S/N required for a value of $p(c)_{\max}=0.80$. The earlier experiment also employed a broadband masker extending from 100 to 3000 Hz having a spectrum level equivalent to 50 dB SPL. A single-interval, "yes/no" task, like the one used in the current study was employed. The thresholds obtained in the earlier study

were adjusted as described in Sec. II B of this paper in order to approximate the performance level of 75% that was targeted by the SIAM procedure.

The close correspondence between the two sets of thresholds that were measured 33 years apart attests not only to the validity of the SIAM procedure employed in the main experiments but also demonstrates that duration-dependent changes in NoSπ thresholds are robust, repeatable, and relatively free of interindividual (and intergenerational) differences.

[1]In accord with comments made by a reviewer, we believe it beneficial for archival consideration to differentiate clearly these types of studies from both monaural and binaural studies that have focused on what has been termed the "overshoot" phenomenon. Overshoot is operationally defined as the relatively higher level of a brief signal that is required for detection when that signal is presented shortly after the beginning, or shortly before the end, of a (longer) masker (e.g., McFadden, 1988). A major difference is that the studies of interest here manipulated the duration of the forward fringes of masking noise but included very short or no backward fringes of masking noise. In contrast, studies of overshoot, typically manipulate simultaneously, and in a complementary fashion, the duration of both forward and backward fringes of masking noise. Specifically, the relative amounts of forward and backward masking fringes change as the onset of the signal is delayed with respect to the onset of a (longer) masker of constant duration. Perhaps more to the point for our purposes, McFadden (1988) found no changes in NoSπ thresholds that could be attributable to overshoot.

[2]The predicted values of threshold do not coincide perfectly with the obtained values in the pulsed condition from which they were derived. This is so because the functions relating S/N to $\Delta\rho$ which were used to determine the criterion value of $\Delta\rho$ in the pulsed conditions were not *perfect* mirror images of the functions relating $\Delta\rho$ to S/N.

Akeroyd, M. A., and Bernstein, L. R. (**2001**). "The variation across time of sensitivity to interaural disparities: Behavioral measurements and quantitative analyses," J. Acoust. Soc. Am. **110**, 2516–2526.

Bernstein, L. R., and Trahiotis, C. (**1997**). "The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation," J. Acoust. Soc. Am. **102**, 1113–1120.

Bernstein, L. R., and Trahiotis, C. (**1999**). "The effects of signal duration on NoSo and NoSπ thresholds at 500 Hz and 4 kHz," J. Acoust. Soc. Am. **105**, 1776–1783.

Bernstein, L. R., Par, Steven van de, and Trahiotis, C. (**1999**). "The normalized correlation: Accounting for NoSπ thresholds obtained with Gaussian and "low-noise" masking noise," J. Acoust. Soc. Am. **106**, 870–876.

Bernstein, L. R., Trahiotis, C., Akeroyd, M. A., and Hartung, K. (**2001**). "Sensitivity to brief changes of interaural time and interaural intensity," J. Acoust. Soc. Am. **109**, 1604–1615.

Blodgett, H. C., Jeffress, L. A., and Taylor, R. W. (**1958**). "Relation of masked threshold to signal-duration for various interaural phase-combinations," Am. J. Psychol. **71**, 283–290.

Bourbon, W. T., Jr., (1966). "Effects of bandwidth and level of masking noise on detection of homophasic and antiphasic tonal signals," Dissertation, University of Texas, Austin, Texas.

Breebaart, J., Par, S. van de, and Kohlrausch, A. (**2001**). "Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters," J. Acoust. Soc. Am. **110**, 1105–1117.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996a**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996b**). "A quantitative model of the 'effective' signal processing in the auditory system. II. Simulations and measurements," J. Acoust. Soc. Am. **99**, 3623–3631.

Diercks, J. K., and Jeffress, L. A. (**1962**). "Interaural phase and the absolute threshold for tone," J. Acoust. Soc. Am. **34**, 981–984.

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (**1986**). "Interaural correlation discrimination: II. Relation to binaural unmasking," J. Acoust. Soc. Am. **79**, 1548–1557.

Egan, J. P. (**1965**). "Masking-level differences as a function of interaural disparities in intensity of signal and of noise," J. Acoust. Soc. Am. **38**, 1043–1049.

Elliott, L. L. (**1965**). "Changes in the simultaneous masked threshold of brief tones," J. Acoust. Soc. Am. **38**, 738–746.

Gilkey, R. H., Simpson, B. D., and Weisenberger, J. M. (**1990**). "Masker fringe and binaural detection," J. Acoust. Soc. Am. **88**, 1323–1332.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Hall, J. W., Tyler, R. S., and Fernandes, M. A. (**1983**). "Monaural and binaural auditory frequency resolution measured using bandlimited noise and notched-noise masking," J. Acoust. Soc. Am. **73**, 894–898.

Hartung, K., and Trahiotis, C. (**2001**). "Peripheral auditory processing and investigations of the 'precedence effect' which utilize successive transient stimuli," J. Acoust. Soc. Am., **110**, 1505–1513.

Hays, W. L. (**1973**). *Statistics for the Social Sciences* (Holt, Rinehart, and Winston, New York).

Heijden, M. van der, and Trahiotis, C. (**1998**). "Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution," J. Acoust. Soc. Am. **103**, 1609–1614.

Jain, M., Gallagher, D. T., Koehnke, J., and Colburn, H. S. (**1991**). "Fringed correlation discrimination and binaural detection," J. Acoust. Soc. Am. **90**, 1918–1926.

Kaernbach, C. (**1990**). "A single-interval adjustment-matrix (SIAM) procedure for unbiased adaptive testing," J. Acoust. Soc. Am. **88**, 2645–2655.

Keppel, G. (**1973**). *Design and Analysis: A Researchers Handbook* (Prentice-Hall, Englewood Cliffs, New Jersey).

Kohlrausch, A. (**1988**). "Auditory filter shape derived from binaural masking experiments," J. Acoust. Soc. Am. **84**, 573–583.

Kohlrausch, A., and Fassel, R. (**1997**). "Binaural masking level differences in nonsimultaneous masking," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. Anderson (Lawrence Erlbaum Assoc., Mahwah, NJ), Chap. 9, pp. 169–190.

Leshowitz, B., and Wightman, F. L. (**1971**). "On-frequency masking with continuous sinusoids," J. Acoust. Soc. Am. **49**, 1180–1190.

McFadden, D. M. (**1966**). "Masking-level differences with continuous and with burst masking noise," J. Acoust. Soc. Am. **40**, 1414–1419.

McFadden, D. M. (**1968**). "Masking-level differences determined with and without interaural disparities in masker intensity," J. Acoust. Soc. Am. **44**, 212–223.

McFadden, D. M. (**1988**). "Absence of overshoot in a dichotic masking condition," J. Acoust. Soc. Am. **38**, 1685–1687.

Meddis, R. (**1986**). "Simulation of mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. **79**, 702–711.

Par, S. van de, and Kohlrausch, A. (**1999**). "Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters," J. Acoust. Soc. Am. **106**, 1940–1947.

Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). "Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform," J. Acoust. Soc. Am. **98**, 1890–1894.

Robinson, D. E., and Trahiotis, C. (**1972**). "Effects of signal duration and masker duration on detectability under diotic and dichotic listening conditions," Percept. Psychophys. **12**, 333–334.

Robinson, D. E., and Jeffress, L.A. (**1963**). "Effects of varying the interaural noise correlation on the detectability of tonal signals," J. Acoust. Soc. Am. **35**, 1947–1952.

Slaney, M. (**1993**). "An efficient implementation of the Patterson-Holdsworth auditory filter bank," Apple Computer Technical Report #35.

Trahiotis, C. (**2005**). (private communication).

Trahiotis, C., Dolan, T. R., and Miller, T. E. (**1972**). "Effect of 'backward' masker fringe on the detectability of pulsed diotic and dichotic tonal signals," Percept. Psychophys. **12**, 335–338.

Wightman, F. L. (**1971**). "Detection of binaural tones as a function of masker bandwidth," J. Acoust. Soc. Am. **50**, 623–636.

Yost, W. A., (**1985**). "Prior stimulation and the masking-level difference," J. Acoust. Soc. Am. **78**, 901–907.

Yost, W. A. (**1988**). "The masking-level difference and overall masker level: Restating the internal noise hypothesis," J. Acoust. Soc. Am. **83**, 1517–1521.

Zwicker, U. T., and Zwicker, E. (**1984**). "Binaural masking-level difference as a function of masker and test-signal duration, Hear. Res. **13**, 215–220.

Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners[a)]

Bom Jun Kwon[b)] and Chris van den Honert
Cochlear Americas, 400 Inverness Parkway, Suite 400, Englewood, Colorado 80112

(Received 11 July 2005; revised 7 February 2006; accepted 8 February 2006)

Bipolar stimulation has been thought to be more beneficial than monopolar stimulation for speech coding in cochlear implants, on the basis of its more restricted current flow. The present study examined whether bipolar stimulation would indeed lead to reduced channel interaction in a behavioral forward masking experiment tested in four Nucleus® 24 users. The masker was fixed on one channel and three masker levels that were balanced for loudness between the configurations were chosen. As expected, masking was maximal when the masker and probe channels were spatially close and decreased as they were separated. However, overall masking patterns did not consistently demonstrate sharper tuning with bipolar stimulation than monopolar. This implies that the spatial extent of a bipolar current field is not consistently narrower than that of an equally loud monopolar stimulus; therefore, it should not be assumed that bipolar stimulation leads to reduced channel interaction. Notably, bipolar masking patterns appeared to display more variations across channels, possibly influenced more by anatomical and neural irregularities near electrode contacts than monopolar masking patterns. The present psychophysical results provide a theoretical basis regarding the widespread use (and success) of monopolar configurations by implant users. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184128]

PACS number(s): 43.66.Ts, 43.66.Dc, 43.66.Ba [JG] Pages: 2994–3002

I. INTRODUCTION

Since multiple-electrode arrays were first introduced in cochlear implant (CI) systems, channel interactions have emerged as a limiting factor in increasing the number of electrodes. With simultaneous stimulation, channel interactions result from summation of current fields. While such summation is avoided with nonsimultaneous stimulation, channel interactions nevertheless exist due to (1) summation of charge accumulation at the neural membrane, (2) refractoriness of neurons previously excited by an adjacent channel, or (3) central interactions. Speech coding strategies that utilize simultaneous stimulation, such as the compressed analog (CA) strategy, are susceptible to current field summation. Sequential pulsatile stimulation, as used in the continuous interleaved sampling (CIS; Wilson *et al.*, 1991), eliminates current field summation, as multiple current fields do not occur simultaneously. By reducing the current field interactions, the CIS strategy has been shown to provide better speech recognition for CI patients, compared with the CA strategy (Wilson *et al.*, 1991).

With nonsimultaneous stimulation of channels, neural interactions might occur at both the peripheral and central stages of the auditory system, as observed in psychophysical tasks such as loudness summation (e.g., Shannon, 1983) or gap detection (Hanekom and Shannon, 1998). Forward masking is another psychophysical task used to study channel interactions in CIs (Shannon, 1983; Tong and Clark, 1986; Lim *et al.*, 1989; Chatterjee and Shannon, 1998; Cohen *et al.*, 1996, 2001; Boëx *et al.*, 2003). The threshold of a probe stimulus could be elevated by a preceding masker stimulus, if the electrodes carrying the masker and probe stimuli are close in space. The threshold shift, or the amount of masking, might be an indicator of channel interaction between the masker and probe, as we would expect a large threshold shift if the masker and probe channels activate a large population of common neurons. Therefore, the amount of masking as a function of the intracochlear spatial distance between the masker and probe channels would reveal the "width" of channel interactions.

In bipolar stimulation, current flows between an active and a return electrode located within the cochlea. The bipolar stimulation "mode" for Nucleus® devices indicates the distance between the active and return electrodes (in terms of electrode separation): current flows between adjacent electrodes in BP mode, current flows between electrodes that are separated by one electrode in BP+1 mode, and so on. In monopolar stimulation, current flows between an active intracochlear electrode and one or two extracochlear return electrodes (a ball electrode placed under the temporalis muscle for MP1 mode; a plate electrode on the casing of the internal receiver for MP2 mode; and both electrodes for MP1+2 mode). Monopolar stimulation generally requires much less current to evoke a hearing sensation, and therefore provides an advantage in terms of power consumption. Physiological measures have shown that "narrow" bipolar configurations produce spatially restricted current fields (Kral *et al.*, 1998), providing more "focused" stimulation of neural populations (van den Honert and Stypulkowski, 1987;

[a)]A portion of this study was presented in poster sessions at the 33rd Neural Prosthesis Workshop, October 2002 (Bethesda, MD), and at the 26th Mid-Winter Research Meeting Association for Research in Otolaryngology, February 2003 (Daytona Beach, FL).

[b)]Electronic mail: bjkwon@gmail.com

 0001-4966/2006/119(5)/2994/9/$22.50

Bierer and Middlebrooks, 2002). Monopolar configurations, on the other hand, produce broader current fields (van den Honert and Stypulkowski, 1987; Bierer and Middlebrooks, 2002). One could presume that bipolar stimulation, with its narrow current fields, would produce less channel interaction than monopolar stimulation. However, the presumed advantages of bipolar stimulation have not been shown in patient performance. Speech recognition with monopolar stimulation has been equal to or better than that with bipolar stimulation and many patients have expressed a preference for monopolar rather than bipolar stimulation modes (Lehnardt *et al.*, 1992; von Wallenberg *et al.*, 1994; Zwolan *et al.*, 1996; Pfingst *et al.*, 1997, 2001; Kileny *et al.*, 1998). Thus, monopolar electrode configurations are most commonly used in many CI speech processors.

Previous studies have shown mixed results regarding the effects of electrode configuration on forward masking. Shannon (1983) observed broader forward masking patterns with the monopolar configuration than with the bipolar configuration, consistent with physiological predictions (van den Honert and Stypulkowski, 1987; Bierer and Middlebrooks, 2002). A recent study by Boëx *et al.* (2003) reiterated Shannon's observation, with Clarion® patients implanted with the HiFocus-I™ array. With Nucleus® 22 CI patients, "pseudo-monopolar" forward masking patterns were measured by Cohen *et al.* (1996, 2001), in which the most basal intracochlear electrode was used as the return electrode in a bipolar pair (monopolar stimulation is not possible with the Nucleus® 22 device, as all electrodes are located inside of the cochlea). Cohen *et al.* (1996, 2001) found that the masking patterns for pseudo-monopolar maskers were not consistently broader than those of bipolar maskers. A similar observation was made by Lim *et al.* (1989), who varied the distance between the active and return electrodes in bipolar maskers (or the "spatial extent"). They found that the masking patterns initially became broader as the spatial extent of the masker was increased; when the spatial extent was sufficiently wide, the masking patterns often exhibited two peaks, rather than one broad peak. Beyond very narrow bipolar configurations, the width of the peak(s) did not increase with wider bipolar configurations. Thus, these previous studies present somewhat conflicting views regarding the effects of electrode configuration.

While these previous studies have shown some effects of electrode configuration on forward masking patterns, it is unclear, as yet, how the *relative* spread of excitation differs between monopolar and bipolar stimulation modes. Another consideration is the effect of stimulation level on forward masking patterns, which have been shown to broaden as the masker level is increased, in both acoustic (e.g., Nelson and Fortune, 1991) and electric hearing (e.g., Chatterjee and Shannon, 1998). It is unclear how the spread of excitation differs with stimulation level, or whether this interacts with the masker electrode configuration. To address these issues, forward-masked excitation patterns were measured as functions of masker stimulation mode and level in Nucleus® 24 CI users. The masker stimulation mode was varied between either bipolar or monopolar. Maskers were presented at loudness-balanced levels corresponding to soft, soft-medium, and medium listening levels.

TABLE I. Subjects' demographic information, etiology of deafness, and CNC word recognition scores in quiet.

ID	Gender	Age (years)	CI use (years)	Etiology	CNC word score
S1	Female	28	5	Unknown	78%
S3	Male	72	5	Noise exposure	46%
S4	Male	55	2	Otosclerosis	63%
S5	Male	74	3	Familial	54%

II. METHODS

A. Subjects and equipment

Four adult postlingually deafened CI patients participated in the present study. All subjects were implanted with the Nucleus® CI24M device (straight electrode array) and each had least 2 years' experience with his/her implant prior to testing. Table I contains relevant subject information, including CNC word recognition scores in quiet (Peterson and Lehiste, 1962), measured prior to the present study with their clinical MAP (MP1+2 mode). The use of human subjects in this study was reviewed and approved by the Western Institutional Review Board.

All stimuli were presented via the Nucleus® Implant Communicator™ (NIC™), a dynamic link library interface between a computer and the SPrint™ speech processor connected to the clinical fitting interface (Programming Control Interface). The computer generated the stimuli and the SPrint™ processor transmitted the encoded radio frequency (RF) pulses to the subject's internal receiver/stimulator.

The forward masking experiments were conducted using custom software written in C. During testing, one computer was dedicated to stimulus generation and test procedure, while another computer was dedicated as subject interface (e.g., subject response, test instructions, feedback, etc.). The two computers were linked via TCP/IP-based custom software.

B. Stimuli

In the Nucleus® 24 implant system, stimulation current intensity is specified in logarithmic units of "current levels (CLs)." The relationship between intensity represented in CLs and the corresponding linear current I in μA is given by

$$I = 10I_o^{\mathrm{CL}/255}, \tag{1}$$

where CL is an integer ranging from 0 to 255 and I_o has a nominal value of 175 μA, which is one tenth of the maximum current of the device (the actual calibration data for individual devices are displayed in Table II). One CL corresponds to 0.176 dB. Increasing the current by 34 CLs is approximately equivalent to doubling it (or a 6-dB increase).

Figure 1 shows the stimulus timing (a) and the masker and probe electrode locations (b) used in the forward masking experiments. All stimuli were biphasic, charge-balanced

TABLE II. Factory-calibrated currents (see text), per channel stimulation rates and modes used in clinical speech processing MAP, and the experimental stimulation rates and modes used for experiments.

Subject	I_o in Eq. (1) (μA)	Stimulation rate, mode (clinically) assigned MAP)	Stimulation rate for experiments	
			Monopolar	Bipolar (mode)
S1	190	1800 Hz, MP1+2	1800 Hz	1800 Hz (BP+1)
S3	175	900 Hz, MP1+2	900 Hz	2400 Hz (BP+2)
S4	178	900 Hz, MP1+2	1800 Hz	2400 Hz (BP+2)
S5	188	250 Hz,MP1+2	2400 Hz	2400 Hz (BP+2)

pulse trains. The masker duration was 500 ms, and the probe duration was 20 ms; the masker-probe delay was 20 ms. The pulse phase width was 25 μs and the interphase gap was 8 μs. The same phase width was used for both monopolar and bipolar conditions, so that equal currents indicate equal amounts of charge. For each monopolar and bipolar condition, masker and probe were presented in the same mode. For the bipolar conditions, due to the relatively high current requirement, the narrowest bipolar spacing (BP mode) could not be used. The stimulation mode was varied across subjects to ensure that sufficient loudness could be achieved for as narrow a bipolar configuration as possible. For S1, sufficient loudness was achieved with BP+1 mode; for the rest of the subjects, BP+2 mode was necessary. Furthermore, relatively high stimulation rates were used with bipolar conditions for each subject, to alleviate the high current level requirement in bipolar stimulation. The exact stimulation rates used for bipolar and monopolar stimulation in each subject are shown in Table II, with corresponding values from their clinical MAPs. Originally, it was intended that the same stimulation rate would be used for both monopolar and bipolar conditions. However, a somewhat lower monopolar rate was chosen for two subjects who did not prefer the high stimulation rate (S3 and S4). The masker was always presented to electrode 7; probe stimuli were presented to electrode 2, 4, 6, 8, 10, 12, 14, 16, or 18. [For brevity, the stimulation channel is denoted by the active electrode with the mode specification—e.g., when the stimulation is at (active) electrode 7, the actual electrode pair is (7,9), (7,10), or (7, *extracochlear electrodes*) with BP+1, BP+2, or MP1+2 mode, respectively.]

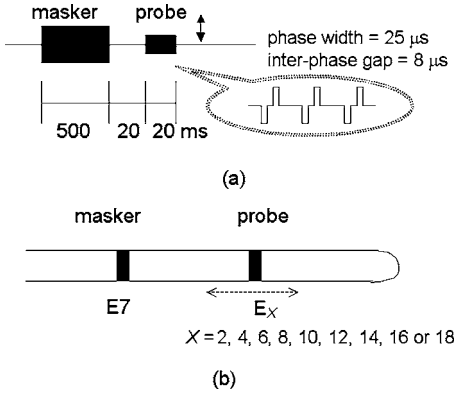


FIG. 1. Schematic diagram of the temporal (a) and spatial (b) properties of the masker and probe stimuli. Note that only the active electrode is denoted in this paper to specify a stimulation channel. For example, electrode 7 with BP+1 mode refers to the current field between the electrode pair (7,9).

Forward masking was measured at three presentation levels, corresponding to soft, soft-medium, and medium loudness levels, which are shown in Table III. Loudness at each presentation level was approximately matched across stimulation modes. First, three levels were established with monopolar stimulation. Then, for each presentation level, the experimenter adjusted the level of the bipolar masker until the subject reported that the monopolar and bipolar maskers were equally loud.

C. Procedure

Forward masking was measured using an adaptive, three-alternative forced-choice procedure (two-down/one-up). The masker was present in all three intervals; the probe existed in only one interval following the masker. The subject's task was to identify the interval containing the probe. The level of the probe was adjusted according to subject response, ultimately converging on 70.7% correct (Levitt, 1971). The adjustment step size was initially 0.88 dB (5 CLs) and was reduced to 0.35 dB (2 CLs) after six reversals. Each run contained 14 reversals, and the mean of the final eight reversals was calculated as the threshold. Each condition was repeated three times and the measurements were averaged. In addition, absolute detection thresholds without the masker were measured for all probe stimuli through the same procedure. For each probe, the threshold shift was calculated as the difference between the masked threshold and the detection threshold without the masker.

III. RESULTS

Figure 2 shows the detection thresholds with and without the masker for each electrode and stimulation configura-

TABLE III. Masker levels, in CL units (min: 0, max: 255), used for the soft, soft-medium, and medium presentation levels. The loudness was balanced between stimulation modes for each level.

Subject	Monopolar			Bipolar		
	Soft	Soft-Med	Medium	Soft	Soft-Med	Medium
S1	129	150	170	182	230	245
S3	160	180	190	200	235	245
S4	148	160	180	205	220	243
S5	155	170	190	217	234	250

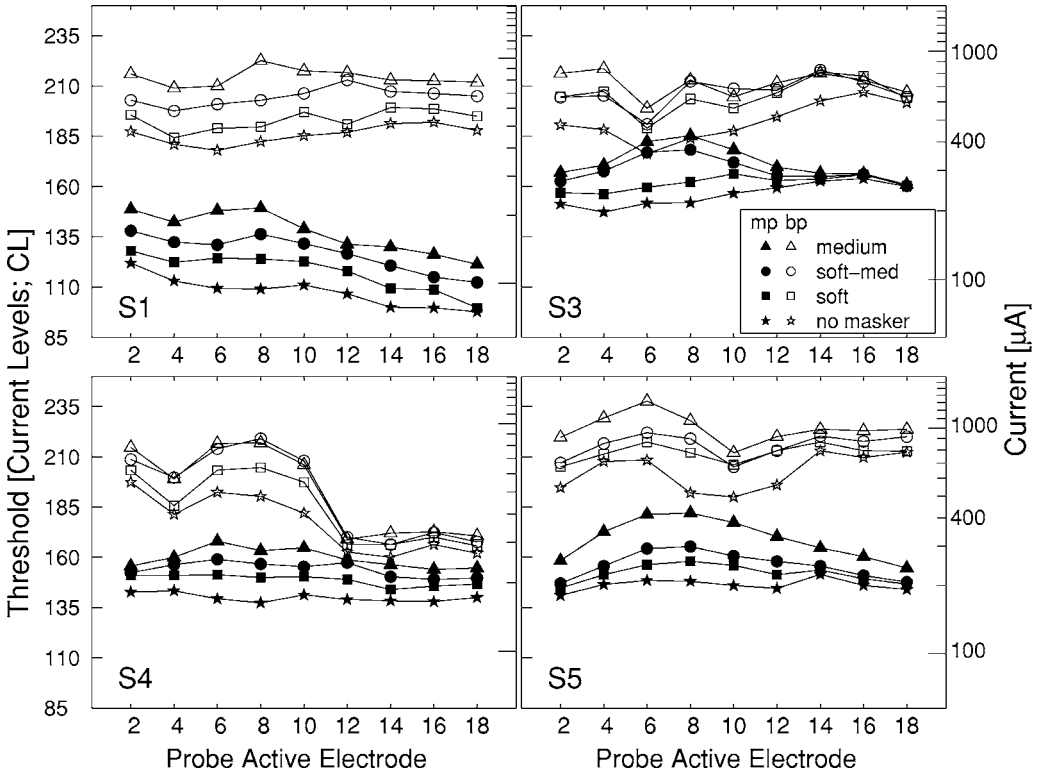


FIG. 2. Detection thresholds with the monopolar and bipolar configurations for each subject. The *x* axis shows the probe location (active electrode). The filled and open symbols indicate masked thresholds at three masker levels for monopolar and bipolar stimulations, respectively.

tion. Thresholds are plotted against the CL scale on the left axis, and log μA on the right for reference. Figure 3 displays threshold shifts, calculated from the masked threshold subtracted by the unmasked threshold in the CL scale. The different symbols represent different masker presentation levels. There was large intersubject variability in terms of the shape of the masking patterns. However, the results are generally consistent with those of previous studies (Cohen *et al.*, 1996; Chatterjee and Shannon, 1998), in that the peak in the forward masking pattern was found near the masker location, and the amount of masking is reduced as the masker and probe become more spatially remote. Masked thresholds

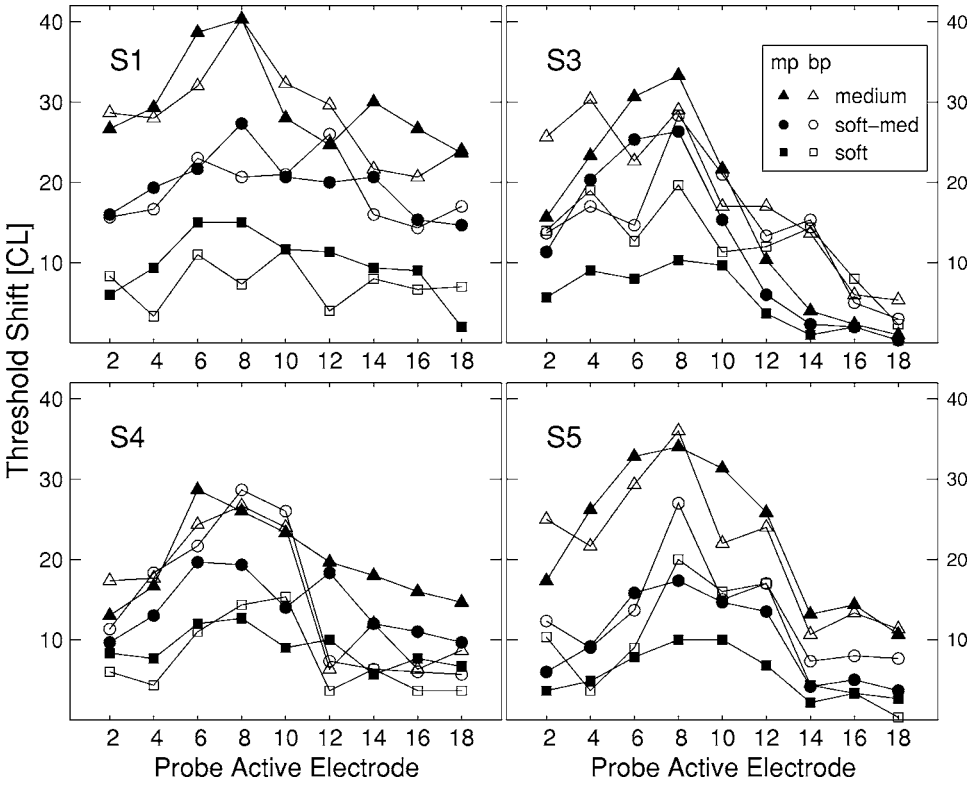


FIG. 3. Threshold shifts (difference between masked and unmasked thresholds) in CLs, plotted as a function of the active electrode of the probe channel. The *x* axis shows the probe location (active electrode). The filled and open symbols indicate masked thresholds at three masker levels for monopolar and bipolar stimulations, respectively.

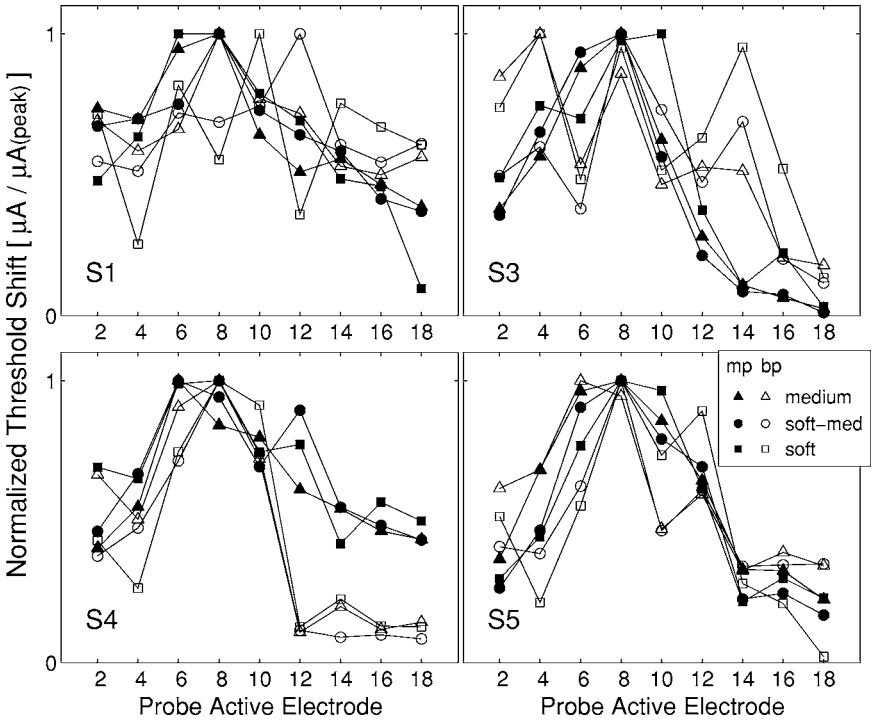


FIG. 4. Normalized threshold shifts for each subject. For each masking pattern, threshold shift was normalized to the peak threshold shift (in μA). Figure format is the same as Fig. 3.

generally increased with masker level. As seen in Fig. 3, the maximum amount of masking at each masker level is comparable between monopolar and bipolar modes in most cases, implying that the loudness balance between stimulation modes was properly done. In terms of absolute detection thresholds (cf. Fig. 2), there was a greater variability in bipolar thresholds across the electrode array, compared with monopolar thresholds, consistent with the results found by Pfingst and Xu (2004). Statistical significance was assessed using a three-way, repeated measures ANOVA to test the effects of the following within-subject factors on threshold shifts: stimulation mode (two levels: monopolar, bipolar), probe electrode location (nine levels), and masker level [three levels: soft (S), soft/medium (SM), medium (M)]. Most importantly, the effect of stimulation mode was not significant ($F_{1,3}=0.32, p=0.61$), and none of interaction terms with stimulation mode was significant (all the p values were greater than 0.2). While both the effects of probe location ($F_{8,24}=16.24, p<0.01$) and masker level ($F_{2,6}=18.64, p<0.01$) were significant, the interaction between them was also significant ($F_{16,48}=7.08, p<0.01$); therefore, a *posthoc* analysis was necessary to assess the effects of interaction. The monopolar and bipolar data were grouped together for each condition and three paired t tests (S-SM, SM-M, and S-M) for the masker level were done at each probe location. With a 95% confidence interval, the effect of masker level was the most significant at probe electrode 6 (all three t-test pairs significantly different), two pairs were significant at probe electrodes 2, 4, 8, and 10, and one or no pair was significant from electrode 12 and the rest. In other words, the interaction term indicated that threshold shifts depended on masker level only for probe locations near active electrode 7 and there was no effect of masker level for remote locations.

As an alternative analysis, threshold shifts were also calculated as the difference in linear μA. Due to the substantial difference in the absolute magnitudes of the stimuli, it would be difficult to plot and compare the result patterns for monopolar and bipolar stimulations in the same graph panel. Therefore, threshold shifts were normalized to the peak masking (in linear μA) in each level and plotted in Fig. 4, similar to the analysis done by Shannon (1983, 1990; see the Appendix). Once normalized, the masking patterns are quite similar across stimulation level within each stimulation mode. There are some deviations in this trend, notably for the bipolar masking patterns in subjects S1 and S3. At the soft and soft-medium presentation levels, the normalized masking patterns are more diffused. Only at the medium presentation level does a clear peak emerge near the masker location. Across stimulation mode, the patterns are also similar, in that the peak of the masking patterns is generally found near the masker location. However, there are also noticeable differences in the masking patterns between the stimulation modes. For subject S4, there was relatively little masking for the apical probe electrodes with bipolar stimulation; with monopolar stimulation, there was relatively more masking across the electrode array. Conversely, for subject S3, there was relatively more masking with bipolar stimulation for the apical probe electrodes at the soft masker level, and more masking for the basal electrodes at the medium level. This trend created double-peaked masking patterns. Local irregularities in neural survival and the geometric relation between implanted electrodes and healthy neural populations most likely contributed to the variance in the threshold patterns measured with bipolar stimulation, as previously described by Pfingst *et al.* (2004). With monopolar

stimulation, threshold patterns were less variable, most likely because of the presumed broader current spread within the cochlea.

IV. DISCUSSION

A. Comparison with previous studies

The peak threshold shifts measured in the present study, 4.7 to 7 dB (or 27 to 40 CLs), are comparable to the 7–11-dB range reported by Chatterjee and Shannon (1998; from their Fig. 1, converted to dB), and the 2–10-dB range reported by Cohen *et al.* (1996). The difference in these measurements could reflect methodological differences and parameter choices (e.g., masker and probe durations, interstimulus delay, etc). Considering that it is not uncommon to observe a substantial variability in forward masking data with cochlear implants in the literature [e.g., a wide range of the time constants for recovery from masking reported by Shannon (1990), Chatterjee and Shannon (1998), and Nelson and Donaldson (2002)], the discrepancy between the present and previous measurements is rather modest. In fact, some qualitative similarities across studies are noteworthy. For example, the peak of the masking pattern was generally located closest to the masker and the amount of masking increased with masker level. The across-site threshold variability was greater for bipolar stimulation than for monopolar stimulation, consistent with the findings of Pfingst and Xu (2004). However, the spread of excitation, as represented by forward masking patterns, was not significantly greater for monopolar stimulation than for bipolar stimulation in the present study. This observation is consistent with the results from Cohen *et al.* (1996, 2001), but somewhat in conflict with previous results from Shannon (1983) and Boëx *et al.* (2003). We note that the argument for broader monopolar masking patterns in those studies is relatively weak, as the observation was made with only one subject (out of two) in both studies. The results from the present study indicate that the difference in masking patterns between stimulation configurations often clearly exists. As seen in the present study, the difference could lead to either broader monopolar masking (e.g., S4) or broader bipolar masking (e.g., S3), depending on the subject and the masker level. Therefore, with the data from the present and previous studies at hand, we cannot conclude that convincing and consistent evidence exists for broader monopolar masking.

To rephrase the findings of the present study, at suprathreshold levels, the relative spread of masking is overall not significantly different between monopolar and bipolar stimulation, while an intersubject variability might exist. As such, the degree of channel interaction is most likely similar between these stimulation modes. The discrepancy between the present psychophysical and previous physiological data (e.g., Bierer and Middlebrooks, 2004) still needs to be reconciled. According to the spike-count model of loudness, the loudness of a stimulus is determined by the number of neurons responding to electrical stimulation. Spatially compact stimulation (e.g., narrow bipolar electrode configurations) is disadvantageous in terms of neural recruitment (Pfingst *et al.*, 1997): relatively high current levels are required to evoke comparable loudness with broader stimulation modes (e.g., wide bipolar or monopolar electrode configurations). It is plausible that the elevated current requirement for sensation (due to the very nature of narrow stimulation) causes the spatial extent to be broadened, ironically compromising its benefit. In physiological studies, the levels of stimuli are determined mainly by the observed neural responses, possibly lower than practically required for behavioral testing; therefore, it would be easier to maintain a focused spatial extent with bipolar stimulation. While loudness growth could differ between bipolar and monopolar stimulation, these stimulation modes should be compared at similar loudness levels, which is difficult to do in physiological studies. Moreover, in Bierer and Middlebrooks (2004), the different degrees of channel interaction observed with different electrode configurations occurred primarily with simultaneous stimulation. The difference in channel interaction was much smaller with nonsimultaneous stimulation.

B. Growth of masking

Another way to characterize the differences between bipolar and monopolar stimulation is to compare the growth of masking (GOM), i.e., threshold shift as a function of masker level. For each subject and at each probe location, the slope of GOM was calculated from the linear regression fit. Figure 5 shows the slope of GOM for each probe location condition, also illustrating a typical example of GOM function in the small figure insert. The slope of GOM, in $\mu A/\mu A$, is related to the effectiveness of a masker: the unity slope indicates that the increase of masker intensity leads to the increase in threshold shift to the same degree. For each stimulation mode, the maximum slope (steepest GOM) generally occurred near the masker location, ranging from 0.44 to 0.75 for monopolar and 0.15 to 0.51 for bipolar. These maximum slopes are comparable to the data range (0.38–0.72) reported by Chatterjee and Shannon (1998). The slopes decreased as the probe became spatially remote from the masker. A two-way repeated measures ANOVA on the GOM slopes, with factors stimulation mode (monopolar, bipolar) and probe electrode location (nine levels), indicated a significant effect of both factors (mode: $F_{1,3}=19.224$, $p=0.022$; electrode: $F_{8,24}=12.517$, $p<0.01$). The interaction between these factors was not significant ($F_{8,24}=2.209, p=0.064$). This indicates that the GOM slopes were generally steeper with monopolar stimulation than bipolar stimulation. The slopes in the bipolar condition were particularly small for S3 and S4, reflecting that as the masker level increased, the threshold shifts were relatively unchanged with bipolar stimulation. These shallow slopes might have in part resulted from irregularities in neural survival (which are also reflected in the previous figures of thresholds and threshold shifts). Possibly, uneven and poor neural survival at the sites of excitation (in the proximity of electrode contacts) compromised the spread of excitation associated with the increase of the masker level. Therefore, this phenomenon is a complex interaction between the profile of surviving neurons and current spread of the stimulation configuration. It is also possible that most surviving neurons in the region, where the neural survival is

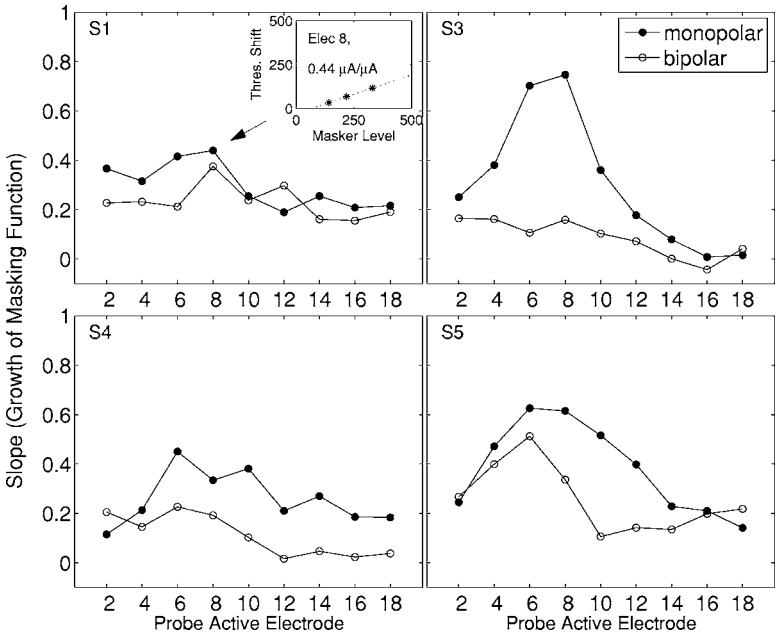


FIG. 5. Slope of linear fits to growth of masking at each probe location. Linear regressions were applied to masked threshold shifts (in μA) as function of masker level (in μA), at each probe location. The filled and open symbols shows the slope values for monopolar and bipolar stimulations, respectively. The small figure insert shows a typical example of growth of masking from S1's monopolar data at electrode 8.

uneven, have relatively high neural thresholds, thus making the increase of masker level less effective in recruiting more neurons in the spread region. At any rate, it is likely that local irregularities such as uneven or poor neural population exist for these two subjects, and they were reflected in the GOM slopes with the bipolar configuration. With monopolar stimulation, because the current spreads more broadly in the cochlea, there is less variance in the detection thresholds, thereby producing a more predictable growth of masking across the probe locations, relative to the masker location.

C. Implications for speech coding

The comparable forward masking patterns between monopolar and bipolar stimulation modes might explain why speech recognition results are generally similar between the two stimulation modes. The presumed better spatial selectivity with bipolar stimulation has not provided consistently better speech recognition results. The smaller power consumption and relative uniformity in thresholds and comfort levels can provide a substantial advantage for monopolar stimulation with current clinical speech processors. Pfingst *et al.* (2004) recently demonstrated a negative correlation between across-site variation in thresholds and speech recognition scores (correlation coefficients ranging from -0.7 to -0.47, depending on testing materials, with statistical significance, $p<0.05$). That would be an additional reason why monopolar stimulation, with its smaller across-site threshold variability, could be preferable to bipolar stimulation. This implies that stimulations that further restrict the current flow (e.g., tripolar stimulation) would presumably produce even more across-site threshold variability, and might not provide a noticeable advantage in speech recognition, as long as the coding strategy employs pulsatile sequential stimulation, such as CIS (Wilson *et al.*, 1991). Of course, regardless of stimulation mode, any speech processing strategy will be limited by the health of the neural populations and their proximity to the implanted electrodes. Finally, narrow electrode configurations might provide some advantages, if the strategy demands increased channel independence to permit simultaneous stimulation across channels (e.g., Clopton *et al.*, 2003). However, such an advantage has yet to be shown.

ACKNOWLEDGMENTS

The authors thank Wendy Parkinson for assisting data collection and her contribution to the project. The first author thanks John J. Galvin III for his critically valuable review and editorial assistance during the manuscript preparation, Robert Shannon and Monita Chatterjee for constructive discussion in the early stages of the study, and also two anonymous reviewers as well as John Grose for helpful comments. Finally, many thanks go to the NIC development team (Bret Swanson, Colin Irwin, and Michael Goorevich) in Cochlear Ltd., Australia.

APPENDIX: LINEAR VERSUS LOGARITHMIC SCALE TO REPRESENT CURRENT THRESHOLDS

It has long been debated as to how current threshold values in CI should be scaled. Some researchers have used a linear current scale, in μA (e.g., Chatterjee and Shannon, 1998; Nelson and Donaldson, 2001, 2002). Others have used a logarithmic current scale, in dB (e.g., Cohen *et al.*, 1996; Lim *et al.*, 1989). In the present study, currents were shown both in terms of CLs (Figs. 2 and 3) and μA (Figs. 4 and 5). Linear current scales may be appropriate because (1) loudness growth has been shown to be linearly related between acoustic sound pressure level (in dB SPL) and electric current (in μA) with patients who had a substantial residual hearing contralateral to the implanted ear (e.g., Eddington *et al.*, 1978; Zeng and Shannon, 1993; Dorman *et al.*, 1993), and (2) the loudness theory of electrical stimulation proposed by Zeng and Shannon (1994) would be better suited with a linear current scale. However, the demonstration of linearity between dB SPL and μA is not as definitive as it appears in

TABLE IV. The ranges of electric currents used for the previous loudness matching studies between acoustic and electric stimuli and the r values of regression for linear and logarithmic fit. The data of ABI 14 from Zeng and Shannon are represented in mV.

	Current range (μA)		r	
Study reference	min	max	Linear	Log
Eddington *et al.* (1978)				
Figure 18, 400 Hz	103	165	0.992	0.984
Zeng and Shannon (1993)				
ABI 17, Fix A	82	495	0.995	0.979
ABI 14	575	843	0.985	0.991
ABI 18	251	445	0.973	0.969
Dorman *et al.* (1993)				
electrode 1	72	130	0.994	0.996
electrode 2	72	140	0.993	0.995

the literature. The authors in the above studies (Eddington *et al.*, 1978; Zeng and Shannon, 1993; Dorman *et al.*, 1993) emphasized the successful linear fitting, but did not present a comparative fitting with a logarithmic scale. Due to the narrow dynamic range of electrical stimulation, the maximum current was less than twice the minimum current in these studies (cf. Table IV) and regression fitting of the data on both linear and log scales would have been equally good. We attempted a logarithmic fitting of the data and present the R values in Table IV, demonstrating that either fitting is indeed equally successful. Nevertheless, the successful demonstration of linear fitting in those studies might have led subsequent researchers to the inference that a linear scale is more proper.

When the range of amplitude under consideration is small, the choice of scale (linear or logarithmic) is of little consequence. However, if the range is sufficiently large, it can affect interpretation of the data. Threshold shifts in μA would have been difficult to compare without the normalization, as done in Fig. 4, because the maximum threshold shift with each mode differs substantially. A large change in amplitude is often necessary for bipolar stimulation to match the perceptual change in small amplitude with monopolar stimulation. Perhaps a logarithmic scale is particularly useful if the range of comparison is large, as the use of a logarithmic scale would mitigate this mismatch of scaling. Thus, a linear scale should be used only when current values have similar absolute magnitudes (e.g., as with a fixed stimulation mode). Of course, normalizing linear threshold shifts to the peak shift (as in Fig. 4) also allows direct comparison of stimuli with large differences in current values. This is essentially the same approach in a log (dB) scale, in that the ratio (rather than the absolute difference) between physical measurements is analyzed.

Bierer, J. A., and Middlebrooks, J. C. (**2002**). "Auditory cortical images of cochlear implant stimuli," J. Neurophysiol. **87**, 478–492.

Bierer, J. A., and Middlebrooks, J. C. (**2004**). "Cortical responses to cochlear implant stimulation: channel interactions," J. Assoc. Res. Otolaryngol. **5**, 32–48.

Boëx, C., Kos, M-I., and Pelizzone, M. (**2003**). "Forward masking in different cochlear implant systems," J. Acoust. Soc. Am. **114**, 2058–2065.

Chatterjee, M., and Shannon, R. V. (**1998**). "Forward masked excitation patterns in multielectrode electrical stimulation," J. Acoust. Soc. Am. **103**, 2565–2572.

Clopton, B. M., Corbett, S. S., and Spelman, F. A. (2003). "Sound processor for high-channel count implants," Conference on Implantable Auditory Prostheses, Pacific Grove, CA, August 2003, p. 103.

Cohen, L. T., Busby, P. A., and Clark, G. M. (**1996**). "Cochlear implant place psychophysics. II. Comparison of forward masking and pitch estimation data," Audiol. Neuro-Otol. **1**, 278–292.

Cohen, L. T., Sanders, E., and Clark, G. M. (**2001**). "Psychophysics of a prototype peri-modiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.

Dorman, M. F., Smith, L., and Parkin, J. L. (**1993**). "Loudness balance between acoustic and electric stimulation by a patient with a multichannel cochlear implant," Ear Hear. **14**, 290–293.

Eddington, D. K., Dobelle, W. H., and Brackman, D. E. (**1978**). "Auditory prostheses research with multiple channel intracochlear stimulation in man," Ann. Otol. Rhinol. Laryngol. **87**, suppl. 53, 5–39.

Hanekom, J. J., and Shannon, R. V. (**1998**). "Gap detection as a measure of electrode interaction in cochlear implants," J. Acoust. Soc. Am. **104**, 2372–2384.

Kileny, P. R., Zwolan, T. A., Telian, S. A., and Boerst, A. (**1998**). "Performance with the 20+2L lateral wall cochlear implant," Am. J. Otol. **19**, 313–319.

Kral, A., Hartmann, R., Mortazavi, D., and Klinke, R. (**1998**). "Spatial resolution of cochlear implants: the electrical field and excitation of auditory afferents," Hear. Res. **121**, 11–28.

Lehnardt, E., Gnadeberg, D., Battmer, R. D., and von Wallenberg, E. L. (**1992**). "Evaluation of different coding strategies and stimulation modes for the Nucleus multichannel cochlear implant," in *Transplants and Implants in Otology II*, edited by N. Yanaihara and J. Suzuki (Kugler, New York), pp. 283–290.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," 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.

Nelson, D. A., and Donaldson, G. S. (**2001**). "Psychophysical recovery from pulse-train forward masking in electric hearing," J. Acoust. Soc. Am. **109**, 2921–2933.

Nelson, D. A., and Donaldson, G. S. (**2002**). "Psychophysical recovery from single-pulse-forward masking in electric hearing," J. Acoust. Soc. Am. **112**, 2932–2947.

Nelson, D. A., and Fortune, T. W. (**1991**). "High-level psychophysical tuning curves: simultaneous masking by pure tones and 100-Hz-wide noise bands," J. Speech Hear. Res. **34**, 360–373.

Peterson, F. E. and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear Disord. **27**, 62–70.

Pfingst, B. E. and Xu, L. (**2004**). "Across-site variation in detection thresholds and maximum comfortable loudness levels for cochlear implants," J. Assoc. Res. Otolaryngol. **5**, 11–24.

Pfingst, B. E., Xu, L., and Thompson, C. S. (**2004**). "Across-site threshold variation in cochlear implants: relation to speech recognition," Audiol. Neuro-Otol. **9**, 341–352.

Pfingst, B. E., Zwolan, T. A., and Holloway, L. A. (**1997**). "Effects of stimulus configuration on psychophysical operating levels and on speech recognition with cochlear implants," Hear. Res. **112**, 247–260.

Pfingst, B. E., Franck, K. H., Xu, L., Bauer, E. M., and Zwolan, T. A. (**2001**). "Effects of electrode configuration and place of stimulation on speech perception with cochlear prostheses," J. Assoc. Res. Otolaryngol. **2**, 87–103.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hear. Res. **12**, 1–16.

Shannon, R. V. (**1990**). "Forward masking in patients with cochlear implants," J. Acoust. Soc. Am. **88**, 741–744.

Tong, Y. C., and Clark, M. (**1986**). "Loudness summation, masking, and temporal interactions for sensations produced by electric stimulation of two sites in the human cochlea," J. Acoust. Soc. Am. **79**, 1958–1966.

van den Honert, C., and Stypulkowski, P. H. (**1987**). "Single fiber mapping of spatial excitation patterns in the electrically stimulated auditory nerve," Hear. Res. **29**, 195–206.

von Wallenberg, E. L., Laszig, R., Gnadeberg, D., Battmer, R. D., Desloo-

vere, C., Kiefer, J., Lehnardt, E., and von Ilberg, C. (**1994**). "Initial findings with a modified Nucleus implant comprised of 20 active intracochlear and 2 extraccochlear reference electrodes," in *Advances in Cochlear Implants*, edited by I. J. Hochmair-Desoyer and E. S. Hochmair (Mantz, Vienna), pp. 186–192.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.

Zeng, F. G., and Shannon, R. V. (**1993**). "Loudness balance between electric and acoustic stimulation," Hear. Res. **60**, 231–235.

Zeng, F. G., and Shannon, R. V. (**1994**). "Loudness-coding mechanisms inferred from electric stimulation of the human auditory system," Science **264**, 564–566.

Zwolan, T. A., Kileny, P. R., Ashbaugh, C., and Telian, S. A. (**1996**). "Patient performance with the Cochlear Corporation '20+2' implant: Bipolar versus monopolar activation," Am. J. Otol. **17**, 717–723.

Numerical study of the effects of inferior and superior vocal fold surface angles on vocal fold pressure distributions

Sheng Li
The Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

Ronald C. Scherer
Department of Otolaryngology—Head and Neck Surgery, University of Cincinnati Medical Center, P.O. Box 670528, Cincinnati, Ohio 45242

MingXi Wan,[a] SuPin Wang, and HuiHui Wu
The Key Laboratory of Biomedical Information Engineering of Ministry of Education, and Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

(Received 30 December 2004; revised 16 February 2006; accepted 20 February 2006)

Vocal fold geometry plays an important role in human phonation. A wide range of inferior and superior vocal fold surface angles has been shown to be present during phonation [Nanayakkara, Master's thesis, Bowling Green State University, Bowling Green, OH (2005)]. This study explored how these angles affect pressure distributions on the vocal folds, and thus how they may affect phonation. The computational code FLUENT was used to obtain pressure distributions for laminar, incompressible flow. Eighteen inferior vocal fold surface angles and nineteen superior vocal fold surface angles were selected for three specific glottal angles, uniform 0°, convergent 10°, and divergent 10°. Minimal glottal diameter (0.01 cm), transglottal pressure (8 cm H_2O), and glottal entrance radius (0.15 cm) were held constant, and the glottal exit radius was constant for each glottal angle. Results suggest that the vocal fold surface pressures are independent of the inferior and superior vocal fold surface angles realistic for human phonation. These results suggest that, in contrast to the important effects of glottal entrance and exit radii, minimal diameter, and angle on intraglottal pressures, the inferior and superior vocal fold surface angles (excluding possible interactive effects with the false vocal folds) do not have an influence on the intraglottal pressures. [DOI: 10.1121/1.2186548]

PACS number(s): 43.70.Aj, 43.70.Bk [AL] Pages: 3003–3010

I. INTRODUCTION

The vocal fold surface is essentially three-sided. One side is the medial glottal surface that changes its orientation during phonation from convergent during glottal opening to divergent during glottal closing. A second side is the inferior surface of the vocal fold that leads from the tracheal wall upward and medially to the glottal entrance. The third side is the superior vocal fold surface that leads from the glottal exit to the lateral wall of the ventricle between the true and false vocal folds. Much has been said about the medial surfaces of the vocal folds and their shape change, but few studies have examined the inferior and superior vocal fold surfaces.

Recent studies by Agarwal *et al.* (2003) and Nanayakkara (2005) indicate that the inferior and superior vocal fold surfaces can take on a wide range of angles during phonation. These authors made measurements on laminagram x-ray tracings from studies by Hollien and Colton (1969) and Wilson (1972, 1976) to come to this conclusion. The laminagraphic images were obtained during phonation, and represent "averages" of the vocal fold contours over short intervals of time during the phonations. The Nanayakkara results suggested that when the data were combined across men and women, modal register and falsetto, louder and softer phonation, and singers and nonsingers, the average inferior vocal fold surface angle was 30.3°, with a two standard deviation range of 13.1° to 47.6°. The average for the superior vocal fold surface angle for similarly combined data was 10.2°, with a two-standard deviation range of −7.4° to 27.7°, where a positive angle refers to the superior vocal fold surface shaped downward laterally. Relative to physical and computer models of phonation, however, the inferior vocal fold surface angle usually has been fixed at some intermediate angle and the superior vocal fold surface angle fixed at some nominal small positive angle or zero degrees (e.g., Gauffin *et al.*, 1983; Pelorson *et al.*, 1994; Alipour *et al.*, 1996). Relative to changes of the inferior and superior angle during phonation, radiographic stroboscopy with excised canine larynxes (Saito *et al.*, 1981; Fukuda *et al.*, 1983) and particle tracking with excised canine larynxes (Baer, 1981) suggest considerable change of the superior vocal fold surface angle near the medial portion of the vocal fold (the lateral portion of the superior surfaces were not shown). This is similar to the results with distributed mass models (Titze,

[a]Electronic mail: mxwan@mail.xjtu.edu.cn

1981) and finite-element models (Alipour *et al.*, 2000), which also show motion of the superior (as well as inferior) vocal fold surface throughout the vibratory cycle. Prior studies suggest, therefore, that there are individual differences in the average phonatory superior and inferior vocal fold surface angles, as well as some change of these angles during the phonatory cycle.

Nevertheless, the relation between inferior and superior vocal fold surface angles and the pressures on the vocal folds for phonation apparently has not been studied empirically or computationally. If variation of the inferior and superior angles creates substantial changes to the pressure distributions on the vocal folds, the subsequent motion of the vocal folds and the shape and size of the glottal volume velocity signal may be significantly affected. Furthermore, physical and computer models for vocal fold motion would need to be extended to include these effects. Thus, the purpose of this study was to examine the effects on the intraglottal vocal fold surface pressures when the inferior and superior vocal fold surface angles were parametrically varied.

II. METHODS

The commercially available computational fluid dynamics package FLUENT (Fluent, Inc., Lebanon, NH) was used to obtain estimates of the normal pressures along the vocal fold surfaces as a function of the inferior and superior vocal fold surface angles. The code has been shown to produce pressure distributions for constant flows in laryngeal ducts that were highly similar to those found empirically using Plexiglas models of the larynx (Scherer *et al.*, 2001a, 2001b; Li *et al.*, 2006). The code is based on the control-volume technique and was used to solve the Navier-Stokes equations for laminar, incompressible airflow physics occurring inside two-dimensional symmetric laryngeal geometries. This use of FLUENT, then, is for constant shapes, not for vibrating vocal folds. Furthermore, the code as used does not produce turbulence. The fluid material was defined as air (density of 1.225 kg/m^3 and viscosity of 1.7894×10^{-5} kg/m s). The grids used in the present simulation contained both structured and unstructured meshes. The mesh size varied from 77×10^3 to 320×10^3 nodes and was made finer in regions where high gradients in pressure or velocity were expected. The boundary conditions of the model included an inlet pressure of 8 cm H_2O and outlet pressure of 0 cm H_2O. All computational runs were performed unsteady to obtain second-order solutions for momentum and pressure that had residuals of less than 10^{-4}. The flow field was assumed to be symmetric across the midline of the glottis in this study, and therefore only the half flow field was modeled.

In this study, three glottal angles, 10° divergent, 0° uniform, and −10° convergent, were used, as they represent the three basic glottal shapes typically expected in phonation, and the angle values are typically expected for most phonation in modal register (as suggested by normal videostroboscopy (Hirano and Bless, 1993) and schematics of vocal fold vibration (Schönhärl, 1960; Stevens and Klatt, 1974; Baer, 1975; Hirano, 1977; Cranen and de Jong, 2000), which show maximum diverging glottis angles from 17° to 67° and convergent glottal angles from 8° to 67°). The 18 inferior vocal fold angles (measured between a horizontal line and the inferior vocal fold surface line) were 87.5°, 85°, 82.5°, 80°, 75°, 70°, 65°, 60°, ***45***°, ***40***°, ***35***°, ***25***°, ***15***°, 10°, 5°, 0°, −5°, and −10°, where the bold and italicized values are within ±2 SDs in the work of Nanayakkara (2005). These were used for each glottal angle. The superior vocal fold surface angle was kept constant at 0° for these runs because this angle is not atypical (see below). These inferior vocal fold surface angles for the divergent glottal angle are shown in Fig. 1(a) and 2(a). Similarly, 19 superior vocal fold surface angles (measured between a horizontal line and the superior vocal fold surface line) were −85°, −80°, −75°, −70°, −65°, −60°, −45°, −35°, −25°, −15°, −10°, ***−5***°, ***0***°, ***5***°, ***10***°, ***15***°, ***25***°, 35°, and 45°, where again the bold and italicized values are within ±2 SDs in the work of Nanayakkara (2005). These also were used for each glottal angle. The inferior vocal fold surface angle was kept constant at 40° for these runs because the value is within the expected range for phonation (see above). A subset of the superior vocal fold surface angles for the convergent glottal angle is shown in Fig. 3(a). It also can be seen from Figs. 1(a), 2(a), and 3(a) that altering the inferior and superior angles changes the size and shape of the vocal fold.

The glottal exit radii were set at 0.108, 0.0987, and 0.0908 cm for the divergent 10°, uniform 0°, and convergent −10°, respectively, consistent with the design for the empirical model M5 (Scherer *et al.*, 2001b). The false vocal folds were not included. Held constant for all cases were the glottal entrance radius (0.15 cm), the minimal glottal diameter (0.01 cm), the anterior-posterior glottal length (1.2 cm), the inlet domain axial length (0.2 cm), the outlet domain axial length (0.5 cm) (see Fig. 4), and the transglottal pressure (8 cm H_2O). The axial glottal length was set at 0.30 cm for all cases, where this length was determined by the projection to the centerline of the center of the entrance radius circle and the projection to the centerline of a line tangent to the exit radius when pointing downstream (see Figs. 1–4).

The angles beyond the range expected for human phonation mentioned above may be unrealistic, but are important because of the suspicion that the intraglottal pressure effects may not be large for the normal range; thus, it would be important to determine when the pressure effects actually come into play. Also, using a wider than normal range of angles permits inquiry into atypical laryngeal geometric conditions.

III. RESULTS

A. Effects of changing the inferior vocal fold surface angle

Figure 1(b) shows the pressure distribution results for the inferior vocal fold surface angles for the 10° divergent glottal angle. The decrease in cross-sectional area leading up to the glottis caused most of the pressure drop due to the flow acceleration (Bernoulli). Near the glottal entrance, along the glottal entrance curvature, viscous and additional flow acceleration effects would decrease the pressures further. The minimum pressure and largest velocity occurred

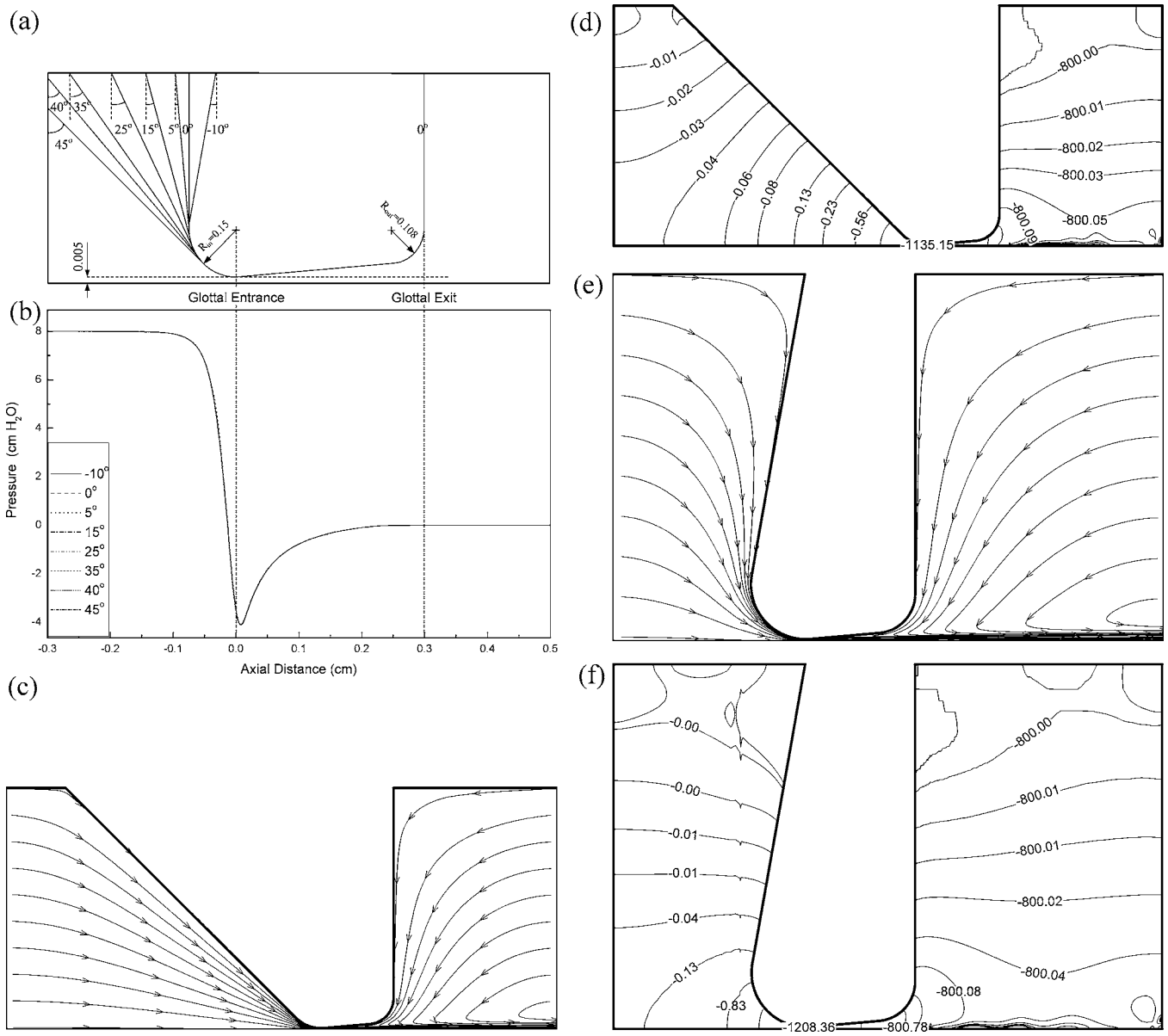


FIG. 1. Schematic of eight inferior vocal fold surface angles for the 10° divergent glottal angle, 0.01 cm minimal glottal diameter, and 8 cm H_2O transglottal pressure drop, and the corresponding pressure distributions (using FLUENT) on the vocal fold surfaces. (a) Outlines of the glottal configurations corresponding to the eight inferior vocal fold surface angles, with the superior vocal fold surface angle held constant at 0° (the entire vocal fold is not shown). (b) Corresponding pressure distributions, which overlap. (c) Streamlines for the 45° condition. (d) Pressure contours for the 45° condition. (e) Streamlines for the −10° condition. (f) Pressure contours for the −10° condition. Distances are in centimeters.

just past the minimum diameter cross section. Downstream of this location the pressure recovered (that is, increased to the prevailing supraglottal pressure, zero or atmospheric) due to the cross-sectional area increase.

The pressure distributions for the inferior vocal fold surface angles shown in Fig. 1(b) suggest that for angles considered realistic for human phonation, the vocal fold surface pressures are essentially identical (for the chosen entrance radius). How large must this angle be, assuming a constant glottal entrance radius, before significant changes to the pressure distributions appear? Figure 2(b) shows the results for inferior vocal fold surface angles between 60° and 87.5°. Differences appeared only for angles greater than 70°: the locations of the minimal diameters remained the same, the minimum pressures became less negative, and the pressure distribution on the inferior vocal fold surface decreased. The increase in minimal pressure for the largest angle shown was 13% of the transglottal pressure. These results suggest, then, that the pressure distributions on the vocal fold surfaces (inferior surface, glottal entrance, and intraglottal surface) for a constant glottal entrance radius for 10° divergent glottal angle remained approximately the same for inferior vocal fold surface angles between −10° and 70°.

It may appear curious that, as Fig. 1(b) indicates, the normal surface pressures are essentially identical for inferior vocal fold surface angles between 45° and −10°. Figures 1(c)–1(f) show the streamlines and pressure contours for the 45° and −10° geometries, respectively. These figures indicate that normal pressures are indeed almost identical because of the size of the entrance radius chosen for this study. A much

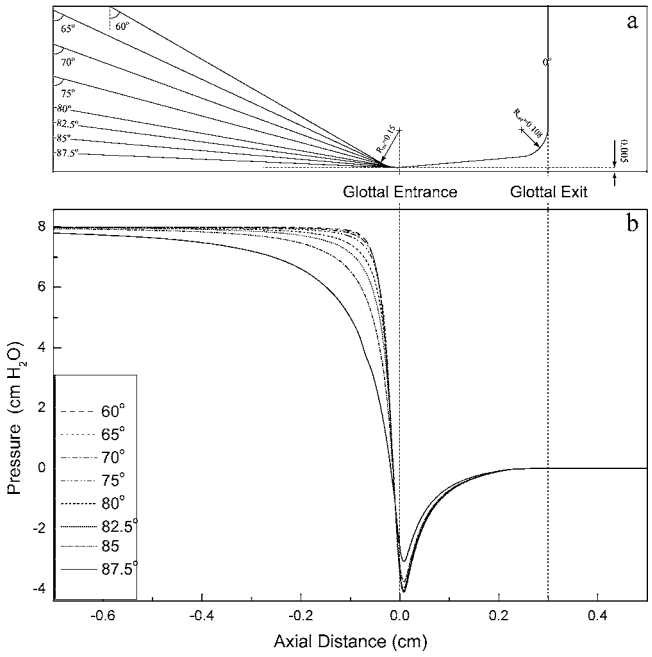


FIG. 2. Schematic of eight larger inferior vocal fold surface angles for the 10° divergent glottal angle, 0.01 cm minimal glottal diameter, and 8 cm H_2O transglottal pressure drop, and the corresponding pressure distributions (using FLUENT) on the vocal fold surfaces. (a) Outlines of the glottal configurations corresponding to the eight inferior vocal fold surface angles, with the superior vocal fold surface angle held constant at 0° (the entire vocal fold is not shown). (b) Corresponding pressure distributions. Distances are in centimeters.

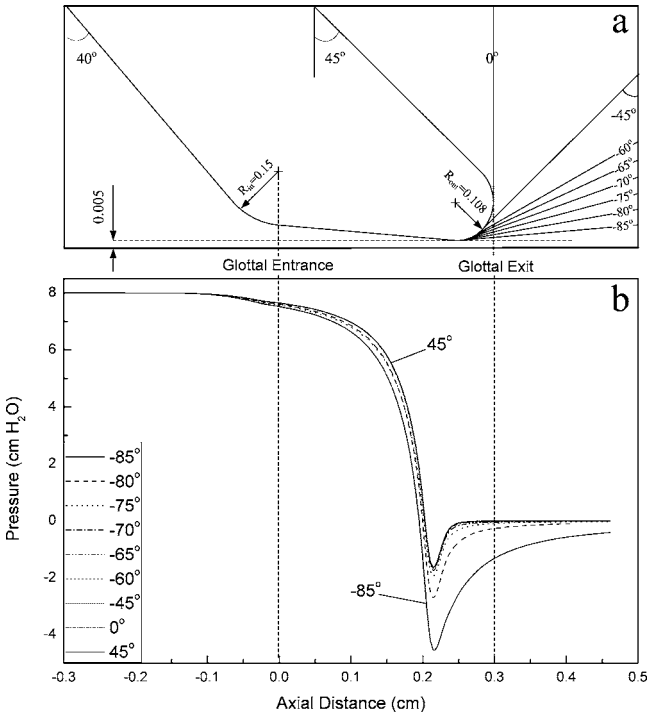


FIG. 3. Schematic of nine superior vocal fold surface angles for the 10° convergent glottal angle, 0.01 cm minimal glottal diameter, and 8 cm H_2O transglottal pressure drop, and the corresponding pressure distributions (using FLUENT) on the vocal fold surfaces. (a) Outlines of the glottal configurations corresponding to the nine superior vocal fold surface angles, with the inferior vocal fold surface angle held constant at 40° (the entire vocal fold is not shown). (b) Corresponding pressure distributions. Distances are in centimeters.

smaller glottal entrance radius might show more sensitivity to the inferior vocal fold surface angle near glottal entrance, but entrance radius was not considered as a parameter in this study (it was in the study by Li *et al.* (submitted), which shows that intraglottal pressures are highly sensitive to the entrance radius for divergent glottal angles).

Similar results for the pressures on the vocal fold surfaces were found for each of the other two glottal shapes, the uniform and 10° convergent glottal angles, for the constant glottal entrance and exit radii. That is, the minimal glottal pressures and the pressure distributions on the vocal fold surfaces were not altered until the inferior vocal fold surface angle was greater than 70°. The axial location of greatest pressure drop for both the uniform and the convergent glottis was near the exit of the glottis (as will be shown below for the convergent glottis).

B. Effects of changing the superior vocal fold surface angle

The superior vocal fold surface angle also appears not to influence the vocal fold surface pressures except for angles beyond those expected for human phonation. Figure 3(b) shows the results for the 10° convergent glottal angle for a constant glottal exit radius (the convergent angle was chosen for display because the large pressure changes near the glottal exit were thought to be potentially especially sensitive to changes in the superior vocal fold surface angle. The surface pressures for all positive angles were nearly identical, meaning that the vocal fold surface pressures were independent of positive superior vocal fold surface angles, until the superior vocal fold surface angle exceeded approximately −65°. The axial location of greatest pressure drop for each of the superior angles was constant (at approximately 0.21 cm axial distance in Fig. 3). As the superior vocal fold surface angle increased negatively beyond −65°, the minimum pressure became more negative and the pressure increase past that point became more gradual (because the superior surface acted like a diffuser, not an abrupt area enlargement), with slight influence on the intraglottal pressures (reduction in the pressure distribution).

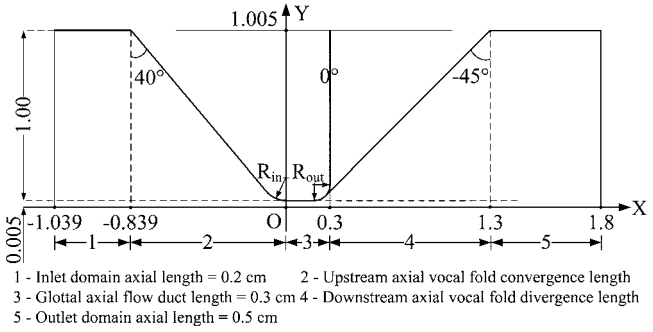


FIG. 4. Schematic of the glottal parameters and their dimensions for the 40° inferior vocal fold surface angle and −45° superior vocal fold surface angle. R_{in} was set at 0.15 cm, and R_{out} was set at 0.108, 0.0987, and 0.0908 cm for the divergent 10°, uniform 0°, and convergent −10°, respectively. The 0.005 cm shows only half of the minimum glottal diameter.

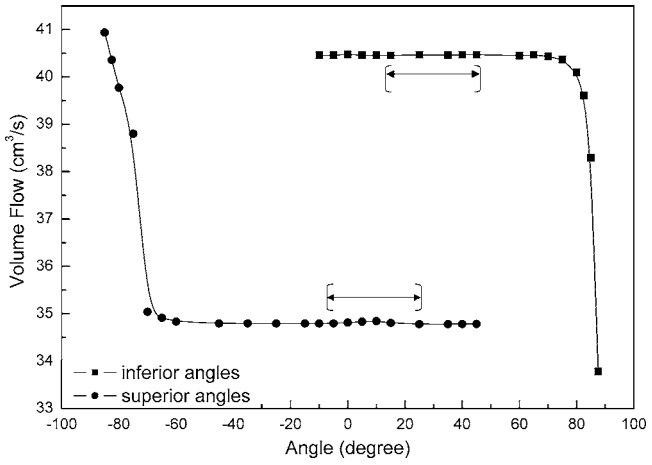


FIG. 5. Flow rates for all of the inferior and superior vocal fold surface angles used in this study. The glottal angle for the inferior angles was 10° divergent, and 10° convergent for the superior angles (0.01 cm minimal glottal diameter, 8 cm H_2O transglottal pressure drop). The brackets indicate the ±2 SDs for phonation given in Nanayakkara (2005).

Again, the trends for the results for the superior vocal fold surface angle were similar for the other two glottal angles, the uniform and 10° divergence. That is, the pressure distributions on the vocal fold were essentially identical for the range of 45° to −65° for all of the uniform and divergent glottis cases, and began to change slightly as the superior vocal fold surface angle increased negatively beyond −65°. The pressure distribution lowered along the supraglottal vocal fold walls as the negative superior angle increased negatively beyond −65°, the larger the negative superior angle, the larger this change. Because the changes were consistent, there is no need to show the figures for the uniform and divergent glottal conditions.

C. Effects of changing the superior and inferior vocal fold surface angles on glottal flow

The effect of the inferior and superior vocal fold surface angles on glottal flow can be seen in Fig. 5. The flow values are given for changes in both inferior and superior vocal fold surface angles, where the divergent 10° glottis was used for the inferior angle values, and the convergent 10° glottis for the superior angle values (the divergent glottis offers less flow resistance yielding higher flows in the bracketed angle regions). The changes in the flow for each series of values correspond to the angle values when changes in surface pressures are seen in the preceding figures. For the inferior vocal fold surface angle, the flow is constant within the region of the ±2 SDs expected for phonation noted earlier (shown by the double-headed arrows), decreasing only when the angle is above approximately 65°. Above this angle the flow resistance increases as the effective length of the inferior vocal fold surface increases. The opposite effect occurs for the superior vocal fold surface angle. The flow is unaffected within the range of ±2 SDs expected in phonation, but for angles more negative than approximately −65° the flow increases, indicating less flow resistance. This is due to the diffuser effect of the expanding supraglottal surfaces allowing pressure rise along them, and therefore less flow resistance for the same transglottal pressure condition.

D. Comparison of computational results to lossless Bernoulli equation

It is of some importance to explore the differences between the computational approach of FLUENT, which includes viscous effects, and the lossless Bernoulli equation. That is, it may be that portions of the pressure distributions are sufficiently identical to the predictions from the Bernoulli equation that the extra cost and sophistication of a computational approach such as FLUENT are not warranted.

For the steady, laminar, incompressible, and lossless airflow, Bernoulli's equation for any two cross sections in the model can be expressed as

$$P_1 + \frac{1}{2}\rho\frac{Q_1^2}{A_1^2} = P_2 + \frac{1}{2}\rho\frac{Q_2^2}{A_2^2}, \tag{1}$$

where P is pressure, ρ is the volume density of air (1.225×10^{-3} g/cm^3), and $(1/2)\rho(Q^2/A^2)$ is the kinetic energy per unit volume (or dynamic pressure). The flow rate, Q, is given by $Q=AV$, where A is the cross-sectional area in the model, and V is the average velocity.

For the models in this study, mass balance gives

$$Q_1 = A_1V_1 = Q_2 = A_2V_2.$$

The values of Bernoulli static pressure drops are therefore

$$P_{\text{static Bernoulli}} = \frac{1}{2}\rho(V_2^2 - V_1^2) = \frac{1}{2}\rho Q^2\left(\frac{1}{A_2^2} - \frac{1}{A_1^2}\right). \tag{2}$$

And, the Bernoulli dynamic pressure values are

$$P_{\text{dynamic Bernoulli}} = \frac{1}{2}\rho\frac{Q^2}{A_2^2}. \tag{3}$$

Four vocal fold surface angle conditions were used for divergent and convergent glottal angles to calculate the lossless Bernoulli pressure distributions, namely, 45° and 87.5° inferior surface angles for the 10° divergent glottal angle, and −45° and −85° superior surface angles for the 10° convergent glottal angle. These conditions were used with FLUENT in Fig. 1(a), Fig. 2(a), Fig. 3(a), and Fig. 4, with the transglottal pressure held constant at 8 cm H_2O and minimal glottal diameter of 0.01 cm. The corresponding volume flows for these four surface angles obtained using FLUENT were used in the Bernoulli equation. These were 40.47 and 33.78 cm^3/s for the 45° and 87.5° inferior surface angles, and 34.78 and 40.93 cm^3/s for the −45° and −85° superior surface angles, respectively (the flows are shown in Fig. 5).

Figures 6 and 7 show the static and dynamic pressure distributions for these four vocal fold surface angles. The figures show that there are large differences between the results of FLUENT and the lossless Bernoulli equation. It can be seen from Fig. 6(a) and Fig. 7(a) that (a) the FLUENT static pressure is lower than the corresponding Bernoulli static pressure; (b) the smaller the cross-sectional area in the model, the larger the pressure difference; and (c) more down-

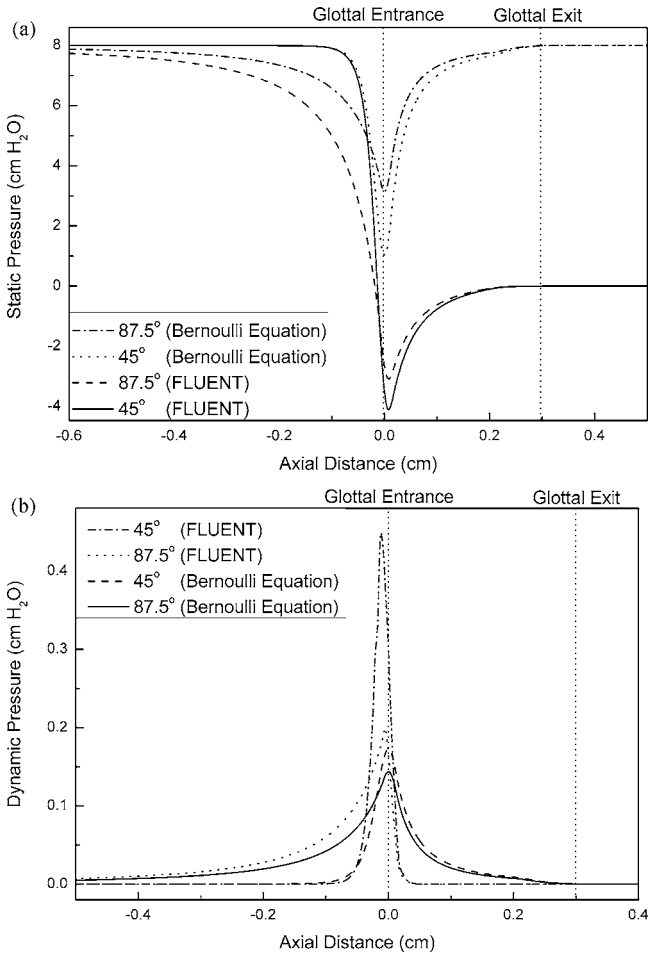


FIG. 6. Static and dynamic pressure distributions on the vocal fold surfaces for 45° and 87.5° inferior glottal angle for the 10° divergent glottal angle, 0.01 cm minimal glottal diameter, and 8 cm H_2O transglottal pressure drop. The data computed by both FLUENT and Bernoulli equation. (a) Static pressure distributions. (b) Dynamic pressure distributions.

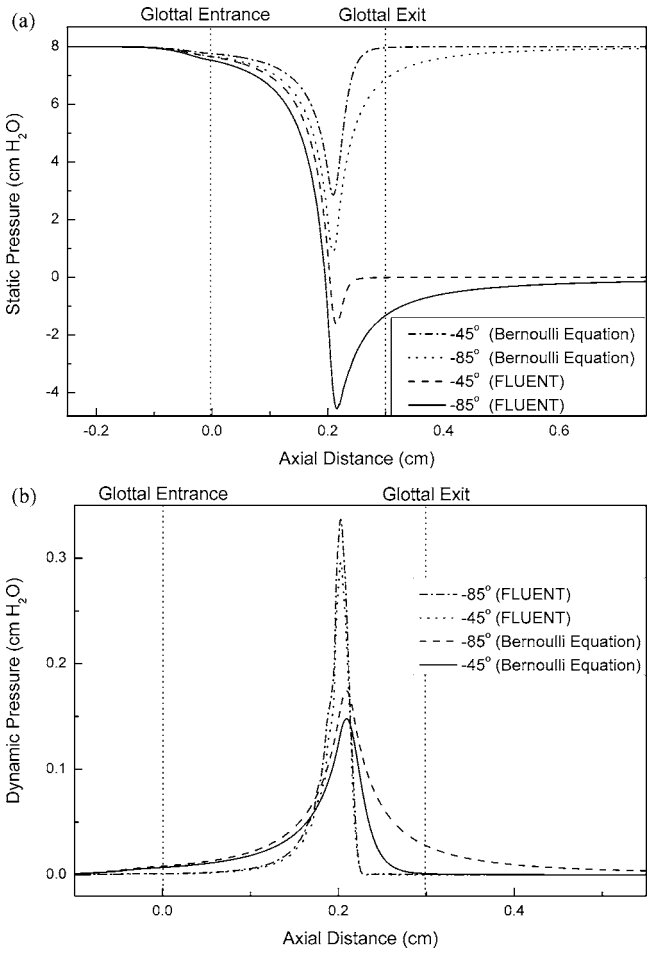


FIG. 7. Static and dynamic pressure distributions on the vocal fold surfaces for −45° and −85° superior glottal angle for the 10° convergent glottal angle, 0.01 cm minimal glottal diameter, and 8 cm H_2O transglottal pressure drop. The data computed by both FLUENT and Bernoulli equation. (a) Static pressure distributions. (b) Dynamic pressure distributions.

stream sections have larger pressure differences. For example, for the 45° inferior surface angle for the divergent case, the pressures begin to differ in an obvious manner upstream of the narrowest point. The difference is about 5 cm H_2O at the minimal diameter, and then, due to the lossless characteristic of the Bernoulli equation, differs by 8 cm H_2O at the exit. The Bernoulli equation does not predict drag losses.

The dynamic pressure distributions, which are shown in Fig. 6(b) and Fig. 7(b), suggest that the dynamic pressure values are relatively quite small compared to the static pressures. For example, for the 45° inferior surface angle case (divergent glottis) for FLUENT, the mean range of static pressure is from 8 cm H_2O to approximately −4 cm H_2O, whereas the dynamic pressure range is from approximately 0.44 cm H_2O to 0 cm H_2O. The peak dynamic pressure according to the Bernoulli equation is about one-third as much. The figures also suggest that the dynamic pressure is more sensitive to the minimal diameter region in FLUENT than for the Bernoulli equation because the former creates not only a higher peak dynamic pressure near the minimal diameter, but also lower values just away from the minimal diameter location.

IV. DISCUSSION

The results of this computational modeling study suggest that, for any specific glottal angle with specified glottal entrance radius, glottal exit radius, and minimal glottal diameter, the pressures on the vocal fold surfaces for a constant transglottal pressure are essentially nearly identical over a wide range of inferior and superior vocal fold surface angles that includes the angles which are realistic for human phonation. Therefore, whereas intraglottal pressures may be highly sensitive to the radius of the glottal contours at entry (for the divergent glottis) and exit (for the convergent glottis) (Scherer *et al.*, 2001a; Li *et al.*, submitted), as well as the glottal angle and minimal glottal diameter, these pressures are not sensitive to the steepness of either the inferior or the superior vocal fold surfaces. This may help "explain" the wide range of angles encountered in human phonation data for the inferior vocal fold surface angle. That is, a variable that has little sensitivity to the all-important vocal fold driving forces may vary over a wide range, as appears to be the case for the inferior vocal fold surface angle. Indeed, the results of Nanayakkara (2005) suggest that the inferior vocal fold surface angle in particular is highly person specific.

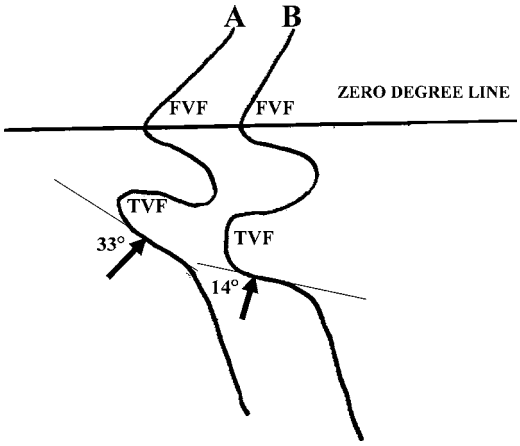


FIG. 8. (A) and (B) each represent an outline of one side of a larynx airway. The two outlines are from different subjects from the corpus of x-ray laminagram tracings of phonation by humans in Nanayakkara (2005). The outlines suggest the potentially wide range of inferior vocal fold surface angles, and therefore their corresponding resultant force vectors to move the vocal folds upward and laterally. The "FVF" and "TVF" in the figure mean "false vocal fold" and "true vocal fold," respectively.

The potential interaction of the superior vocal fold surface angle with the false vocal folds and the possible resonance utility of the ventricles (Agarwal, 2004; Zhang *et al.*, 2004), and the exit glottal jet flow behavior not given by the FLUENT approach, may make the superior vocal fold surface more sensitive to phonatory aerodynamics and acoustics not pursued in this study, although as shown here perhaps not sensitive *itself* to altering pressures on the vocal fold surfaces.

There may well be anatomical and surgical cases where the inferior and superior vocal fold surfaces take on the larger angles studied here, whereby the pressures would be altered from normal expectation. These should be taken into consideration for their potential effects on phonation.

As the figures of this report (except Figs. 5–7) suggest, when the inferior and superior vocal fold surface angles take on different values, the size and shape of the vocal fold itself changes. Thus, even though the surface pressures may not differ, the biomechanical vocal fold tissue characteristics, mass and shape in particular, will alter. With pressures that do not change when inferior and superior vocal fold surface angles change, the altered vocal fold may nonetheless take on different two-dimensional oscillatory behaviors. In particular, one might expect a greater vertical force on the vocal fold when the inferior vocal fold surface angle is smaller (more horizontal), as suggested in Fig. 8, where two x-ray laminagraphic tracings [from the corpus used in the Nanayakkara (2005) study] are shown, one having an inferior vocal fold surface angle of 14°, the other 33°. Future research modeling should alter the inferior and superior vocal fold surface angles to determine their effects on vocal fold oscillation and the resultant volume velocity signal.

The viscous and kinetic pressure losses are more realistically evident from the FLUENT computational approach than from the lossless Bernoulli equation. The Bernoulli equation underpredicts intraglottal pressure changes and should be used with great caution in phonatory modeling.

The modeling here was of the half-field type, forcing symmetry of the laryngeal duct, flows, and pressures. Assuming that the inferior vocal fold surface angle does not change appreciably during a cycle of phonation, the results appear to suggest that if there were asymmetries (right vocal fold, left vocal fold differences) in the inferior vocal fold surface angle, there most likely would *not* be significant intraglottal pressure changes due to that asymmetry *per se*, leaving asymmetries of the glottal shape (oblique glottal angles, contour differences due to lesions) more responsible for intraglottal pressure differences (Scherer *et al.*, 2001b, 2002). However, if the asymmetries were extreme enough, arising for example from edema, webbing, or lesions on the inferior surface of the vocal folds, the geometric alterations may be severe enough to alter the driving pressures along the vocal folds, and should be studied for this effect; the change in biomechanical characteristics, if involved with the tissue in motion, still may have a more profound effect on vocal fold oscillation. Because the superior vocal fold surface angle is part of the ventricle, it is more difficult to reach the same conclusion that the resulting vocal fold surface pressures are independent of the superior vocal fold surface angle throughout a phonatory cycle; the results here only suggest that if the false vocal folds were not present, the superior angle itself would not affect the vocal fold surface pressures.

V. CONCLUSIONS

A wide variety of inferior and superior vocal fold surface angles for uniform, 10° convergent, and 10° divergent glottal angles was explored separately in this study by using the computational fluid dynamics code FLUENT. The results suggest that the intraglottal pressures are essentially independent of realistic inferior and superior vocal fold surface angles for each of the glottal angle cases. However, the effects on vocal fold motion as a result of altering vocal fold biomechanical characteristics because of changes in the inferior and superior vocal fold surface angles needs study. For inferior vocal fold surface angles greater than approximately 70°, and for superior vocal fold surface angles lower negatively than about −65°, the pressures on the vocal fold surfaces began to alter, suggesting that these effects would be of importance to vocal fold kinematics and aerodynamics for a variety of abnormal and medical/surgical considerations.

Although the results of this study appear broadly applicable to aerodynamic and modeling concerns of the vocal folds and phonation, the limitations of the modeling here (laminar not turbulent or transitional code, the lack of the presence and effects of the false vocal folds, the assumption of symmetry, the use of a limited range of intraglottal angles, and only one minimal diameter and transglottal pressure) suggest a conservative acceptance of the results. Further study of a more complete model would more accurately elucidate the phonatory effects of the inferior and superior vocal fold surfaces.

ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (Grant Numbers 30070212 and

69925101) and the National Institutes of Health (Grant Number 2 R01 DC03577). We would like to thank Bogdan Kucinschi for his computational help.

Agarwal, M. (**2004**). "The false vocal folds and their effect of translaryngeal airflow resistance," Ph.D. dissertation, Bowling Green State University, Bowling Green, OH.

Agarwal, M., Scherer, R. C., and Hollien, H. (**2003**). "The false vocal folds: Shape and size in frontal view during phonation based on laminagraphic tracings," J. Voice **17**, 97–113.

Alipour, F., Berry, D. A., and Titze, I. R. (**2000**). "A finite-element model of vocal fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.

Alipour, F., Fan, C., and Scherer, R. C. (**1996**). "A numerical simulation of laryngeal flow in a forced-oscillation glottal model," Comput. Speech Lang. **10**, 75–93.

Baer, T. (**1975**). "Investigation of phonation using excised larynxes," Ph.D. thesis, MIT, Cambridge.

Baer, T. (**1981**). "Observation of vocal fold vibration: Measurement of excised larynges," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo), pp. 119–133.

Cranen, B., and de Jong, F. (**2000**). "Laryngostroboscopy," in *Voice Quality Measurement*, edited by R. D. Kent and M. J. Ball (Singular, San Diego), pp. 257–267.

Fukuda, H., Saito, S., Kitahara, S., Isogai, Y., Makino, K., Tsuzuki, T., Kogawa, N., and Ono, H. (**1983**). "Vocal fold vibration in excised larynges viewed with an x-ray stroboscope and an ultra-high-speed camera," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 238–252.

Gauffin, J., Binh, N., Ananthapadmanabha, T. V., and Fant, G. (**1983**). "Glottal geometry and volume velocity waveform," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. M. Bless and J. H. Abbs (College-Hill, San Diego), pp. 194–201.

Hirano, M. (**1977**). "Structure and vibratory behavior of the vocal folds," in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and F. S. Cooper (University of Tokyo Press, Tokyo), pp. 13–27.

Hirano, M., and Bless, D. M. (**1993**). *Videostroboscopic Examination of the Larynx* (Singular, San Diego).

Hollien, H., and Colton, R. H. (**1969**). "Four laminagraphic studies of vocal fold thickness," Folia Phoniatr. **21**, 179–198.

Li, S., Scherer, R. C., Wan, M., Wang, S., and Wu, H. (**2006**). "The effect of glottal angle on intraglottal pressure," J. Acoust. Soc. Am. **119**, 539–548.

Li, S., Scherer, R. C., Wan, M., Wang, S., and Wu, H. (submitted). "The effect of entrance radii on intraglottal pressure distributions in the divergent glottis," J. Acoust. Soc. Am.

Nanayakkara, I. (**2005**). "Vocal fold superior and inferior angles measured from radiographic images," Master's thesis, Bowling Green State University, Bowling Green, OH.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasisteady flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Saito, S., Fukuda, H., Isogai, Y., and Ono, H. (**1981**). "X-ray stroboscopy," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo), pp. 95–106.

Scherer, R. C., De Witt, K. J., and Kucinschi, B. R. (**2001a**). "The effect of exit radii on intraglottal pressure distributions in the convergent glottis (L)," J. Acoust. Soc. Am. **110**, 2267–2269.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kuchinschi, B., and Afjeh, A. (**2001b**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Scherer, R. C., Shinwari, D., De Witt, K. J., Zhang, C., Kucinschi, B. R., and Afjeh, A. A. (**2002**). "Intraglottal pressure distributions for a symmetric and oblique glottis with a uniform duct (L)," J. Acoust. Soc. Am. **112**, 1253–1256.

Schönhärl, E. (**1960**). *Die Stroboskopie in der praktischen Laryngologie* (Thieme, Stuttgart).

Stevens, K. N., and Klatt, D. H. (**1974**). "Current models of sound sources for speech," in *Ventilatory and Phonatory Control Systems*, edited by B. Wyke (Oxford University Press, Oxford), pp. 279–292.

Titze, I. R. (**1981**). "Biomechanics and distributed-mass models of vocal fold vibration," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo) pp. 245–270.

Wilson, J. E. (**1972**). *A study of the conformation of the laryngo-pharynx in singers during alteration of intensity, frequency, and vowel*, D.Music thesis, Indiana University, Bloomington [Indiana University Library MZ1. W748; ProQuest Dissertation and Theses Publication Number AAT 7223514].

Wilson, J. E. (**1976**). "Variations of the laryngo-pharynx in singing," The NATS Bulletin, December, 20–22, 31.

Zhang, Z., Mongeau, L., Frankel, S., Thompson, S., and Park, J. (**2004**). "Sound generation by steady flow through glottis-shaped orifices," J. Acoust. Soc. Am. **116**, 1720–1728.

An experimental analysis of the pressures and flows within a driven mechanical model of phonation

Bogdan R. Kucinschi and Ronald C. Scherer
Department of Otolaryngology-Head and Neck Surgery, University of Cincinnati Medical Center, P. O. Box 670528, Cincinnati, Ohio 45267-0528

Kenneth J. DeWitt and Terry T. M. Ng
Mechanical, Industrial and Manufacturing Engineering Department, University of Toledo, 2801 West Bancroft, Toledo, Ohio 43606

(Received 25 July 2005; revised 17 February 2006; accepted 17 February 2006)

The production of voice is related to the flow of air through the glottis, whose time-dependent shape is defined by the motion of the vocal folds and the translaryngeal pressure. A scaled dynamically similar experimental apparatus that mimics the motion of the vocal folds was designed and built, such that both the glottal diameter and glottal angle change during a motion cycle. This motion is more realistic than in other reported dynamic models. The motion of the folds can be driven at different frequencies. The glottal flow takes place at a constant inlet pressure, mimicking the lung pressure. The transglottal pressure difference and flow rate were measured over the motion cycle. Satisfactory agreement was obtained for identical cases by numerically solving the two-dimensional, incompressible Navier-Stokes equations. Both experimental and numerical data showed that the glottal flow rate and transglottal pressure were affected by the oscillation frequency of the vocal folds. Flow visualization showed that the glottal flow patterns, which are a potential source of aero-acoustic sound, are influenced by the oscillation frequency. However, glottal flow resistance depended to a lesser extent on vocal fold oscillation frequency for the portion of the cycle when the glottis was divergent. [DOI: 10.1121/1.2186429]

PACS number(s): 43.70.Bk, 43.70.Aj [BHS] Pages: 3011–3021

I. INTRODUCTION

The production of voice, or phonation, is closely related to the oscillation of the vocal folds that induces a modulation of the glottal air flow. The oscillation of the folds is a quasi-periodic process, the characteristics of which (i.e., range of motion and frequency) depend on the interaction between fluid and structure (i.e., vocal folds). The average oscillation frequency is approximately 125 Hz for males and 200 Hz for female speakers (Titze, 1994).

The relationship between the transglottal pressure drop and volumetric flow rate is important relative to the oscillation of the vocal folds. A significant number of studies, both theoretical and experimental, have been dedicated to the investigation of the relationship between pressure and flow equations for situations relevant to speech production. Some theoretical models were shown to predict satisfactorily the flow-induced oscillations of the vocal folds. Relatively simple theoretical mechanical models, coupled with simplified flows, namely the two-mass (e.g., Ishizaka and Flanagan, 1972) and the three-mass model (Story and Titze, 1995), are to be listed in this category, although the current trend is to employ more complex tissue models (e.g., Alipour and Scherer, 2000; Alipour *et al.*, 2000). However, the complexity of the motion of the vocal folds has deterred the use of moving experimental models. Static models were used instead, with shapes that correspond to different position stages of oscillation of the folds (e.g., van den Berg *et al.*, 1957; Gauffin *et al.*, 1983; Scherer *et al.*, 1983; Alipour *et al.*, 1996a). The experimental data obtained with these physical models were used to validate computational models of glottal flow (e.g., Guo and Scherer, 1993). Important pressure-flow equations were thus determined for a multitude of significant static cases. Most of the static studies were performed for symmetric glottal configurations, although asymmetric configurations have also been investigated (Scherer *et al.*, 2001). The pressure and velocity profiles were investigated for the case of a hemilarynx by Alipour and Scherer (2002).

There is a limited amount of reported experimental work investigating the unsteady phenomena that occur during phonation. The work of Pelorson *et al.* (1994, 1995) investigated both experimentally and theoretically the unsteady flow through static models of the glottis when the flow was driven by an abrupt pressure variation. They used the measured glottal pressures and velocities to validate a flow model based on the unsteady Bernoulli equation and boundary layer equations. The theoretical model showed good agreement with the experimental data for the transient period. The flow model was coupled with a modified two-mass model (Pelorson *et al.*, 1994), and the predicted time derivative of glottal volume velocity was in good agreement with the experimental data (obtained by inverse filtering for real speech). There are few studies reported that utilize dynamic vocal fold models. Deverge *et al.* (2003) reported a dynamic vocal fold model which is composed of a set of rigid vocal fold replicas, one of which was given an oscillatory motion; the model allows for the collision of the vocal folds. The experimental results were compared against the theoretical predictions of

simplified flow models (steady Bernoulli equation corrected for friction, boundary layer solution for steady flows, and Reynolds equation). It was found that the effects of flow unsteadiness are not important for the studied case, such that the steady theoretical models could satisfactorily predict the pressure at a certain location inside the glottis. The unsteady Reynolds equation was found to be useful only close to collision, when the viscous effects dominate the flow. This model does not consider the vocal tract length, and the overall pressure was not kept constant. A dynamic model that uses a set of elastic replicas has been reported by Zhang *et al.* (2002), for which the acoustic pressure of waves radiated upstream and downstream of the orifice was measured. The acoustic source was modeled based on the steady Bernoulli equation, thus neglecting the unsteady phenomena. It was found that the experimental data were well predicted by using the quasi-steady approximation (i.e., assuming that the flow through the time-varying orifice can be regarded as a sequence of steady flows through stationary orifices of identical configuration). The dipole sources were identified as the dominant sound sources for pulsating flows at frequencies typical for phonation. A simple mechanical, real-life size model of the vocal folds has also been reported by Barney *et al.* (1999). In this model a pair of shutters, driven by a vibration generator, mimicked the vocal folds. The flow in the apparatus was provided by a constant flow source. The acoustic signal at the exit of the vocal tract was monitored by a pair of microphones and was used to estimate the pressure wave amplitudes in the shutter plane. The velocity at certain points in the duct was measured by hot-wire anemometry. The experimental results were compared with the predictions of the plane-wave classic acoustic theory of speech production. It was revealed that the plane wave acoustic theory failed to correctly predict the radiated sound field and that there was an additional acoustic source associated with the vortices generated at the shutter exit. This indicates that the flow patterns downstream of the vocal folds may be relevant to sound production. Recently, Thomson *et al.* (2005) reported a dynamic self-oscillating model that does not include the vocal tract, and a complementary numerical model in order to investigate the energy transferred from the flow to the vocal folds during one oscillation cycle. The motion of the folds is dependent on the properties of the material used to fabricate the model vocal folds, which was assumed to be reasonably similar to human tissue. It was found that greater pressures occur inside the glottis during the time of positive energy flow to the solid rather than during the time of negative energy flow. This asymmetry in the energy flow maintained the self-sustained oscillations of the vocal folds.

Prompted by the lack of a dynamic mechanical model that is able to realistically mimic the time-dependent glottal shape, a more realistic mechanical model was built. The oscillation of the vocal folds was forced [a similar approach was used by Alipour *et al.* (1996b), in their numerical study of the glottal flow]. The motion of the vocal folds is approximated by a plan-parallel motion, composed of translations and rotations. The generated motion is compatible with two of the lower order eigenmodes of the vocal folds. Titze and Strong (1975) analyzed the vocal folds as an elastic, compressible medium and obtained the lowest-order vibration modes that are excited in most cases of phonation. Berry and Titze (1996) obtained the three lowest order eigenmodes for the vocal folds considered as both a compressible and, more realistically, an incompressible tissue. The z-10-like (vertical) mode does not play a critical role in vocal fold dynamics. The other two modes, however, can be used to describe the vocal fold motion: x-11, which was shown to produce divergent-convergent shapes during oscillation, and x-10, which describes the medial-lateral movement of the folds. The surface of the vocal folds, between the upper and the lower curved margins, is approximately straight. It was also shown that the eigenfrequencies for these two modes are similar, which facilitates a 1:1 entrainment of the modes during oscillation. These considerations justify the design of a dynamic model, which mimics the motion of the vocal folds with a straight surface, and changes both the glottal diameter and glottal angle to obtain divergent-convergent configurations. The glottal angle can be controlled and synchronized with the glottal diameter. The flow is driven by a constant overall pressure that mimics the lung pressure, which is reported to be relatively constant during phonation. The dynamic model allows for the measurement of the time-dependent flow rate and transglottal pressure for a range of overall (lung) pressures and driving frequencies. The walls of the apparatus are transparent, such that flow visualization inside the glottis can be performed (to be more extensively reported in a future paper). Flow visualization is necessary to establish the character of the flow inside and downstream of the glottis.

The measured values of flow rates and transglottal pressures are compared with the results of numerical simulations performed using the CFD code FLUENT, which allows solution of the Navier-Stokes equations governing the flow through the moving glottis. The results of the numerical simulations are regarded as more reliable than those that can be obtained using simplified flow models based on the Bernoulli equation or boundary-layer approximation.

II. EXPERIMENTAL APPARATUS

A schematic of the experimental apparatus used to obtain the experimental results presented in this paper is shown in Fig. 1. The apparatus is a dynamic mechanical model of the vocal folds, which can replicate the time-dependent shape of the glottis. The replicas of the vocal fold surfaces were made of a deformable material (latex). The latex folds are made of strips of Thera-Band exercise band (Thera-Band Gmbh), with an initial thickness of 0.79 mm $\left(\frac{1}{32}\text{ in.}\right)$. Compared to the real-life dimensions, the model is scaled up by a factor of 7.5.

The difference between the dimensions of the real-life vocal folds and those of the model involves a difference between the parameters related to the flow. The relationship between different quantities in the real-life vocal folds and the model is established using dynamic similitude. Dynamic similitude is used to experimentally study the physical phenomena occurring in a system in which it is too difficult to use the real-life geometry for measurements due to either the

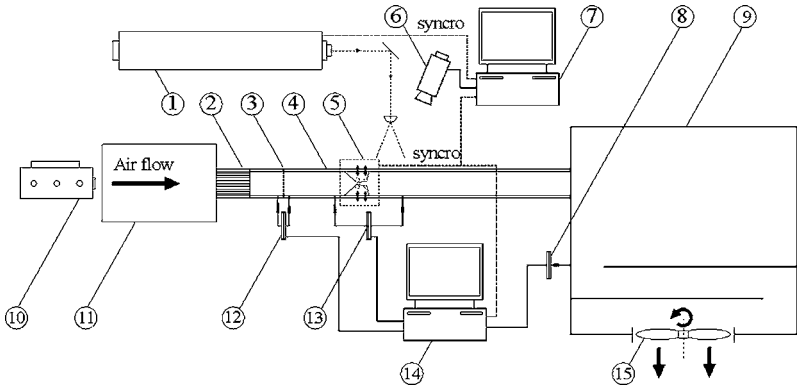


FIG. 1. Schematic of the experimental device: (1) pulse laser, (2) flow straightening section, (3) flowmeter net, (4) plexiglass tunnel, (5) dynamic vocal folds, (6) digital camera, (7) computer for flow visualization, (8) pressure sensor for reservoir, (9) constant pressure reservoir, (10) smoke generator, (11) entry reservoir, (12) pressure sensor for flow measurement, (13) transglottal pressure sensor, (14) computer for data acquisition, and (15) exhaust fan.

magnitudes of the flow variables or the geometric size. It is also used in cases where the governing equations (i.e., continuity, Navier-Stokes, and energy equations) cannot be solved for the original system. This concept also involves geometric similitude, where all geometric dimensions are scaled. The important dimensionless groups are obtained by using the dimensionless Navier-Stokes equations along with the dimensionless boundary and initial conditions (Batchelor, 1967). All the important dimensionless groups (Reynolds number, pressure coefficients, Strouhal number) must have the same value in the real system and the model system. Dynamic and geometric similitude ensure that the same dimensionless pressure and velocity distributions will occur in each system. For the system of concern in this study, geometrical similarity, the Reynolds number, the dimensionless pressure, and the Strouhal number must be kept the same. The choice of a scale factor of 7.5 was made in order that the apparatus could be accurately constructed. The larger dimensions also helped to insert the driving mechanism of the vocal folds, as well as to perform flow visualizations. Since enlarging the geometry causes a reduction in magnitude in model velocities and pressures, it is necessary to ensure that the flow variables can be accurately measured in the scaled system. Scaling ratios from 5 to 7.5 could have been used in the present work.

The Reynolds number is defined as $\mathrm{Re}=D_g U/\nu$, where D_g is the minimal glottal diameter, $U=Q/D_g W$ is the average flow velocity in the cross section of minimum diameter, Q is the volumetric flow rate at inlet, W is the anterior-posterior dimension of the vocal fold, 1.33 cm real-life (corresponding to a width of 10 cm in the physical model), and $\nu=15\times10^{-6}$ m^2/s is the kinematic viscosity of air. The range of Reynolds numbers for the reported experiments is 400–1800. The equality of the Reynolds number between the real-life case and the model requires that the velocities in the model must be 7.5 times smaller than in real-life, while the volume flow rates are 7.5 times larger. The equality of pressure coefficients leads to the pressures in the model being 7.5^2 times smaller than in real-life, where the pressure coefficient is defined as $\Delta p/(0.5\rho u^2)$; here p is the pressure, ρ is the density of air, and u is the fluid velocity. The Strouhal number is defined as $\mathrm{St}=fD_g/u$, where f is the oscillation frequency of the vocal folds (on average around 100 Hz for males and 200 Hz for females). Equality of the Strouhal number requires that the model frequency be 7.5^2 times smaller than the real-life frequency. In order to avoid confusion, all of the quantities reported in the following sections will be expressed in real-life units.

The total length of the experimental device is similar to that of the human vocal tract, which allows for the presence of the inertance effects of the air in the sub- and supra-glottal segments in the vocal tract. The length of the vocal tract downstream of the vocal folds (i.e., between the vocal folds and the exit reservoir) is 14.8 cm real-life, while the length of the trachea, between the entry reservoir (corresponding to the lungs) and the vocal folds, is 3.9 cm real-life.

The motion of the vocal folds is driven by a computer-controlled mechanism. The driving mechanism is composed of two identical lever mechanisms, each powered by a stepper motor (450SM-B3-HM, Aerotech Inc.). The motion of the two motors is coupled by software, allowing for flexibility in prescribing the glottal motion. Each lever mechanism drives a set of cylinders (see Fig. 2), which helps shape the elastic (deformable) material into a controlled configuration. The thickness of the deformable material (i.e., 0.79 mm) was taken into account to determine the geometrical configuration of the glottis. Repeated tests showed that the variation of thickness along a motion cycle led only to negligible alterations of the glottal configuration, as compared to an ideal material whose thickness is constant.

Both the upper and lower sets of cylinders are moved symmetrically with the centerline. The symmetry of the mechanism was verified on zoomed-in pictures with a camera mounted above the apparatus (a graded plate of Plexiglas

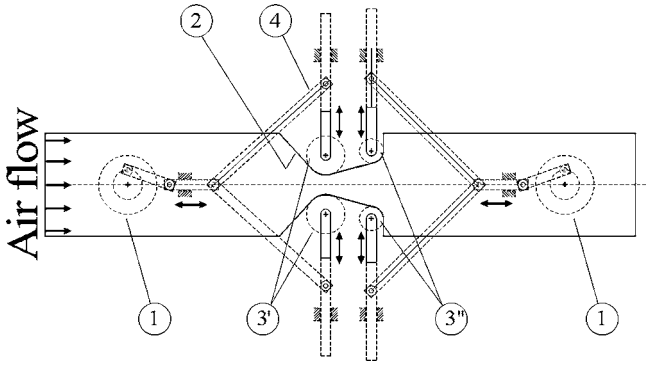


FIG. 2. Schematic of the driving mechanism: (1) stepper motors, (2) deformable material, (3′) upstream cylinders, (3″) downstream cylinders, and (4) levers.

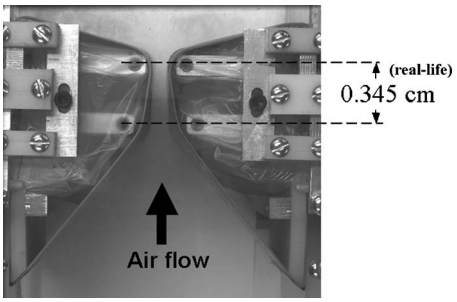


FIG. 3. Top view of the dynamic vocal folds.

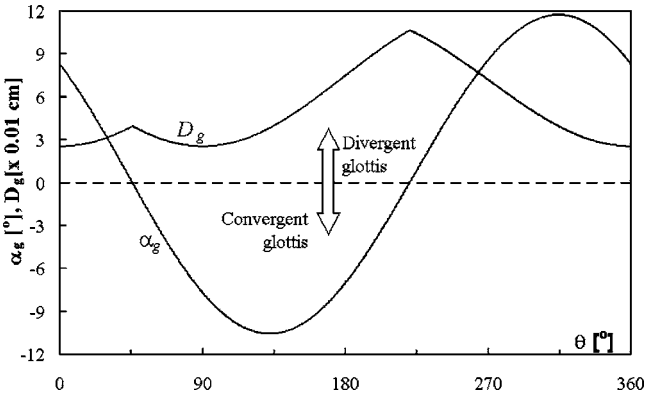


FIG. 5. Glottal diameter and glottal angle during an oscillation cycle.

was installed on top of the mechanism, with a clearly marked centerline). Due to the high stiffness of the mechanism, spring preloaded action levers, and low forces induced by the latex band, the measured double-sided uncertainty of the minimum diameter (measured at 0, 90, 180 and 270°) was less than 0.076 mm (0.003 in.). The top-view close-up of the vocal folds is presented in Fig. 3. The radii shown here include the contribution of the thickness of the deformable material. For this experiment, the mechanism was set such that the motion of the downstream set of cylindrical deformers lags the upstream set by 90°. This value is consistent with the phase delay between two flesh points located at the upper and lower margins of the vocal folds, measured by Titze *et al.* (1993) on excised canine larynges. Also, Berry *et al.* (1994) showed that a 90° phase difference exists between the two dominant eigenmodes of the vocal folds, i.e., an x-11-like mode, responsible for the alternately divergent-convergent shapes, and an x-10-like mode, responsible for the lateral motion of the vocal folds. The resulting motion mimics the motion of the vocal folds during an oscillation cycle (Hirano, 1977; Hofmans and Groot, 2003). For reasons dependent on the design of the driving mechanism, the beginning of a cycle corresponded to the upstream set of deformers being in the closest position (i.e., smallest glottal gap), with the glottis in a divergent posture. Conventionally, the start of the oscillation cycle is usually depicted when the folds are located in the closest position (normally in contact, i.e., minimum gap is zero) just before glottal opening, and the glottal angle is convergent, which means that there is a 90° difference in the reference (driving) angle between the mechanism and the real-life vocal folds. Figure 4 presents certain positions of the vocal folds versus the driving angle, θ, which is the angle of rotation of the motor driving the upstream set of deforming cylinders. The resulting motion cycle is similar to the schematic presented by Hirano (1977). The positive glottal angles correspond to the divergent shapes, while the negative ones represent convergent shapes. One restriction of the current layout of the apparatus is that it does not allow for the contact of the latex vocal folds, since the collision may result in damage to the mechanism. Thus, the minimum glottal gap was set to a positive value. The values of the minimal glottal gap (in cm real-life) are also given in Fig. 4. The smallest glottal gap was 0.025 cm, and the largest 0.104 cm, throughout the cycle. The variation of glottal gap and glottal angle during one cycle of oscillation is shown in Fig. 5. It is noted that the minimal glottal gap is the same concept as the projected glottal diameter or area in human phonation.

The variables of control are the amplitude of motion of the vocal folds (by setting up the amplitude of motion of the cylinders 3′ and 3″ in Fig. 2), the phase between the upstream and downstream margins of the vocal folds, the frequency of oscillation, and the overall driving pressure (corresponding to lung pressure). One important feature of the flow through the model is that it is driven by a constant pressure source, which models the quasi-constant lung pressure during phonation. The alternate condition, namely constant flow rate through the model, is more convenient to impose from a design standpoint, but it does not reflect physical reality as well. The apparatus was designed such that the flow resistance of the glottis is much larger than the

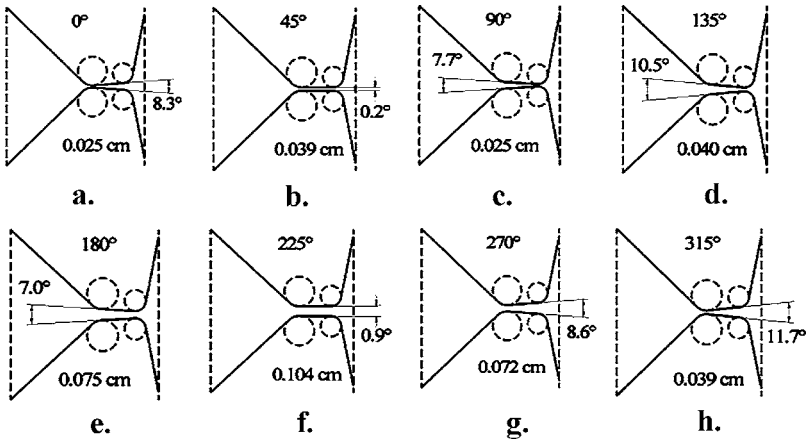


FIG. 4. Glottal shape for certain values of the driving angle, θ.

flow resistance anywhere else in the system. Another particularity of the apparatus is that the air is drawn by a lower pressure created downstream, as this was found more convenient from a design standpoint. The experiments were performed for a set of three lung pressures, i.e., 4, 8, and 12 cm H_2O real-life. Because the pressures in the model are smaller by a factor of 7.5^2 than real-life, the driving pressures used in the model were 7, 14, and 21 Pa. The measurement of such small pressures necessitates the use of extremely accurate pressure sensors (Validyne DP103-10, Validyne Engineering) for the transglottal pressure difference. The frequencies used for the experimental tests were 0.5, 1, 1.5, and 2 Hz, corresponding to real-life frequencies of 28.1, 56.2, 84.4, and 112.5 Hz, respectively. Also, tests were performed for static configurations in the cycle of glottal motion. The volume flow rate was measured via a custom-made pneumotach. Thus, the pressure drop over a metallic net (see Fig. 1) was measured by a Validyne DP103-08 pressure sensor, which is more sensitive than the DP103-10 sensor. The flow resistance of the metallic net was extremely low, about one order of magnitude less than that of the vocal folds. Thus, the pressure drop in the model over the net was oscillating in the vicinity of 0.5, 1, and 1.5 Pa, for driving pressures of 7, 14, and 21 Pa, respectively. As the Validyne pressure transducers have a poor frequency response even at low frequency, special corrections for amplitude and phase were used in order to obtain reliable results. The signal acquired from the pressure sensors was decomposed by means of discrete Fourier transform into harmonic components, whose amplitude and phase were corrected depending on the frequency. The signal was subsequently recomposed in order to obtain the correct measurement pressure. This technique required the construction of an auxiliary experimental device that permitted the comparison of the output signal of the sensor with a reference variable pressure of known amplitude and frequency, such that appropriate correction coefficients could be determined. The auxiliary device consisted of a cam (mounted on the axle of a computer-controlled stepper motor running at prescribed rotational velocities) that drove a piston in order to generate a known, time-variable pressure in a reservoir (considered as a reference pressure). The pressure in the reservoir was measured by using the Validyne sensors, and subsequently compared against the reference pressure, such that the attenuation and lag were calculated at each frequency, and appropriate correction factors were established. This method was applied, using two appropriate cams, for both DP103-10 and DP103-08 pressure sensors [for details about the procedure, see Kucinschi (2004)].

Flow visualization with glycol-based smoke was performed by using a second acquisition system. Two laser sheets were shed from both upstream and downstream of the folds, such that the glottal region could be observed regardless of the glottal angle. A specific glottal position was investigated by synchronizing the pulse of the laser by means of a position signal provided by the apparatus.

III. NUMERICAL SIMULATIONS

In order to obtain the pressure and velocity distributions in the glottis, the Navier-Stokes equations have to be solved numerically. Since the velocities present low values, leading to a small Mach number, $\mathrm{Ma} < 0.1$, the flow can be considered to be that of an incompressible fluid. The large ratio of the width of the model (1.33 cm real-life) to the glottal gap (0.025–0.104 cm real-life) allows the flow through the glottis to be considered two-dimensional. Also, the viscosity can be assumed constant, in which case the form of the Navier-Stokes equations to be solved is

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla p + \mu \nabla^2 \mathbf{u} \tag{1}$$

where $\mathbf{u} = (u_x, u_y)$ and μ is the dynamic viscosity of the air.

The numerical simulations related to the moving vocal folds were performed using the commercial CFD code FLUENT, which is an unstructured, finite-volume-based solver that incorporates a dynamic mesh capability. FLUENT has been successfully used to study the glottal flow (Scherer *et al.*, 2001). Flows can be modeled where the shape of the domain boundaries is time dependent. The flow during the present set of simulations was considered laminar everywhere in the flow domain. The decision to adopt this model was based on the experimental evidence. Flow visualizations showed that the flow inside the glottis (the zone where the velocity gradients present the highest values) is laminar for most of the range of the driving angle, θ, for transglottal pressures of up to 15 cm H_2O real-life. Also, for static cases analyzed in a previous study (Scherer *et al.*, 2001), the results obtained with the laminar model were in excellent agreement with experimental results in terms of flow rate and wall pressures, which is not unexpected since the maximum value of the Reynolds number was around 2500.

A two-dimensional computational domain was used and represented as a hatched area, Fig. 6(a), where the dimensions are provided in cm real-life, although the mesh was generated in terms of the actual dimensions of the experimental apparatus. The boundary conditions were imposed in

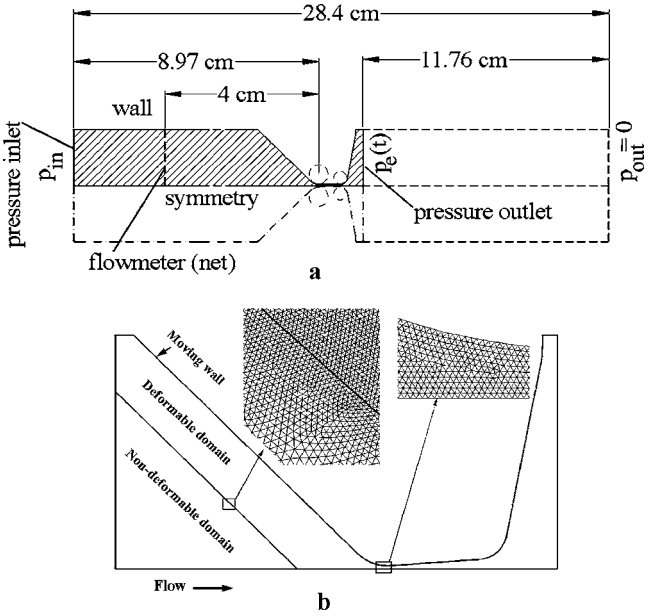


FIG. 6. Computational domain and boundary conditions: (a) computational domain and (b) computational grid.

terms of spatially constant pressures at both inlet and exit. The inlet pressure was taken as constant, while the exit pressure was time dependent, i.e., $p_e = p_e(t)$. In order to further decrease the computational effort, the area downstream of the vocal folds was excluded from the computational domain and the time-dependent exit pressure was estimated as

$$p_e(t) \approx p_{\text{out}} + I(dQ_e/dt), \tag{2}$$

where I is the inertance of the fluid downstream of the vocal folds, $I = \rho l/A$, ρ is the density of the air, and l and A are the length and the cross-sectional area of the downstream section, respectively. This equation is similar to that used by Titze and Story (1997) for the case when the transglottal pressure does not require modeling since it is solved numerically, while the flow resistance of the model vocal tract, which is a constant cross-sectional tunnel, is negligible. The inertive pressures in the model are correctly scaled (i.e., they are 7.5^2 times smaller than the ones in real-life, since the time scale in the model is 7.5^2 times larger than in real-life). The flow rate at the exit of the glottis, Q_e, was related to the inlet flow rate, Q_{in}, and the time rate of change of the glottal volume, V_g:

$$Q_e = Q_{\text{in}} - 2\pi f(dV_g/d\theta). \tag{3}$$

This equation is derived from the equation of continuity for an incompressible fluid, $Q_e = Q_{\text{in}} - (dV_g/dt)$, where $dV_g/dt = (d\theta/dt)(dV_g/d\theta)$ and $d\theta/dt = \omega$, $\omega = 2\pi f$, f being the oscillation frequency. The derivative of the glottal volume, V_g, with respect to the driving angle, θ, was calculated as the glottal volume is known from the kinematics of the mechanism. One restriction for this kind of lumped approach is that the transversal dimension of the model should be smaller than $\lambda/4$, where λ is the wavelength of the acoustic signal (Hall, 1987). As $\lambda = c/f$, $c = 340$ m/s, and f is of the order of 2 Hz (corresponding to a real-life frequency of 112.5 Hz), this condition is fulfilled for the cases approached here. The time derivative dQ_c/dt at the current iteration was calculated based on the values of the flow rate at previous iterations. Since slight numerical oscillations were present that may lead to important errors in estimating the time derivative of the flow rate, it was calculated as the slope of the linear regression of the previous four time steps. This procedure is successful provided the time step is sufficiently small.

The computational grid used for the numerical simulations contained approximately 328,000 triangular cells and, using the symmetry, represents only half of the analyzed geometric configuration. It was thus assumed that the jet was symmetric, although a certain asymmetry of the jet occurs for divergent angles (as revealed by flow visualizations). However, the relatively small divergent angles that occur in the case under scrutiny allow the consideration of the assumption of a symmetric jet for the numerical simulations. A large number of cells were necessary in order to assure a correct numerical representation of the gradient in the glottal region. The mesh is updated automatically by the code at each time step, based on the new shape of the moving boundary (i.e., vocal fold wall). The motion of a boundary is imposed by specifying the position of each node on the boundary through an equation of the type $z = z(x, y, t)$, known from the kinematics analysis of the driving mechanism of the experimental model. A starting mesh needs to be provided to the code, and the domain geometry at subsequent time steps is generated based on the equation of the boundary. An image of the initial computational domain, with details of the grid, is shown in Fig. 6(b). A particularity of the code is that it uses unstructured mesh (triangular cells) for the deformable zones. The mesh density has to be uniform in the deformable zone, even in the convergent entry region. In order to decrease the computational effort, part of the entry region was modeled as a non-deformable domain, where variable mesh density was used, which allowed for a decrease of the number of computational cells.

Ten grids were created and tested on the 8 cm H_2O, 1 Hz case, in order to insure that the results are grid independent. The final selected computational grid led to both good accuracy and a reasonable computational time and was used for the remainder of the cases. A time step of $\Delta t = 10^{-4}$ s (corresponding to a real-life time step of 1.7×10^{-6} s) was used for most of the simulations, leading to a number of time steps per cycle in the range of 5,000–20,000. A first-order implicit time marching algorithm was used, which provides good accuracy for the time step used in the simulations. The accuracy was also assured by imposing small values for the scaled residuals for continuity and momentum (10^{-4}) as convergence criteria per time step. A number of 50 to 100 iterations per time step were necessary for convergence. Details of the numerical algorithms implemented in FLUENT, as well as the methods employed for dynamic mesh handling, are documented elsewhere (FLUENT, 2005). It was observed that the process was cyclical and reached the steady-periodic state after about two complete cycles, regardless of the driven oscillation frequency. Generally, the results from the fourth cycle were retained as reliable results for a typical oscillation cycle.

IV. RESULTS AND DISCUSSION

Flow rates and transglottal pressures were measured over a complete vocal fold motion cycle, both for the static case and for dynamic cases at different frequencies. Two overall pressures were used for the measurements, namely 4 and 8 cm H_2O real-life. The overall pressure, which was constant during the experiments, corresponded to the lung pressure. For the static case, the measurements were performed for static configurations corresponding to fixed driving angles θ. The values of θ cover a complete motion cycle, i.e., from 0° to 360°, with a step of 10°.

The inlet flow rate, Q_{in}, is represented over an oscillation cycle in Fig. 7 for 4 cm H_2O and Fig. 8 for 8 cm H_2O, for the four oscillation frequencies (28.1, 56.2, 84.4, and 112.5 Hz). The experimental values, both for static and dynamic cases, were plotted together with the values obtained from the numerical simulations. For the dynamic cases, the driving angle was related to the oscillation frequency, f, and the current time of measurement/simulation, t, through the equation

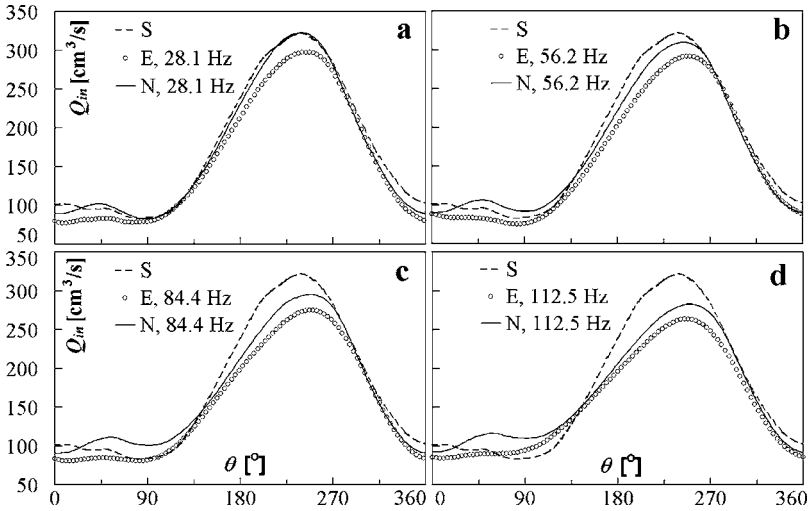


FIG. 7. Inlet flow rate over a cycle for a nominal (lung) pressure of 4 cm H_2O: (S)=static experimental, (E)=experimental, and (N)=numeric.

$$\theta = 360\,^\circ tf, \quad t \in [0, \ldots, T], \quad T = 1/f. \tag{4}$$

The experimental values are in general lower than the numerical predictions for the driving angles 0°–120° and 135°–270°, where for the latter range the flow rate strongly increases and then decreases. The difference between the numerical results and the experimental data was caused by the simplified flow model and geometry used in the numerical simulations (see Sec. III).

As the glottal diameter and the glottal angle are coupled through the geometry of the mechanism, their variation depends on the configuration of the apparatus. Thus, for θ from 0° to 90°, the glottal diameter, D_g, is roughly constant (around 0.025 cm real-life, which is the minimum value along the cycle, with a local maximum of 0.039 cm at 45°), while the glottal angle is changing continuously from 8.3° divergent to 7.7° convergent (see Fig. 5). As a general observation, the input flow rate for θ=0°–90° presents the smallest values along the cycle, regardless of the frequency of the folds and of the overall pressure. For the static case, the minimum flow rate was obtained at one-fourth of the cycle (θ=90°), when the glottal angle was convergent. Slightly higher values were measured for θ=0°, when the glottal angle had a divergent angle of roughly the same magnitude and the glottal diameter had the same value. This difference in flow was due to the larger friction drag that occurs for convergent glottal configurations and small diameters, when viscous effects are important compared to the divergent configuration of small angle. One can conclude that for the static case, when the lung pressure was constant, the smallest value of the flow rate occurred for a convergent angle and smallest glottal diameter. The dynamic results show the flow rate slightly smaller in magnitude than the static results, although the decrease in flow rate with increasing angle is still evident except for the 4 cm H_2O, 112.5 Hz case in Fig. 7(d) and the oscillatory behavior in Fig. 8(a) where the flow rate increases.

For driving angles, θ, in the 90°–180° range, the folds remain convergent, changing the converging angle from 7.7° to 10.5° to 7°, while the glottal diameter increases from 0.025 to 0.040 to 0.075 cm. The tendency of the inlet flow rate was to increase with the glottal diameter, valid for both static and dynamic cases. For driving angles, θ, in the 180°–270° range, the glottal angle changes from a convergent (7°) to a divergent configuration (8.6°), while the glottal diameter changed slightly, in the range from 0.075 to 0.104 to 0.072 cm real-life. The configurations for both θ=180° and θ=270° had nearly the same glottal diameter, but for θ=270°, where the glottal angle was divergent, the flow rate

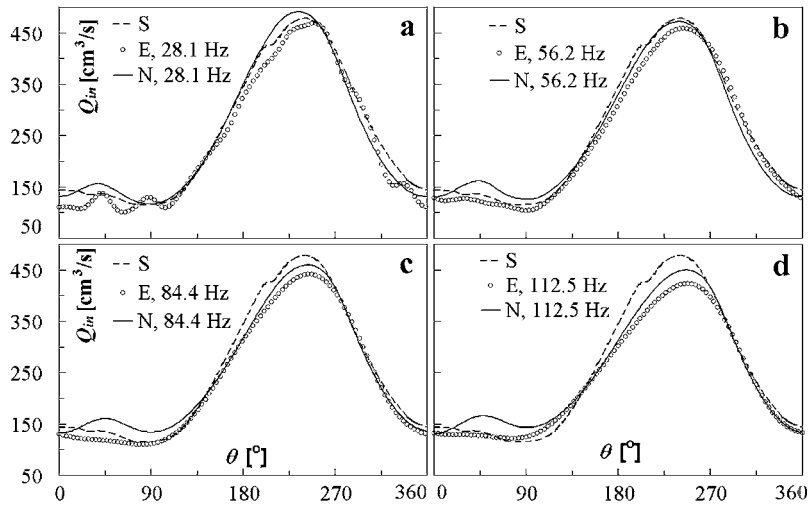


FIG. 8. Inlet flow rate over a cycle for a nominal (lung) pressure of 8 cm H_2O: (S)=static experimental, (E)=experimental, and (N)=numeric.

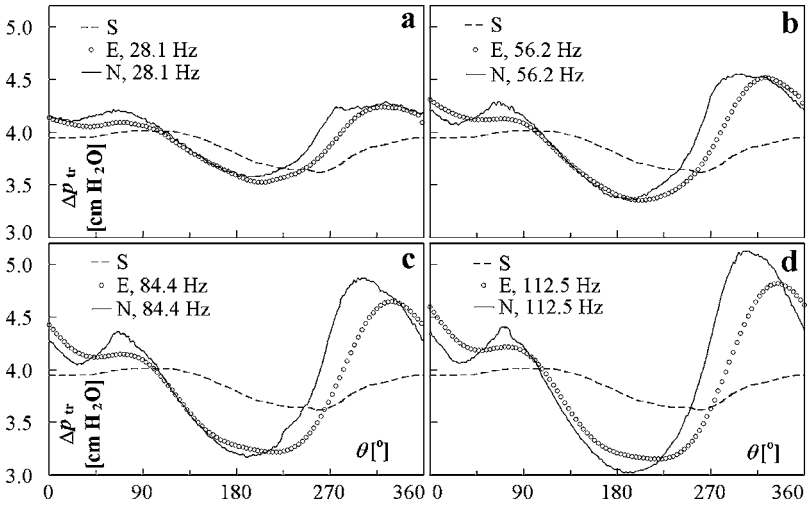


FIG. 9. Transglottal pressure for a nominal (lung) pressure of 4 cm H_2O: (S)=static experimental, (E)=experimental, and (N)=numeric.

was larger. This observation is valid for both the static and dynamic cases. Thus, a divergent glottal angle produces a larger flow rate for the same diameter, consistent with the same trend observed for θ in the 0°–90° range for the static case. The largest static flow rates occurred at θ=240° for both 4 and 8 cm H_2O, having values of 322 and 479 cm^3/s, respectively. For the last quarter of the cycle (θ =270°–360°) the glottis remained divergent (the glottal angle changing slightly in the range from 8.6° to 11.7° to 8.3°), while the glottal diameter returned to the minimal value of 0.025 cm real-life. As expected, the flow rate decreased in this portion of the cycle for both static and dynamic cases. In general, the static flow rates over the range from 135° to 270° were higher than the dynamic values.

For the dynamic cases, a decrease of the maximum flow rate was observed with an increase of the oscillation frequency of the vocal folds as shown in Figs. 7 and 8. This tendency was revealed by both the measurement data and the numerical results. The change in flow rate with frequency was due to the inertive effects of the fluid in the region upstream (corresponding to the trachea), downstream of the vocal folds (corresponding to the vocal tract), and within the glottis. However, as the mass of the air in the glottis is much smaller than that in the rest of the apparatus (less than 0.2%), one can infer that its effect was negligible. The inertance of the air upstream and downstream of the vocal folds created a time-changing pressure at the glottal inlet and exit, respectively, which opposes the variation of the flow rate. The magnitude of this effect increases with the frequency, since $dQ/dt=2\pi f(dQ/d\theta)$. One can remark here that the inertive effects could not have been revealed if the fluid were driven by a constant flow rate or if the regions of the apparatus corresponding to the trachea and vocal tract were removed. In fact, the inertive effects are probably larger in reality, since the average male vocal tract length is 17.5 cm (Titze, 1994), while the vocal tract length in the current apparatus was about 12 cm. The general shape of the variation of the reported flow rate over a cycle is consistent with the real glottal flow pulse shape (Rothenberg, 1981). The effect of inertia of the air column is to skew the waveform to the right (Titze, 1994), i.e., the glottal flow presents a slow rise followed by an abrupt fall. Obviously, if the vocal folds collide and the flow rate were temporarily zero, the time derivative of the flow rate would subsequently have large values, and hence large inertive effects would be expected. For the cases reported here, the glottal flow never reached zero, such that a strong waveform skewing was not obtained.

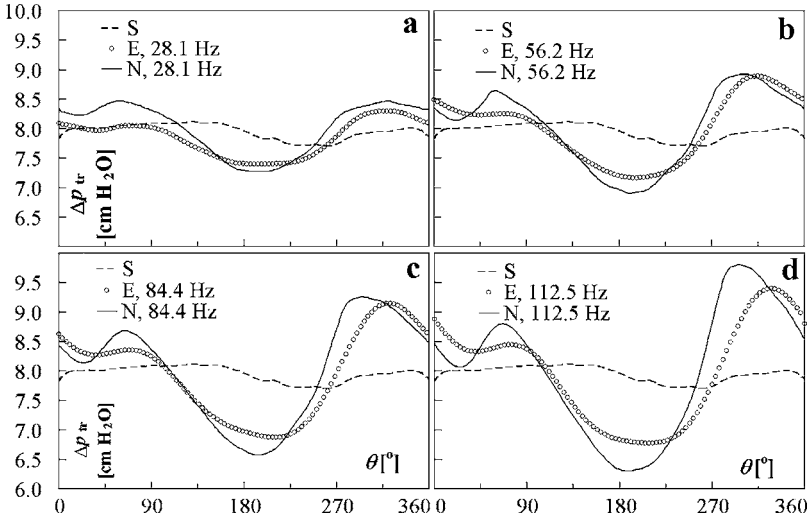


FIG. 10. Transglottal pressure for a nominal (lung) pressure of 8 cm H_2O: (S)=static experimental, (E)=experimental, and (N)=numeric.

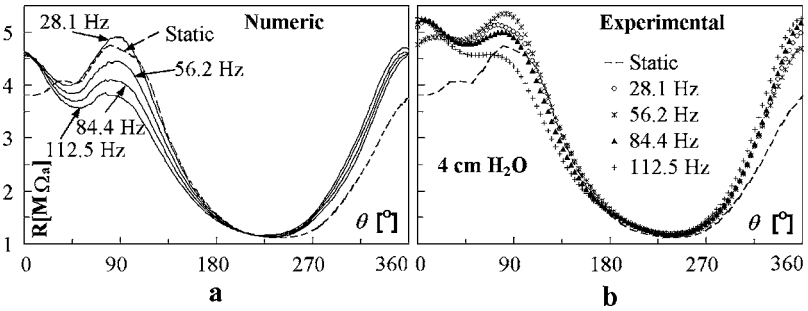


FIG. 11. Glottal flow resistance for a nominal (lung) pressure of 4 cm H_2O.

The trends of the transglottal pressure difference, Δp_{tr}, are represented in Fig. 9 for 4 cm H_2O and Fig. 10 for 8 cm H_2O, for the same frequencies (i.e., 28.1, 56.2, 84.4, and 112.5 Hz). For the static case, a larger Δp_{tr} occurs at $\theta=90°$ (convergent glottis) than at $\theta=0°$ (divergent glottis), for the same (minimum) glottal diameter, because of the higher resistive viscous effects. The minimum static transglottal pressure occurs near $\theta=255°$, which is just after the location corresponding to the maximum glottal diameter ($\theta=225°$). The static pressures show a smaller variation with phase angle than do the dynamic experimental and numerical values.

The difference between the measured and the calculated dynamic transglottal pressures is caused by the simplifying assumptions used in the numerical simulations. Small numerical oscillations are observed for the numerical pressure curves, caused by the estimation of exit boundary pressures based on the flow rate derivatives [see Eq. (2)]. The numerical model predictions show similar variations with driving angle as the experimental values, although there is a magnitude difference which increases with frequency at the higher driving angles. For the numerical 4 cm H_2O case, there is an overprediction for $30° < \theta < 90°$ and $225° < \theta < 315°$ for the lowest frequency. Higher frequencies show the numerical predictions leading the experimental values for $\theta > 225°$. For 8 cm H_2O the overprediction also exists for $\theta > 225°$, as well as an underprediction for $90° < \theta < 225°$ for the higher frequencies.

The magnitude (maximum to minimum) of the experimental Δp_{tr} is strongly dependent on the oscillation frequency, due to the inertive effects of the air columns upstream and downstream of the vocal folds. This difference increases with the frequency, from 0.8 cm H_2O at 28.1 Hz for 4 cm H_2O to 1.5 cm H_2O at 112.5 Hz, and for 8 cm H_2O, 0.9 cm H_2O at 28.1 Hz to 2.6 cm H_2O at 112.5 Hz. The most typical phonatory condition is the 112.5-Hz case, for which the percentage extent relative to the average subglottal pressure was 37.5% for 4 cm H_2O and 32.5% for 8 cm H_2O, consistent with data obtained in humans with regard primarily to subglottal pressure fluctuations because of the foreshortened subglottal tract (Kitzing and Löfqvist, 1975). It is noted that the transglottal pressure waveform is slightly displaced to the right with the increase in oscillation frequency.

The instantaneous flow resistance, R, defined as the ratio of the transglottal pressure difference, Δp_{tr}, to the inlet flow rate Q_{in}, is represented in Fig. 11 for 4 cm H_2O and Fig. 12 for 8 cm H_2O, for both the static and dynamic cases. R is expressed in $M\Omega_a$ ($\Omega_a = \text{kg m}^{-4}\,\text{s}^{-1}$). Both the numerical [Figs. 11(a) and 12(a)] and experimental [Figs. 11(b) and 12(b)] are shown and there is reasonable agreement between them. Although both the flow rate and the transglottal pressure depend on the whole system composed of fluid in the trachea, glottis, and vocal tract, their ratio defines the aerodynamic behavior of the vocal folds. Using the experimental values, Figs. 11(b) and 12(b) illustrate that the glottal flow resistance changes with glottal angle (glottal shape) and also shows a dependence on the vocal fold oscillation frequency over certain parts of the cycle of driving angles. For a driving angle range of $225° < \theta < 270°$, which corresponds to a uniform glottis at 225° becoming 8.6° divergent at 270° (Fig. 4), the resistance is at its lowest value at 240° and then increases at 270°, but does not show any dependence on the oscillation frequency. As the glottal divergence angle increases to 11.7° at 315°, and then decreases to 8.3° at 360° or 0° in the cycle, the flow resistance continues to increase and shows a decreasing magnitude with increasing vocal fold frequency. The resistance continues to increase and oscillate from the 360° or 0° divergent position through the glottal angles of 45° (glottis slightly divergent) and 90° (glottis convergent). The resistance reaches its maximum value at 90° at the lowest frequency and shows a rather large decrease in magnitude with increasing frequency. Even though the pressure at that position (90°) is higher for higher frequencies,

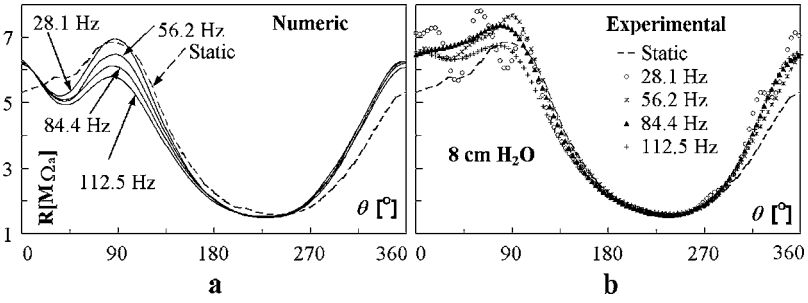


FIG. 12. Glottal flow resistance for a nominal (lung) pressure of 8 cm H_2O.

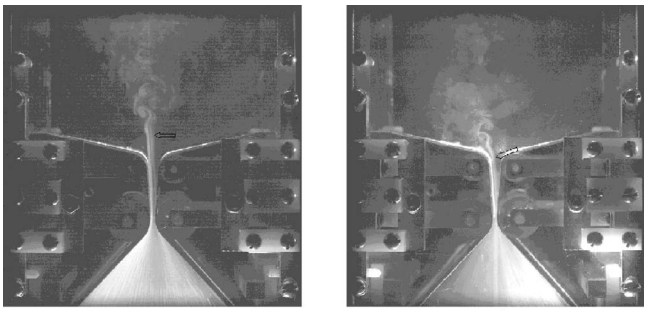


FIG. 13. Glottal flow visualization, 8 cm H_2O, position 0°, static versus dynamic (56.2 Hz). The arrow indicates the transition point.

the flow rate increases at a greater rate, leading to a decrease in the flow resistance. These changes should not be interpreted as a consequence only of the inertia of the air within the glottis, but also as the effect of the increase of the time rate of change of the volume of the glottis [see Eq. (3)]. At this position the glottal diameter is at its minimum value (0.025 cm) for a convergent angle, and the viscous effects are the most important over the cycle. Continuing the cycle to 135°, 180°, and back to 225°, the glottis is convergent, the overall resistance falls, and the dependence on frequency decreases back to being negligible for this set of convergent angles. The minimum resistance, which occurs at $\theta=240°$ (where the glottal diameter is larger and thus the viscous effects are less important) is 1.2 $M\Omega_a$ for 4 cm H_2O and 1.6 $M\Omega_a$ for 8 cm H_2O.

The trends of the flow resistance discussed above show that the aerodynamic behavior of the vocal folds, for any glottal shape, is mainly governed by viscous effects and depends to a lesser extent on the oscillation frequency except for $45° < \theta < 135°$. The flow patterns inside and downstream of the vocal folds, however, are dependent upon the oscillation frequency of the vocal folds, especially when the glottal angle is divergent and the minimum throat thickness is decreasing ($\theta=225°-0°$). For these configurations, flow visualization showed that the Kelvin-Helmholtz shear layer instabilities occur closer to the glottal exit once the frequencies were increased. A static flow visualization picture is shown in Fig. 13(a), for $\theta=0°$ (i.e., glottal angle 8.3° divergent) and a lung pressure of 8 cm H_2O, while Fig. 13(b) presents the visualization obtained at 56.2 Hz at the same moment in the oscillation cycle. For the static case, a laminar jet occurred that separated about one-third of the way upstream in the glottis from the glottal exit on the nonflow (right) wall. The jet was not symmetric, but adhered to one of the walls, and was bi-stable (i.e., the jet could switch sides, triggered by a disturbance in the flow). The jet remained laminar until transition (Kelvin-Helmholtz instability) vortices occurred, followed by turbulent dissipation. For the dynamic case [Fig. 13(b)], a bi-stable, laminar, glottal jet can be observed, but the transition occurred inside the glottis. The same behavior of the flow could be observed for a large number of oscillation frequencies over this range of glottal angles, and, interestingly, the transition points tended to move more upstream with an increase in frequency. The effect of this change in flow patterns can be seen at 0° as a decrease in the flow rate (Figs. 7 and 8) and an increase in the transglottal pressure at 0° (Figs. 9 and 10). The vortex shedding close to solid boundaries can provide a secondary source of acoustic energy. Barney *et al.* (1999) showed that, for a shutter-type dynamic mechanical model, the vortices in the shutter area provide significant acoustic sources. The relationship between the flow patterns and the acoustic field was not within the scope of the present paper, but will be examined in a separate report.

V. CONCLUSIONS

The purpose of the present study was to investigate both experimentally and numerically the relationship between the flow rate and transglottal pressure difference for the moving vocal folds. A complex experimental apparatus was designed for this purpose, which mimics with acceptable realism the oscillation of the human vocal folds.

It was also sought to determine, at least in the framework of the present experimental conditions (i.e., glottal configurations and lung pressures), to what extent the flow rates and transglottal pressure are dependent upon the oscillation frequencies of the vocal folds. Flow visualizations were performed (although not fully reported here) in order to determine the character of the glottal flow (i.e., laminar, transitional, or turbulent). The visualizations showed the existence of a laminar jet in the glottis, except near the exit of the glottis for certain diverging glottal angles and frequencies. The jet was always laminar in the throat region where the viscous effects are important.

For the static configurations, it was observed that the flow rates were lower for the convergent than for divergent glottal angles for roughly the same small glottal diameter, due to the viscous drag. As expected, the flow rate also increased with glottal diameter when the glottal angle was kept roughly constant.

The measurements showed that both the flow rate and transglottal pressure are affected by the frequency of the vocal folds. The inertia of air in the trachea and vocal tract, together with the unsteady flow rate, produced a time-changing pressure both downstream and upstream of the glottis, which was reflected in frequency-dependent flow rates and transglottal pressure differences. Although the collision of the modeled vocal folds was not permitted, the flow waveform was consistent with the shape of the glottal pulse reported in the literature (e.g., Rothenberg, 1981). The increase in oscillation frequency was reflected by a decrease in flow amplitude. Skewing (i.e., shift to the right) of the waveform relative to the static flow waveform existed for all frequencies.

The glottal flow resistance was calculated as the ratio of the transglottal pressure difference to the flow rate. A sizeable decrease of the flow resistance with increase in frequency was observed at about one-quarter of the oscillation cycle, where the glottal diameter was small and a large change in glottal volume takes place. For the remainder of the oscillation cycle, the flow resistance was practically independent of the frequency for $135° < \theta < 270°$ and showed a smaller decrease in resistance with increasing frequency for

$270° < \theta < 360°$. Also, the flow resistance changed very little with the lung pressure in the part of the cycle where the glottal diameters were large. However, it changes significantly with lung pressure for small glottal diameters, where viscous effects are important.

A satisfactory agreement of the results of the numerical simulations with the experimental data was obtained. The differences were caused by the limitations of the flow model used in the simulations, i.e., laminar flow model, glottal jet assumed symmetric, and simplified model of inertia for the vocal tract. However, the conclusion is that the CFD code FLUENT can provide reasonable predictions of the flow parameters.

Flow visualization was not the scope of the present paper, and a detailed analysis of the glottal flow patterns is not presented here. Nevertheless, the flow pictures showed that the flow patterns were altered between the static and dynamic regimes for divergent angle cases ($\theta = 225° - 0°$) where shear-layer instabilities and consequently vortex structures can occur inside the glottis for the dynamic case. Since vortex shedding can become a sizeable source of sound, it can be inferred that the oscillation frequency of the vocal folds may also be related to the generation of aero-acoustic sound. An additional study is necessary to investigate this aspect.

ACKNOWLEDGMENT

This research was support by NIH Grant No. R01 DC03577.

Alipour, F., and Scherer, R. C. (**2000**). "Vocal bulging effects on phonation using a biophysical computer model," J. Voice **14**(4), 47–483.

Alipour, F., and Scherer, R. C. (**2002**). "Pressure and Velocity profiles in a mechanical hemilarynx model," J. Acoust. Soc. Am. **112**, 2996–3003.

Alipour, F., Berry, D., and Titze, I. R. (**2000**). "A finite-element model of vocal fold vibration," J. Acoust. Soc. Am. **108**, 3003–3012.

Alipour, F., Fan, C. W., and Scherer, R. C. (**1996b**). "A numerical simulation of laryngeal flow in a forced-oscillation glottal model," Comput. Speech Lang. **10**(2), 75–93.

Alipour, F., Scherer, R. C., and Knowles, J. (**1996a**). "Velocity distributions in glottal models," J. Voice **10**(1), 50–58.

Barney, A., Shadle, C. H., and Davies, P. O. (**1999**). "Fluid flow in a dynamic mechanical model of the vocal folds and tract. I. Measurements and theory," J. Acoust. Soc. Am. **105**, 444–455.

Batchelor, G. K. (**1967**). *Fluid Dynamics* (Cambridge U. P., New York), pp. 212–215.

Berry, D. A., and Titze, I. R. (**1996**). "Normal modes in a continuum model of vocal fold tissues," J. Acoust. Soc. Am. **100**, 3345–3354.

Berry, D. A., Herzel, H., Titze, I. R., and Krischer, K. (**1994**). "Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions," J. Acoust. Soc. Am. **95**, 3595–3604.

Deverge, M., Pelorson, X., Vilain, C., Lagrée, P. Y., Chentouf, F., Willems, J., and Hirschberg, A. (**2003**). "Influence of collision on the flow through in-vitro rigid models of the vocal folds," J. Acoust. Soc. Am. **114**, 3354–3362.

FLUENT (**2005**). *FLUENT* 6.2 User Guide, FLUENT Inc.

Gauffin, J., Binh, N., Ananthapadmanabha, T., and Fant, G. (**1983**). "Glottal Geometry and Volume Velocity Waveform," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, edited by D. Bless and J. Bless (College-Hill, Boston), pp. 194–201.

Guo, C. G., and Scherer, R. C. (**1993**). "Finite element simulation of glottal flow and pressure," J. Acoust. Soc. Am. **94**, 688–700.

Hall, D. E. (**1987**). *Basic Acoustics* (Harper & Row, New York).

Hirano, M. (**1977**). "Structure and vibratory behavior of the vocal folds," in *Dynamic Aspects of Speech Production*, edited by M. Sawashima and F. Cooper (Univ. of Tokyo, Tokyo), pp. 13–27.

Hofmans, G. C. J., and Groot, G. (**2003**). "Unsteady flow through in-vitro models of the glottis," J. Acoust. Soc. Am. **113**, 1658–1675.

Ishizaka, K., and Flanagan, J. L. (**1972**). "Synthesis of Voice Sounds from a Two-Mass Model of the Vocal Cords," Bell Syst. Tech. J. **51**(6), 1233–1268.

Kitzing, P., and Löfqvist, A. (**1975**). "Subglottal and oral air pressures during phonation—preliminary investigation using a miniature transducer system," Med. Biol. Eng. **13**(5), 644–648.

Kucinschi, B. R. (**2004**). "An analysis of the flow through a driven mechanical model of the vocal folds," Ph.D. dissertation, University of Toledo, Toledo, OH.

Pelorson, X., Hirschberg, A., Wijnands, A. P., and Baillet, H. (**1995**). "Description of the flow through in-vitro models of the glottis during phonation," Acta Acust. (Beijing) **3**, 191–202.

Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y. (**1994**). "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model," J. Acoust. Soc. Am. **96**, 3416–3431.

Rothenberg, M. (**1981**). "Acoustic interaction between the glottal source and vocal tract," in *Proceedings of the Kurume Vocal Fold Physiology Conference*, edited by K. N. Stevens and N. Hirano (Univ. of Tokyo, Tokyo), pp. 305–328.

Scherer, R. C., Titze, I. R., and Curtis, J. F. (**1983**). "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," J. Acoust. Soc. Am. **73**, 668–676.

Scherer, R. C., Shinwari, D., DeWitt, K., Zhang, C., Kucinschi, B., and Afjeh, A. (**2001**). "Intraglottal pressure profiles for a symmetric and oblique glottis with a divergence angle of 10 degrees," J. Acoust. Soc. Am. **109**, 1616–1630.

Story, B., and Titze, I. R. (**1995**). "Voice simulation with a body-cover model of the vocal folds," J. Acoust. Soc. Am. **97**, 1249–1260.

Thomson, S. L., Mongeau, L., and Frankel, S. H. (**2005**). "Aerodynamic transfer of energy to the vocal folds," J. Acoust. Soc. Am. **118**, 1689–1700.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice Hall, Englewood Cliffs, NJ).

Titze, I. R., and Story, B. H. (**1997**). "Acoustic interactions of the voice source with the lower vocal tract," J. Acoust. Soc. Am. **101**, 2234–2243.

Titze, I. R., and Strong, W. J. (**1975**). "Normal modes in vocal cord tissues," J. Acoust. Soc. Am. **57**, 736–744.

Titze, I. R., Jiang, J. J., and Hsiao, T. Y. (**1993**). "Measurement of mucosal wave propagation and vertical phase difference in vocal fold vibration," Ann. Otol. Rhinol. Laryngol. **102**, 58–63.

van den Berg, J. W., Zantema, J. T., and Doornenball, P. (**1957**). "On the air resistance and the Bernoulli effect of the human larynx," J. Acoust. Soc. Am. **29**, 626–631.

Zhang, Z., Mongeau, L., and Frankel, S. H. (**2002**). "Experimental verification of the quasi-steady approximation for aerodynamic sound generation by pulsating jets in tubes," J. Acoust. Soc. Am. **112**, 1652–1663.

Enhanced contrast for vowels in utterance focus: A cross-language study

Jessica F. Hay, Momoko Sato, Amy E. Coren, Cheryl L. Moran, and Randy L. Diehl
Department of Psychology and Center for Perceptual Systems, University of Texas at Austin, 1 University Station A8000, Austin, Texas 78712-0187

(Received 3 May 2005; revised 10 February 2006; accepted 10 February 2006)

The present study examined several potential distinctiveness-enhancing correlates of vowels produced in utterance focus by talkers of American English, French, and German. These correlates included possible increases in vowel space size, in formant movement within individual vowels, and in duration variance among vowels. Each language group enhanced the distinctiveness of vowels in [+focus] context but used somewhat differing means to achieve this. All three groups used spectral differences, but only German talkers used durational differences, to enhance distinctiveness. The results suggest that the amount of distinctiveness enhancement of a vowel property in [+focus] context is positively related to the between-category variation of that property in [−focus] context. Thus, consistent with the theory of adaptive dispersion, utterance clarity appears to vary directly with information content. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184226]

PACS number(s): 43.70.Bk, 43.70.Fq, 43.70.Kv [ARB] Pages: 3022–3033

I. INTRODUCTION

In this study, we examined several possible distinctiveness-enhancing correlates of focally prominent syllables in American English, French, and German. The aim was to assess cross-language commonalities and differences in how phonological distinctiveness of vowels is increased in [+focus] contexts.

Jakobson *et al.* (1963) famously remarked that "we speak to be heard in order to be understood" (p. 13). In his theory of adaptive dispersion, Lindblom (1986, 1990) elaborated this point as follows: the aim of talkers is to produce utterances that are intelligible to listeners, but to do so with as little effort as necessary. In other words, talkers try to achieve sufficient, rather than maximal, distinctiveness in articulating vowels and consonants, and they thus tend to vary their utterances from reduced (hypospeech) forms to clear (hyperspeech) forms according to the communication conditions that apply.

A variety of findings are consistent with Lindblom's account. Communication conditions that have been shown to influence how clearly talkers articulate speech sounds include factors pertaining to the communication channel, the listener, and the information content of the message.

Communication channel and listener factors: An early study by Lombard (1911) demonstrated that talkers produce clearer and louder speech in the presence of background noise, a result that has since been replicated and extended (Lane *et al.*, 1970; Lane and Tranel, 1971). As the distance to the listener increases, talkers increase their vocal effort, although not enough to compensate fully for the distance effect on signal intensity (Traunmüller and Eriksson, 2000). In conversational speech, talkers tend to speak more clearly when the line of sight between talker and listener is obstructed, and the listener is thus deprived of visual speech reading cues, than when the talker and listener are able to see each other (Anderson *et al.*, 1997). Picheny *et al.* (1985, 1986) instructed talkers to produce either ordinary conversational speech or clear speech as if "trying to communicate in a noisy environment or to an impaired listener" (Picheny *et al.*, 1985, p. 97). The clear speech was produced at a slower rate, a higher intensity, and with less vowel and consonant reduction than the conversational speech. As expected, the clear speech was more intelligible to hearing impaired listeners than the conversational speech.[1] Talkers tend to produce more widely dispersed vowel categories, yielding less acoustic overlap among categories, when speaking to non-native speaking listeners than when speaking to native speaking listeners (Knoll and Uther, 2004). Such results suggest that talkers modulate their speech clarity according to demands imposed by the communication channel and by the listener.

Information content of the message: Perhaps the strongest evidence that talkers aim for sufficient—rather than maximal—distinctiveness is pervasive findings that utterance clarity tends to vary positively with information content: (a) Words produced in isolation are more accurately identified than the same words excised from running speech (Pollack and Pickett, 1963). The excised word tokens are more reduced, presumably because of their greater predictability (i.e., lower information content) in the original contexts. (b) Words excised from stereotyped phrases or adages (e.g., "A stitch in time saves [*nine*]") are less intelligible than the same words excised from less redundant contexts (e.g., "The number that you will hear is [*nine*]") (Lieberman, 1963; see Hunnicutt, 1985, for a partial replication and qualification of this result). (c) Words produced for the first time in the course of a monolog are more intelligible when excised from their contexts than words produced for the second time (Fowler and Housum, 1987). (d) Vowels are more clearly articulated in words that occupy dense lexical neighborhoods and that are low in frequency of occurrence relative to their nearest phonological neighbors (Wright, 2003). Higher neighborhood density and lower frequency of occurrence yield higher information content, and both are associated

 0001-4966/2006/119(5)/3022/12/$22.50

with lower word recognition accuracy (Luce, 1986). (e) Vowels are more reduced in closed class function words (e.g., the auxiliary verb "can") than in open class content words (e.g., the noun "candy") (van Bergem, 1993). (f) Vowels in lexically unstressed syllables are more reduced (i.e., shorter in duration and less distant from schwa) than vowels in lexically stressed syllables (van Bergem, 1993). Cross-linguistically, vowels in stressed syllables tend to carry more information because of the greater variety of vowels permitted in stressed contexts (Altman and Carter, 1989). (g) Consonants at the beginning of prosodic domains such as words or phrases are more clearly articulated and less confusable than later-occurring consonants, which are generally more predictable (Jun, 1993; Byrd, 1994; Browman and Goldstein, 1995; Redford and Diehl, 1999; Keating *et al.*, 1999; Cho and Keating, 2001). (h) Vowels in [+focus] context (i.e., in portions of an utterance that contain *new*, as opposed to *given*, information) are produced with greater separation between vowel categories than those in [−focus] context (van Bergem, 1993). (i) In languages with vowel length distinctions, duration differences between short and long vowels tend to be enhanced in [+focus] context. Such enhancement has been observed in Swedish (Heldner and Strangert, 2001), Arabic (de Jong and Zawaydeh, 2002), and Serbian (Smiljanic, 2004). (j) Phonetic correlates of [voice] contrasts show enhanced acoustic distinctiveness in [+focus] context (Cho and McQueen, 2005).

In examples (a)–(j), information content was varied and/or assessed qualitatively, that is, in the form of ordinal comparisons. Taking a more quantitative approach, van Son and Pols (2003) used a large sample of spoken Dutch to estimate the information (in bits) carried by individual phonemes in word and sentence context. Across several speaking styles, they found positive correlations between information content and two measures of speech clarity—duration and spectral distinctiveness. Although these correlations were highly significant statistically, the *R* values were fairly small (<0.3), owing perhaps to the many other sources of durational and spectral variance in natural productions.

As this last study suggests, talker compensation for varying communication conditions is far from perfect. In some cases (e.g., varying background noise, varying distance separating talker and listener, and varying degrees of hearing loss), there may be physical or perceptual limits to maintaining uniform levels of intelligibility. Moreover, in cases of varying information content, talkers may lack the cognitive means to estimate accurately the information available to listeners on a segment-by-segment basis, and thus articulatory compensation may be rather crude. For example, Billerey-Mosier (2000) reported no systematic difference in vowel duration as a function of whether the vowel occurred at the carrier word's uniqueness point (i.e., the earliest point at which the word differs from all lexical neighbors sharing the same initial sequence of phonemes) or after the uniqueness point, where the vowel carries no information. Also, in the study by Anderson *et al.* (1997) cited earlier, the reduction in talker clarity that occurred with an unobstructed line of sight between talker and listener was not closely linked to the moment-to-moment looking behavior of the listener. Finally, talkers may sometimes have difficulty distinguishing what is given information for themselves and what is given information for the listener. For example, Bard *et al.* (2000) reported that when talkers named map landmarks in spontaneous speech to two successive listeners, the key words were produced less clearly on repetition even though the second listener had not heard the original version. Despite such findings, however, it is evident that Lindblom's notion of sufficient contrast accounts at least qualitatively for an impressive range of data.

The present study examines how production of vowels is affected by whether or not they occur in utterance focus, a variable associated with differences in information content. Declarative sentences can typically be analyzed into two kinds of grammatical constituents: those carrying information already known to the listener (given information) and those carrying information not previously known to the listener (new information). The feature [+focus] applies to the latter kind of constituent, whereas [−focus] applies to the former.[2] In general, [+focus] corresponds to the constituent occupying the location of a wh-phrase in the answer to a wh-question:

Q: Who stole the money?

A: $[\text{Bob}]_{\text{F}}$stole the money. (1)

The grammatical constituent marked as [+focus] usually includes the syllable that is intonationally most prominent in the utterance. Linguists have offered varying accounts of how focal prominence is assigned within the focused constituent (Chomsky and Halle, 1968; Ladd, 1980, 1996; Gussenhoven, 1983; Selkirk, 1984; Baart, 1987; Bolinger 1989). Typically, it is assumed to be assigned to the primary stressed syllable of the head (i.e., most important lexical item) of the focused constituent.

Phonetic investigations of various languages have described several correlates of focal prominence. Relative to stressed syllables in [−focus] context, stressed syllables in [+focus] context tend to be longer (Nooteboom, 1972; Cooper *et al.*, 1985; Eefting, 1991; Sluijter and van Heuven, 1995; Turk and Sawusch, 1997; Strangert and Heldner, 1998; Heldner and Strangert, 2001), to be more intense (Campbell, 1995; Sluijter and van Heuven, 1996a; Heldner *et al.*, 1999; Heldner, 2003), to have more high-frequency emphasis (Campbell, 1995; Sluijter and van Heuven, 1996b; Heldner *et al.*, 1999; Heldner, 2003), and to have a greater and more rapid change in fundamental frequency (O'Shaughnessy, 1979; Lyberg, 1979; 't Hart *et al.*, 1990). All of these correlates of focus are likely to contribute to intonational prominence and thus to draw the listener's attention to the most informative portion of an utterance (Ladd and Cutler, 1983).

Consistent with Lindblom's notion of sufficient contrast, other correlates of focal prominence may have a distinctiveness-enhancing role, in addition to any attention-directing function they serve. One such correlate is size of the vowel space: [+focus] syllables tend to show a greater separation of vowel categories within a formant frequency ($F1 \times F2$) space (van Bergem, 1993). It is also possible that [+focus] syllables are associated with greater formant move-

ment within each vowel and/or greater duration variance across vowel categories. Several investigators (e.g., Nearey and Assmann, 1986; Hillenbrand *et al.*, 1995; Hillenbrand and Nearey, 1999) have reported that American English vowels tend to exhibit "vowel-inherent spectral change" (VISC), whereby formant frequencies change from a relatively steady-state region near the onset of the vowel to a different set of values near the offset of the vowel. As the label implies, the frequency changes are assumed to be characteristic of the vowel itself and not merely a by-product of coarticulation with initial and final consonants. Including VISC in the design of synthetic American English vowel stimuli significantly improves their identifiability by listeners (Hillenbrand and Nearey, 1999), and including it in the parametric description of natural American English vowels significantly improves categorization performance by statistical pattern recognition models (Nearey and Assmann, 1986). Similar pattern recognition results have been reported for monophthongs of Australian English (Watson and Harrington, 1999). Because VISC contributes to vowel identity in at least one language, it may be subject to enhancement in [+focus] context, yielding greater acoustic distinctiveness among vowels. Apart from spectral differences, vowel categories tend to vary in duration, and in many languages, vowel length is phonemic (Lehiste, 1970). It was noted earlier that such vowel length distinctions have been shown to be enhanced in [+focus] context for several languages (Heldner and Strangert, 2001; de Jong and Zawaydeh, 2002; Smiljanic, 2004).

In the present study, we examined vowel space size, amount of formant movement, and duration variance as possible distinctiveness-enhancing correlates of [+focus] syllables in American English, French, and German. The aim was to assess cross-language commonalities and differences in how phonological distinctiveness of vowels is increased in [+focus] context. The requirement of sufficient (rather than maximal) contrast allows for the possibility that different languages may exploit varying means of distinctiveness enhancement. Such variation may, of course, reflect differences across languages in the phonological contrasts used, but it may also reflect different weights assigned to the various correlates of a given contrast. For example, a feature correlate that carries more information in language A than in language B may be more heavily exploited for distinctiveness enhancement in language A.

II. METHOD

A. Talkers

American English talkers were 12 undergraduates (six male and six female) from the University of Texas at Austin, all of whom spoke with a Central Texas dialect. They ranged in age from 18 to 22 years old. French talkers were 12 undergraduates (six male and six female) from the Université René Descartes, Paris, France. All spoke with a standard Parisian dialect, and they ranged in age from 18 to 29 years old. German talkers were 14 adults (seven male and seven female) who were currently living in or visiting the central Texas area and who ranged in age from 15 to 38. They were from various dialect regions in Germany, with five speaking the Franconian dialect, three speaking other regional dialects, and six speaking only standard German. However, all of the talkers reported being proficient in standard German. American English talkers were recruited from the introductory psychology subject pool at the University of Texas at Austin, and were given credit towards partial fulfillment of course requirements. French and German talkers were paid $12/hour for their participation.

B. Speech materials

The three groups of talkers produced native-language vowels in a word or nonsense word context embedded in a language-appropriate sentence frame. Eleven American English vowels (/æ, ɛ, ɪ, $e^{ɪ}$, i, ɑ, ʌ, ʊ, $o^{ʊ}$, u, ɜ˞/) were produced in /b_t/ context within the sentence frame, "The large /*b*V̲*t*/ is green." Nine French vowels (/i, ɛ, a, ɑ, o, y, ɔ, œ, u/) were spoken in /p_t/ context within the sentence frame, "La grande /*p*V̲*t*/ est bleue," which translates as "The large /*p*V̲*t*/ is blue." Fifteen German vowels (7 short, /ɑ, ɛ, ɪ, ɔ, œ, ʊ, ʏ/ and eight long, /ɑː, ɛː, eː, iː, oː, øː, uː, ʏː/) were spoken in /b_tə/ context within the sentence frame "Matthias sagt /bVtən/," which translates as "Matthias says /bVtɔn/." Whereas the German vowel set corresponded to the entire German vowel inventory, excluding diphthongs, the American English and French sets were subsets of their respective inventories. The English set lacked the vowel /ɔ/ (which is usually not distinguished from /ɑ/ in Central Texas dialect) and the diphthongs /$ɑ^{ɪ}$/, /$ɑ^{ʊ}$/, and /$ɔ^{ɪ}$/; the French set lacked the close mid vowels /e/ and /ø/ (which are similar to their open mid counterparts) and all nasalized vowels.

C. Elicitation and recording procedure

Talkers were given an orthographic version of each sentence they were to produce (see the Appendix for orthographic representations of test words). They were then presented with a written question about the sentence that was designed to encourage them to place the focus on either the target word (i.e., the word containing the target vowel), or on the final adjective in the sentence. In English, for example, when talkers read the question, "The large WHAT is blue?," the target word represented new information for the hypothetical person who had asked the question (i.e., the target word was marked [+focus]). When talkers read the question, "The large /*b*V̲*t*/ is WHAT?," the target word represented given information for the hypothetical question asker (i.e., the target word was marked [−focus]). Talkers were instructed to produce the entire sentence when responding to the question, and their responses were recorded. They were also instructed to answer the questions at a normal rate of speech.[3] For a given vowel, talkers produced each focus condition before moving on to the next vowel. The order of focus condition was randomized between vowels, and the order of vowels was randomized between talkers. Productions by English and German speakers were digitally recorded at a sample rate of 22 050 Hz using a Shure SM48 microphone and a locally developed signal processing program (WAVAX) implemented on a PC. Utterances of French speakers were recorded using a Shure HW501 headworn dy-

namic microphone and a Sony Mini Disc MZ-R70 recorder and were then imported into WAVAX at the same sampling rate. All recordings were made either in a sound-attenuated booth or in a quiet room.

D. Acoustic measurements

Vowel tokens were analyzed acoustically using the Multi-Speech Signal Analysis Workstation, Model 3700, Version 2.4 (Copyright© 2001 Kay Elemetrics Corp). Temporal measurements were made from inspection of wave forms and wideband spectrograms (323 Hz bandwidth). Vowel durations were measured as the interval between the first and last significant glottal pulses of a given token. As noted earlier, one possible means of distinctiveness enhancement of vowels in [+focus] context is to increase systematic variation in vowel duration (e.g., between tense and lax vowels in English or between long and short vowels in German). Accordingly, for all three languages examined in this study, we calculated the variance in vowel durations for each talker and condition.

Frequencies of the first two formants (F1 and F2) were measured using pitch synchronous LPC (18 coefficients) with a 20 ms Blackman analysis window centered at each of three relative temporal locations—25%, 50%, and 75% of vowel duration. LPC derived formant tracks were superimposed on the spectrograms. In cases where the formant tracks reflected spurious peaks or where actual formants were missed, formant frequencies were marked manually using visual inspection of the spectrograms and of 512-point FFT spectra. The F1 and F2 measurements were used to determine the size of the vowel space and the amount of VISC for each talker.

E. Auditory modeling and measurements

Formant frequency measurements are very useful for making comparisons with other studies of vowels. However, the motivating idea of this study—sufficient contrast—must ultimately be evaluated in auditory, rather than purely acoustic, terms. Accordingly, an auditory model was used to produce auditory spectra of each vowel token at the 25%, 50%, and 75% relative time locations. These spectral data were analyzed using principal components analysis (PCA), and the first two extracted components (PC1 and PC2) served to define a two-dimensional space (analogous to an $F1 \times F2$ space) within which vowel space size and amount of VISC could be calculated.

The auditory spectra were plotted in units of loudness (Sones) at each 0.2 Bark step, for English and French, and at each 0.1 Bark step for German, between 0 and 16 Bark (corresponding to the frequency range between 0 and 3200 Hz). (The use of a smaller Bark step for German, corresponding to a smaller analysis bin, was necessitated by the relatively low fundamental frequency values of several male talkers, which would otherwise have produced more than one harmonic per analysis bin.) These auditory spectra were derived using a modified version of the auditory model described by Bladon and Lindblom (1981), which was itself adapted from Schroeder *et al.* (1979). This model was judged to be suitable for our purposes because it was based on human psychophysical measurements rather than on physical or electrophysiological data from animal listeners. Although the Bark frequency scale of the model was retained in the present application, we replaced the Bark-based auditory filters with filters having somewhat narrower equivalent rectangular bandwidth (ERB) values as described by Moore and Glasberg (1983) and Glasberg and Moore (1990).

The Bladon and Lindblom (1981) model operates on schematic vowel spectra each consisting of a list of harmonic frequencies and amplitudes. In order to apply the model to actual vowel signals, it is necessary to preprocess the signals, converting actual spectra into schematic spectra. This was done as follows: A Hanning window was applied to a selected temporal region of the vowel consisting of at least three glottal periods, and an FFT was calculated. (Hanning windowing results in a spectrum consisting of triangular shaped harmonics.) A set of frequency-domain template functions were created each consisting of a series of triangular pulses of unit height and spaced at equal frequency intervals, with the size of the frequency interval varying in very small steps across the template set. A given FFT spectrum was multiplied by each template function of the set, and the template yielding the highest product was used to estimate the harmonic frequencies of the FFT spectrum. The corresponding harmonic amplitudes were then determined directly from the FFT spectrum. Visual inspection of sample cases confirmed that this procedure yielded highly accurate results.

F. Calculating vowel space size and amount of VISC

For both the acoustic ($F1 \times F2$) and auditory ($PC1 \times PC2$) spaces, two methods were used to calculate the size of the vowel space. These methods are described here in terms that apply equally to the acoustic and auditory spaces. The first method was to calculate the area of the triangle (using Heron's formula) defined by the vowels /i/, /ɑ/, and /u/, with the analysis window centered at the 50% time point of each vowel trajectory. An advantage of this quantity is that it corresponds to a typical graphic representation of vowel space size in two dimensions. The second method was to calculate the average Euclidean distance between every pair of vowels in the entire vowel set. Relative to the first method, this one yields a more comprehensive measure of vowel dispersion. The quantities calculated by these methods are referred to here as size (area) and size (dispersion), respectively.

Two methods were also used to calculate the amount of VISC according to both the acoustic and auditory measures. The first was to sum the Euclidean distances traversed by a given vowel (in the $F1 \times F2$ space or the $PC1 \times PC2$ space) between the 25% and 50% time points and between the 50% and 75% time points and then to average these sums across the entire vowel set. This quantity, which is referred to as VISC (magnitude), approximates the degree of spectral change across the vowel nucleus while limiting somewhat the contribution of the initial and final consonants to spectral dynamics. It can reasonably be viewed as a valid measure of

the amount of VISC. However, by itself this quantity does not provide for an adequate test of the notion of sufficient contrast. The reason is that there is no guarantee that a greater amount of VISC, defined in this way, will correspond to greater acoustic or auditory distinctiveness among vowels.

The second method for calculating the amount of VISC was an attempt to remedy this limitation. It was analogous to the size (dispersion) method for calculating vowel space size. For each stimulus, the signed differences in F1 and F2 (or PC1 and PC2) were calculated between the 25% and 50% time points and between the 50% and 75% time points. For each of the two time intervals, every vowel token was represented in a space whose dimensions corresponded to these signed differences (i.e., $F1_{diff} \times F2_{diff}$, or $PC1_{diff} \times PC2_{diff}$), and the average Euclidean distance between every pair of vowel tokens was determined. Finally, these two averages were themselves averaged across the two time intervals. This quantity, which is referred to as VISC (dispersion), yields a measure of vowel distinctiveness that is independent of size (dispersion).

III. RESULTS

For each language and for each measure, a 2 (focus condition) $\times 2$ (sex of talker) repeated-measures ANOVA was performed. In the case of German, ANOVAs for size (area) were performed on the short vowel and long vowel sets separately, whereas those for size (dispersion) and all other measures were performed on the entire vowel set.

A. Vowel space size

1. Acoustic space: F1 × F2

a. Size (area): For all three languages, there was a significant effect of focus condition (English: $[F(1,10)=10.70,\ p<0.01,\ \eta_p^2=0.517]$, French: $[F(1,10)=15.86,\ p<0.005,\ \eta_p^2=0.613]$, German long: $[F(1,12)=13.46,\ p<0.005,\ \eta_p^2=0.529]$ and German short: $[F(1,12)=15.79,\ p<0.005,\ \eta_p^2=0.568]$), with [+focus] vowels having a larger size (area) than [−focus] vowels. Sex of talker was not significant in either English or French but was significant for both German long $[F(1,12)=48.58,\ p<0.001,\ \eta_p^2=0.802]$ and short $[F(1,12)=21.04,\ p<0.005,\ \eta_p^2=0.637]$ vowels. Figures 1–4 show the effects of focus condition and sex of talker on the size of the acoustically defined vowel space for each of the three languages.

b. Size (dispersion): Figure 5(a) shows the effect of focus condition on the acoustically defined measure of size (dispersion). This effect was significant for English $[F(1,10)=33.10,\ p<0.001,\ \eta_p^2=0.768]$, French $[F(1,10)=41.61,\ p<0.001,\ \eta_p^2=0.806]$, and German $[F(1,12)=37.14,\ p<0.001,\ \eta_p^2=0.756]$, with [+focus] vowels having greater size (dispersion) than [−focus] vowels. The effect of sex of talker was significant only in German ($[F(1,12)=19.12,\ p<0.005,\ \eta_p^2=0.614]$), with German females having greater size (dispersion) than German males.

2. Auditory space: PC1 × PC2

a. Size (area): For English and French, there was a significant effect of focus condition on the size of the auditorily defined vowel space ($[F(1,10)=10.37,\ p<0.01,\ \eta_p^2=0.509]$ and $[F(1,10)=9.78,\ p<0.05,\ \eta_p^2=0.499]$, respectively), with the [+focus] vowels having a larger vowel space than [−focus] vowels. German short $[F(1,12)=19.91,\ p<0.005,\ \eta_p^2=0.624]$, but not long vowels ($p=0.072$), were produced with significantly more auditory separation in the [+focus] condition. Sex of talker was not a significant variable in any of the three languages. The results patterned very similarly for the acoustically and auditorily defined vowel spaces. Table I displays correlations, for each of the three languages, between the first and second principal components of the auditorily defined space and F1 and F2.

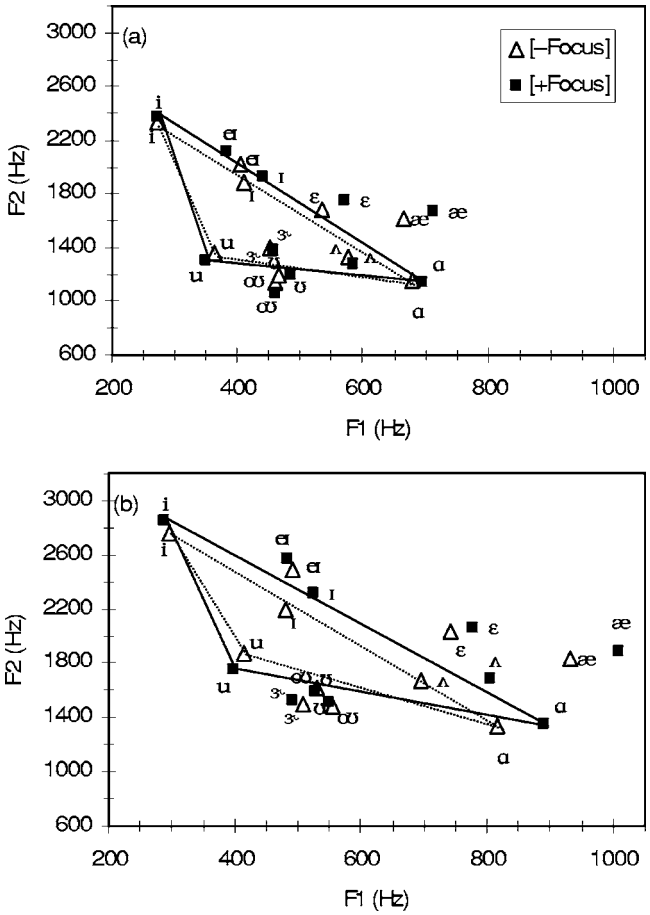


FIG. 1. Effect of focus condition on size (area) for English vowels: (a) males, (b) females. Dotted and solid lines represent the /i/-/ɑ/-/u/ vowel triangles for the [−focus] and [+focus] conditions, respectively.

b. Size (dispersion): Figure 5(b) displays the effect of focus condition on the auditorily defined measure of size (dispersion). For all three languages, [+focus] vowels showed greater size (dispersion) than [−focus] vowels (English: $[F(1,10)=49.76,\ p<0.001,\ \eta_p^2=0.833]$, French $[F(1,10)=15.04,\ p<0.005,\ \eta_p^2=0.601]$, German $[F(1,12)=15.14,\ p<0.005,\ \eta_p^2=0.558]$. As in the case of the auditorily defined size (area) measure, sex of talker was not a significant variable for any of the three languages.

B. VISC

1. Acoustic space: F1 × F2 for VISC (magnitude); $F1_{diff} \times F2_{diff}$ for VISC (dispersion)

a. VISC (magnitude): Figure 6(a) shows the effect of focus condition on VISC (magnitude) for the three languages. Focus condition was a significant variable only for

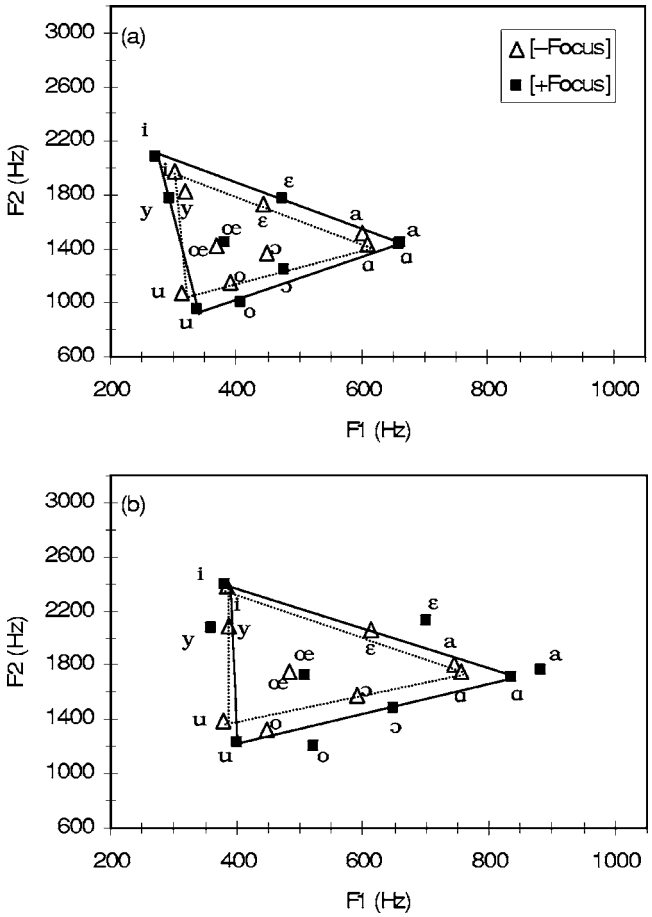


FIG. 2. Effect of focus condition on size (area) for French vowels: (a) males, (b) females. Dotted and solid lines represent the /i/-/ɑ/-/u/ vowel triangles for the [−focus] and [+focus] conditions, respectively.

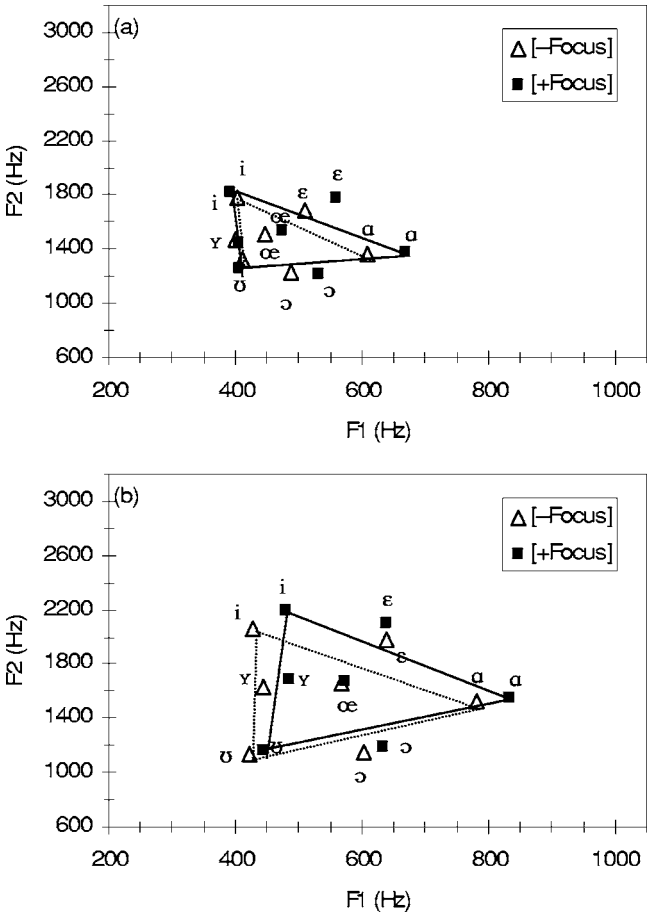


FIG. 3. Effect of focus condition on size (area) for German short vowels: (a) males, (b) females. Dotted and solid lines represent the /i/-/ɑ/-/u/ vowel triangles for the [−focus] and [+focus] conditions, respectively.

German [$F(1,12)=14.54$, $p<0.005$, $\eta_p^2=0.548$], with [+focus] vowels showing greater VISC (magnitude) than [−focus] vowels. For all three languages, female talkers showed significantly greater VISC (magnitude) than male talkers (English: [$F(1,10)=13.57$, $p<0.005$, $\eta_p^2=0.576$], French: [$F(1,10)=14.87$, $p<0.005$, $\eta_p^2=0.598$], and German: [$F(1,12)=24.76$, $p<0.001$, $\eta_p^2=0.674$]). There were no significant interactions between focus condition and sex of talker for any of the three languages.

b. VISC (dispersion): Figure 6(b) exhibits the effect of focus condition on VISC (dispersion) for the three languages. As in the case of VISC (magnitude), German was the only language to show significantly greater VISC (dispersion) for [+focus] vowels than [−focus] vowels [$F(1,12)=11.87$, $p<0.01$, $\eta_p^2=0.497$]. For each of the three languages, there was a significant effect of sex of talker on the measure of VISC (dispersion) (English: [$F(1,10)=26.12$, $p<0.001$, $\eta_p^2=0.723$], French: [$F(1,10)=10.36$, $p<0.01$, $\eta_p^2=0.509$], and German: [$F(1,12)=33.64$, $p<0.001$, $\eta_p^2=0.737$]), with females showing greater acoustic separation than males. There were no significant interactions between focus condition and sex of talker for any of the languages.

2. Auditory space: PC1×PC2 *for VISC (magnitude);* $PC1_{diff}\times PC2_{diff}$ *for VISC (dispersion)*

a. VISC (magnitude): The effect of focus condition on the auditorily defined measure of VISC (magnitude) is shown in Fig. 6(c). English, French, and German all showed greater VISC for [+focus] vowels than [−focus] vowels ([$F(1,10)=111.11$, $p<0.001$, $\eta_p^2=0.917$], [$F(1,10)=16.41$, $p<0.005$, $\eta_p^2=0.621$], and [$F(1,12)=15.09$, $p<0.005$, $\eta_p^2=0.557$], respectively). None of the languages showed an effect of sex of talker.

As was true for size (area), correlations between the first two principal components of the auditorily defined space and F1 and F2 were generally significant. These are shown in Table I.

b. VISC (dispersion): Figure 6(d) displays the effect of focus condition on the auditorily defined measure of VISC (dispersion). All three languages showed a significant effect of focus condition (English [$F(1,10)=77.05$, $p<0.001$, $\eta_p^2=0.885$], French [$F(1,10)=9.18$, $p<0.05$, $\eta_p^2=0.478$], and German [$F(1,12)=14.10$, $p<0.005$, $\eta_p^2=0.540$]), with [+focus] vowels showing greater separation than [−focus] vowels. There was no significant effect of sex of talker for any of the languages.

Note that for both VISC (magnitude) and VISC (dispersion), the auditorily defined measures yielded more cross-linguistically consistent effects of focus condition than the acoustically defined measures.

In contrast, the auditorily defined measures of both vowel space size and VISC showed, if anything, less sensitivity to the sex of talker variable. This can perhaps be attributed to the use of the Bark scale in the auditory represen-

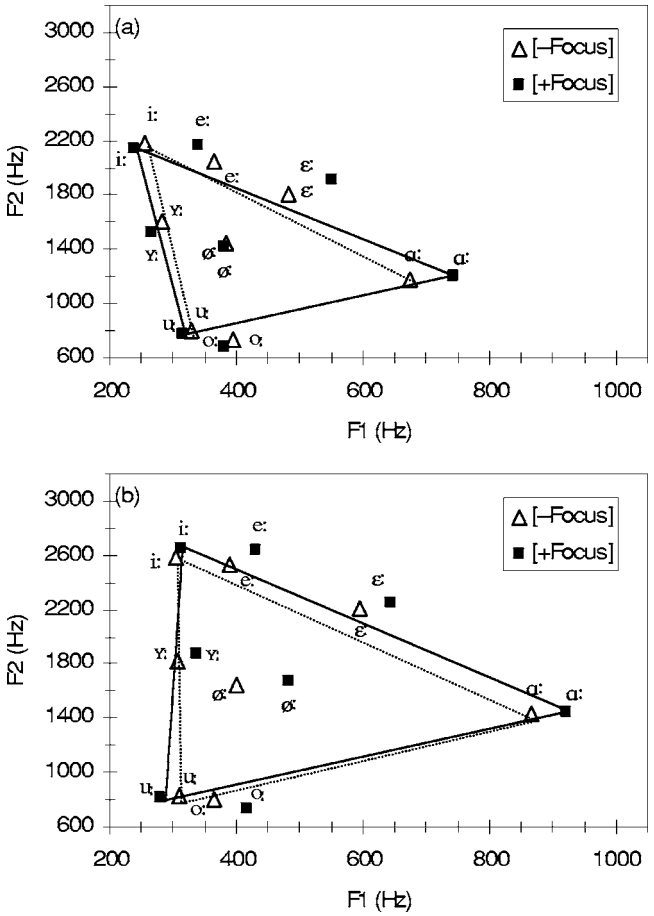


FIG. 4. Effect of focus condition on size (area) for German long vowels: (a) males, (b) females. Dotted and solid lines represent the /i:/-/ɑ:/-/u:/ vowel triangles for the [–focus] and [+focus] conditions, respectively.

tations of vowels, the effect of which is partly to normalize frequency differences between male and female talkers (Syrdal and Gopal, 1986).

Although the measures of VISC (magnitude) and VISC (dispersion) are logically independent, they should be highly correlated provided that vowel tokens are widely distributed in the $F1_{diff} \times F2_{diff}$ space or the $PC1_{diff} \times PC2_{diff}$ space. Table II shows these correlations for each language and for both the acoustic and auditory measures.

C. Duration variance

Neither English nor French showed significant effects of focus condition on vowel duration variance. However, in German there was more duration variance for [+focus] vowels than for [–focus] vowels [$F(1,12)=17.01$, $p<0.005$, $\eta_p^2=0.586$]. There was no significant effect of sex of talker on duration variance in any of the languages. Figure 7 shows the effects of focus condition on duration variance for each of the three languages.

TABLE I. Correlations between auditory principal components and formant frequencies. The asterisks represent that correlation is significant at the 0.01 level (2-tailed).

		Vowel space		VISC	
		PC1	PC2	PC1	PC2
English	F1	0.046	0.471*	0.446*	0.018
English	F2	0.184*	−0.450*	−0.472*	0.171*
French	F1	0.662*	0.289*	0.651*	0.275*
French	F2	−0.359*	0.422*	−0.386*	0.401*
German	F1	0.627*	0.159*	0.623*	0.129*
German	F2	−0.261*	0.452*	−0.295*	0.438*

D. Comparing the English tense/lax distinction and the German long/short distinction

In the analyses described above, each language was treated separately owing to differences among English, French, and German in the size and character of the vowel inventories. However, for a subset of the measurements, it is instructive to perform additional analyses in which language is a factor. This subset includes data for all seven long/short pairs of vowels in German (/ɑ:, ɑ/, /e:, ɛ/, /i:, ɪ/, /o:, ɔ/, /ø:, œ;/, /u:, ʊ/, /ʏ:, ʏ/) as well as for four tense/lax vowel pairs in America English (/ɑ, ʌ/, /e^{ɪ}, ɛ/, /i, ɪ/, /u, ʊ/). The members of each of these vowel pairs are potentially confusable because they occupy similar or adjacent regions in F1×F2 space, although in both languages (and especially in German) this problem is mitigated because of durational differences.

Three sets of measurements were analyzed using 2 (focus conditions)×2 (language) repeated-measures ANOVAs. (For the purpose of these analyses, the data were collapsed over the sex of talker variable.) The first two sets of measurements were equivalent to the auditory versions of size (dispersion) and VISC (dispersion), respectively, except that the stimulus tokens being compared did not comprise the entire vowel set but instead were restricted to the two mem-

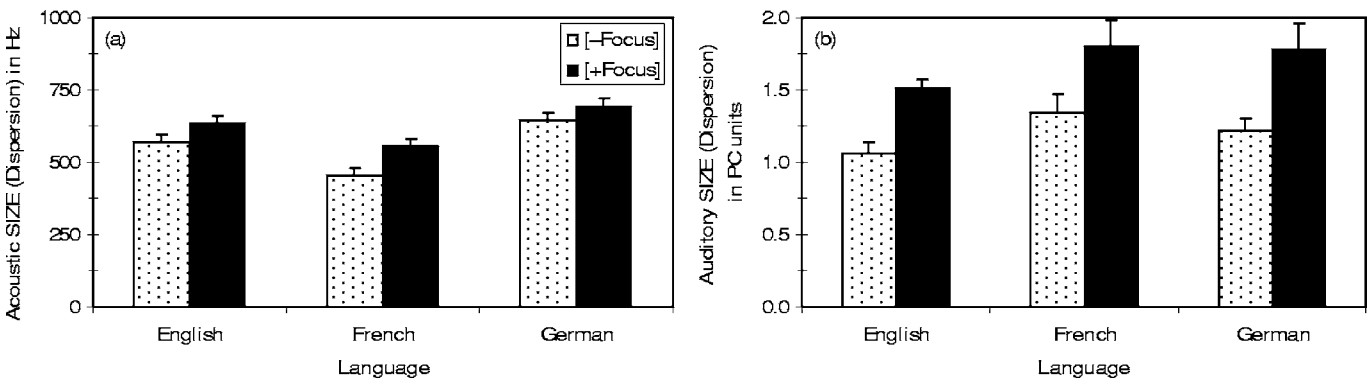


FIG. 5. Effect of focus condition on the (a) acoustically defined and (b) auditorily defined measures of size (dispersion) for English, French, and German. PC units refer to unit distances in the PC1×PC2 space. Error bars represent the standard error of the mean.

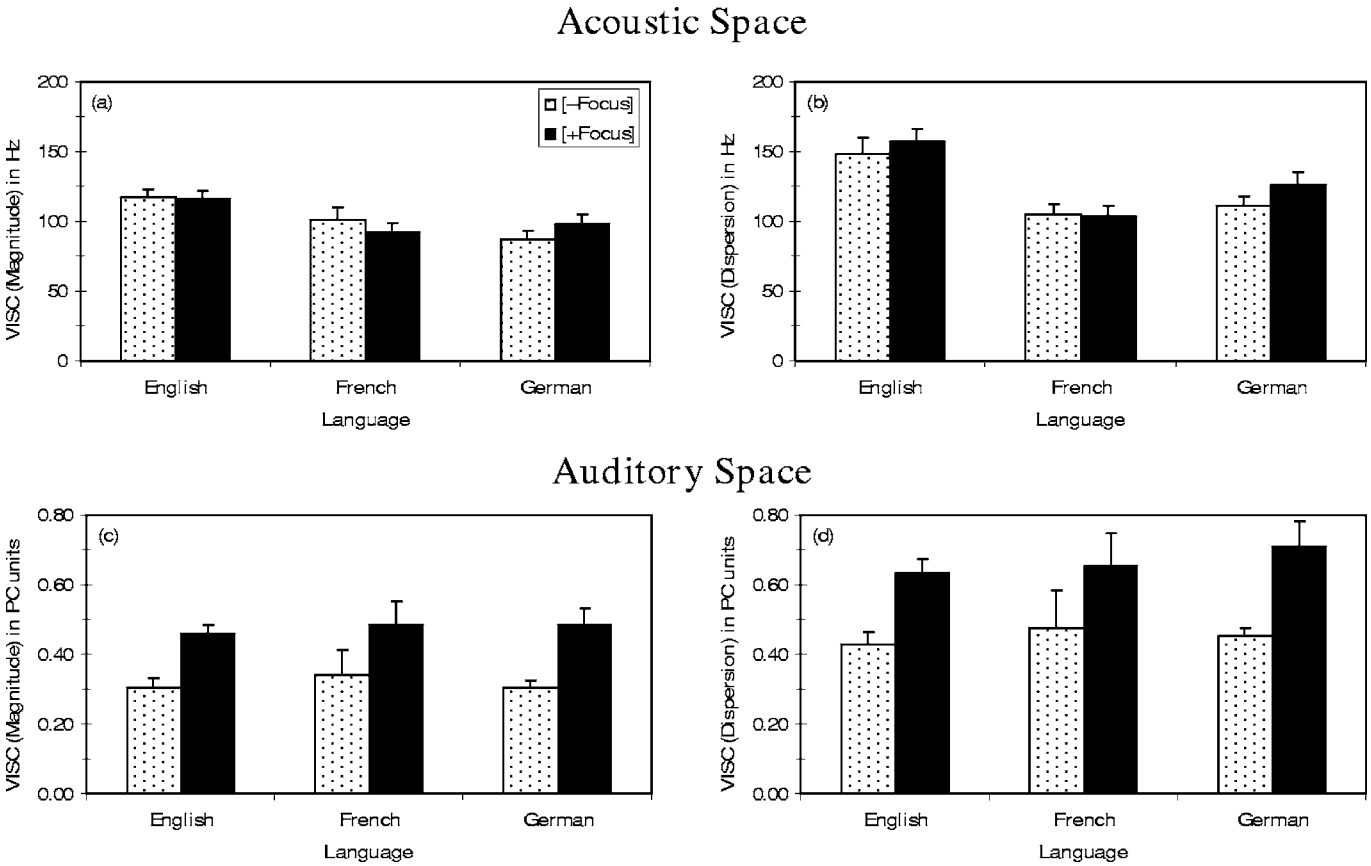


FIG. 6. Effect of focus condition on the acoustically defined measures of (a) VISC (magnitude) and (b) VISC (dispersion) and on the auditorily defined measures of (c) VISC (magnitude) and (d) VISC (dispersion), for English, French, and German. For the size (dispersion) measure, PC units refer to unit distances in the PC1 × PC2 space; for the VISC (dispersion) measure, PC units refer to unit distances in the $PC1_{diff} \times PC2_{diff}$ space. Error bars represent the standard error of the mean.

bers of each long/short or tense/lax pair. The third set of measurements consisted of the duration ratios between long/short or tense/lax pair members.

Table III shows the mean values of size (dispersion), VISC (dispersion), and duration ratio between English tense/lax and German long/short vowel pair members in [+focus] and [−focus] conditions. There were significant effects of focus condition for the modified size (dispersion) [$F(1,24)=31.89$, $p<0.001$, $\eta_p^2=0.571$] and VISC (dispersion) [$F(1,24)=27.13$, $p<0.001$, $\eta_p^2=0.531$] measures, with both yielding larger values in [+focus] condition. For neither measure was there a significant effect of language or a significant focus condition × language interaction.

For the duration ratio measure, there were significant main effects of focus condition [$F(1,24)=4.81$, $p<0.05$, $\eta_p^2=0.167$] and language [$F(1,24)=102.43$, $p<0.001$, $\eta_p^2=0.810$], with vowel pairs in [+focus] condition and in German showing greater duration ratios, but there was no interaction between these variables. Analyses of simple effects showed that focus condition was a significant variable in German [$F(1,13)=8.04$, $p<0.05$, $\eta_p^2=0.382$], but not in English ($p=0.497$). The larger duration ratios for [+focus] vowels in German help to account for the increased overall duration variance reported earlier for German vowels in [+focus] condition. Although the tense/lax duration ratios in English were much smaller than the long/short duration ratios in German, the former were nevertheless significantly greater than 1.0 ($p<0.001$) in both [+focus] and [−focus] conditions.

It is noteworthy that the relative effect of focus condition was considerably greater for the two dispersion measures than for the duration ratio measure. This was no less true in German than in English.

TABLE II. Correlations between VISC (magnitude) and VISC (dispersion) in $F1_{diff} \times F2_{diff}$ and $PC1_{diff} \times PC2_{diff}$ spaces for English, French, and German. The asterisks represent that correlation is significant at the 0.01 level (2-tailed).

	Acoustic ($F1_{diff} \times F2_{diff}$ Space)	Auditory ($PC1_{diff} \times PC2_{diff}$ Space)
English	0.908*	0.964*
French	0.859*	0.990*
German	0.966*	0.994*

IV. DISCUSSION

The present study examined several possible acoustic and auditory correlates of vowels in [+focus] context produced by talkers of American English, French, and German. Consistent with Lindblom's theory of adaptive dispersion, each language group enhanced the distinctiveness of vowels in the portion of an utterance used to signal new (as opposed to given) information, and, as allowed by the principle of sufficient contrast, the three language groups used varying

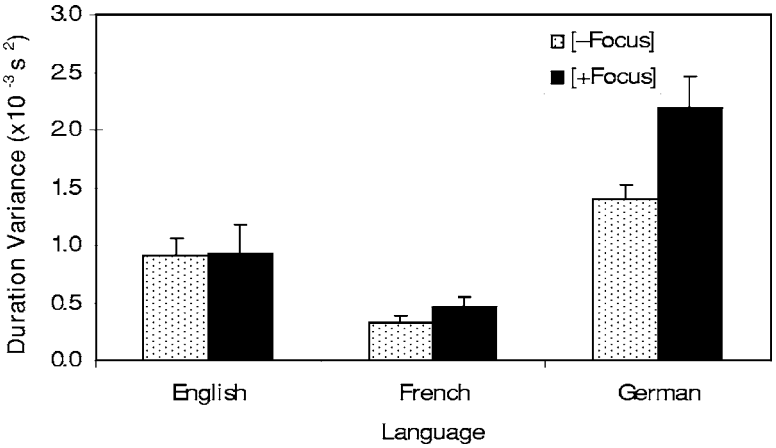


FIG. 7. Effect of focus condition on duration variance in English, French, and German. Error bars represent the standard error of the mean.

means of enhancing distinctiveness. In producing vowels in [+focus] context, all three language groups increased spectral differences among vowels, but only the German talkers increased vowel duration differences. Table IV lists the measures for which the effect of focus condition was significant in each language.

The acoustic and auditory versions of the spectral measures generally yielded similar patterns of results; however, the auditory versions yielded a higher number of significant effects of focus condition on the measures of VISC. We interpret these results as a validation of our auditory measures of spectral distance. Although the vowel space is most commonly defined acoustically, namely, in terms of the frequencies of the first several formants, there are several advantages in using an auditorily defined space. First, as noted earlier, the notion of sufficient contrast is more directly evaluated in terms of auditory rather than acoustic distance. Second, automatic formant tracking procedures are error-prone and often need to be supplemented with manual editing. In contrast, the auditory space used in the present study is derived entirely by algorithmic means. Third, there is little reason to believe that, unlike artificial systems, the human auditory system some how manages to circumvent the inherent difficulties in extracting formant frequencies reliably. It seems more plausible to assume that the auditory system extracts parameters that are rich in phonetic information but that can be computed algorithmically from the speech signal. Among possible candidates for such parameters, we would include stimulus values within a PCA-defined auditory space such as that used in the present study.[4] Previous factor analyses of auditory spectra have yielded principal components that account well for the variance among vowel sounds (Plomp *et al.*, 1967; Klein *et al.*, 1970) and that are reasonably highly correlated with formant frequencies (Nearey and Kiefte, 2003). Such correlations were also found in this study. Factor analytic approaches in these cases may thus be described as yielding implicit information about formant values without requiring actual formant tracking.

Finally, we return to the topic of cross-language variation in the means by which vowel distinctiveness is enhanced in [+focus] contexts. One possible account of this variation is based on the distinction between phonologically contrastive and noncontrastive, or secondary, properties. Consider, for example, the case of vowel duration differences. In German, such differences are used to signal phonological contrasts between long and short vowels of similar quality (e.g., /ɑː/ versus /ɑ/). However, in English, vowel duration differences are generally viewed as a type of noncontrastive variation that is correlated with and conditioned by some primary phonological contrast (e.g., tense versus lax vowels and nonlow versus low vowels). As in other languages that make contrastive use of vowel length (Heldner and Strangert, 2001; de Jong and Zawaydeh, 2002; Smiljanic, 2004), duration ratios of German long/short vowel pairs were significantly increased in [+focus] context. However, no such change occurred for English tense/lax vowel pairs. This pat-

TABLE III. Mean auditory size (dispersion), VISC (dispersion), and duration ratio for English tense/lax and German long/short vowel pairs. For the size (dispersion) measure, PC units refer to unit distances in the PC1 × PC2 space; for the VISC (dispersion) measure, PC units refer to unit distances in the $PC1_{diff} \times PC2_{diff}$ space.

	English		German	
	[+Focus]	[–Focus]	[+Focus]	[–Focus]
Size (dispersion) in PC units	1.056	0.761	1.266	0.719
VISC (dispersion) in PC units	0.578	0.396	0.688	0.396
Duration ratio	1.238	1.183	2.126	1.981

TABLE IV. Significant effects of focus condition on distinctiveness enhancement in vowels. * represents significant at the 0.05 level; ** significant at the 0.01 level; *** significant at the 0.001 level.

	English	French	German
Acoustic size(area)	**	**	**
Acoustic size(dispersion)	***	***	***
Auditory size(area)	**	**	Long=ns, Short***
Auditory size(dispersion)	***	**	**
Acoustic VISC(magnitude)	ns	ns	**
Acoustic VISC(dispersion)	ns	ns	**
Auditory VISC(magnitude)	***	**	**
Auditory VISC(dispersion)	***	*	**
Duration variance	ns	ns	***

tern of results might suggest that in [+focus] context talkers exaggerate only those vowel differences that are phonologically contrastive.

However, this account of cross-language differences in vowel contrast enhancement faces several difficulties. First, some noncontrastive properties of phonemes do appear to be subject to enhancement in [+focus] contexts. In modern theories of distinctive features (for a review, see Diehl and Lindblom, 2004), VISC has not generally been recognized as being phonologically contrastive. It is more likely to be viewed as a secondary correlate of certain primary vowel features (e.g., tenseness or height). Yet, in the present study, talkers from all three language groups exhibited increased VISC (magnitude) or increased VISC (dispersion) in [+focus] context, by either the acoustic measure, the auditory measure, or both. Analogously, it has been reported that duration differences of English stressed vowels before [+voice] and [−voice] stop consonants are enhanced in [+focus] context (de Jong, 2004), even though such differences have typically been viewed as a secondary correlate of the syllable-final [voice] distinction. Thus, status as a phonologically contrastive feature of a language does not appear to be a necessary condition for enhanced distinctiveness in utterance focus.

Another difficulty in appealing to the contrastive/noncontrastive distinction to account for cross-language differences in vowel contrast enhancement is that the distinction itself is difficult to apply in practice and may be theoretically problematic. For example, German vowels differing in length have been variously described by linguists as tense/lax contrasts or as long/short contrasts (Hawkins, 1992). Phonetically, the members of these vowel pairs differ both in spectral properties and in duration (Fischer-Jørgensen, 1990), and it is unclear which of these differences should be viewed as contrastive and which as noncontrastive (or secondary). Because phonological features may be quite abstract, mapping onto diverse phonetic correlates each having perceptual relevance (Kingston and Diehl, 1994), the assignment of contrastive status to one correlate and secondary status to others may be quite arbitrary.

An alternative hypothesis is that the amount of contrast enhancement of a vowel property in [+focus] context is positively related to the between-category variation of that property in [−focus] context. Although the spectral measures did not provide evidence for or against this hypothesis (because the language groups did not differ very much according to these measures in either the [−focus] or the [+focus] conditions), the duration variance and duration ratio measures did yield results consistent with the hypothesis. Duration variance among [−focus] vowels was much larger in German than in English and French, and significant enhancement of duration variance in [+focus] context occurred only for German.[5] Similarly, duration ratios between German long/short vowel pairs were much larger than those between English tense/lax vowel pairs in [−focus] condition, and only German vowels exhibited significant enhancement of duration ratios in [+focus] condition. Some previous cross-language results also appear to be consistent with the above hypothesis. For example, de Jong and Zawaydeh (2002) and de Jong (2004) found that vowel duration differences conditioned by a following [voice] distinction are greater in English than in Arabic (see also Port *et al.*, 1980, and Flege and Port, 1981), and that [+focus] contexts yield a reliable enhancement of these differences only in English. A tendency to enhance properties of phonemes that show greater between-category variance may be viewed as yet another instance of the positive relationship, predicted by the theory of adaptive dispersion, between utterance clarity and information content.

ACKNOWLEDGMENTS

This work was supported by research grants Nos. R01 DC00427-13, -14, -15 from the National Institute of Health, National Institute on Deafness and Other Communication Disorders, to the fifth author (R.L.D.) Special thanks to Juan Segui and Ludovic Ferrand of the CNRS and Université René Descartes, Laboratoire de Psychologie Expérimental, Paris, France, for providing facilities to conduct the French portion of the study. Some of these findings were reported at the 143rd Meeting of the Acoustical Society of America, Pittsburgh, June 4, 2002 and at the 146th Meeting of the Acoustical Society of America, Austin, November 11, 2003. The authors thank Ann Bradlow for very helpful comments on earlier versions of this paper.

APPENDIX

The following orthographic representations of test words were used:

English material

bat bæt	*bet* bɛt	*bit* bɪt	*bait* beˈt	*beet* bit
baht bat **	*butt* bʌt	*bout* bʊt rhymes w/'put' **	*boat* boᵘt	*boot* but
bert bɜˑt **				

French material

pites pit **	*pète* pɛt 'to break' (present tense)	*patte* pat 'paw'	*pâte* pɑt 'pasta'	*peaute* pot **
pute pyt 'hooker'	*pote* pɔt 'friend' (colloquial)	*peute* pœt **	*poute* put **	

German material

Batten bɑtən **	*Betten* bɛtən 'beds'	*Bitten* bɪtən 'to ask'	*Botten* bɔtən **	*Bötten* bœtən **
Butten bʊtən **	*Bütten* bʏtən 'tubs'	*Baten* bɑːtən 'asked'	*Bäten* bɛːtən 'asked' (subj)	*Beten* beːtən 'pray'
Bieten biːtən 'to offer'	*Boten* boːtən 'boats'	*Böten* bøːtən **	*Buten* buːtən **	*Büten* bʏːtən **

**Nonsense words.

[1]Under noisy conditions, clear speech has also been found to be more intelligible than reduced speech, other things being equal, for normal hearing adults (Tolhurst, 1957) and children with learning disabilities (Bradlow *et al.*, 2003).

[2]Apart from signaling new information, [+focus] may also be used to highlight thematic relations between a constituent and its context (Nooteboom and Kruyt, 1987).

[3]For the American English and French groups but not the German group, the same questions were also answered using fast and slow rates of utterance. For a given vowel, talkers produced each focus condition at every rate before moving on to the next vowel. The order of focus condition and utterance rate was randomized between vowels. Because focus condition was the key variable of interest in the present study and because the German group produced utterances only at the normal rate, we include here only results for the normal utterance rate. In general, the pattern of results obtained for the normal utterance rate was very similar to the patterns for the other two utterance rates.

[4]The demand for efficient neural coding of signals suggests that some form of factor analysis (e.g., PCA or independent components analysis) may be implicit in the design of sensory/perceptual systems (Field, 1987; Simoncelli and Olshausen, 2001; Lewicki, 2002).

[5]A potential concern with the participant selection in the present study is that, unlike the American English and French groups, the German group was currently visiting or residing in Autsin, Texas, rather than living in their native language community. However, if exposure to English affected the German vowel productions, this was clearly not sufficient to eliminate differences between English and German modes of contrast enhancement.

Altman, G., and Carter, D. M. (**1989**). "Lexical stress and lexical discriminability: Stressed syllables are more informative, but why?," Comput. Speech Lang. **3**, 265–275.

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., and Doherty-Sneddon, G. (**1997**). "Limited visual control of the intelligibility of speech in face-to-face dialogue," Percept. Psychophys. **59**, 580–592.

Baart, J. J. G. (**1987**). "Focus, syntax, and accent placement," Ph.D. thesis, Leyden University.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (**2000**). "Controlling the intelligibility of referring expressions in dialogue," J. Mem. Lang. **42**, 1–22.

Billerey-Mosier, R. (**2000**). "Segment realization in relation to a carrier word's uniqueness point," J. Acoust. Soc. Am. **108**, 2479–2480.

Bladon, R. A. W., and Lindblom, B. (**1981**). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am. **69**, 1414–1422.

Bolinger, D. (**1989**). *Intonation and Its Uses* (Stanford University Press, Stanford).

Bradlow, A. R., Kraus, N., and Hayes, E. (**2003**). "Speaking clearly for children with learning disabilities: Sentence perception in noise," J. Speech Lang. Hear. Res. **46**, 80–97.

Browman, C. P., and Goldstein, L. (**1995**). "Gestural syllable position effects in American English," in *Producing Speech: Contemporary Issues. For Katherine Harris*, edited by F. Bell-Berti and L. J. Raphael (AIP Press, New York), pp. 19–33.

Byrd, D. (**1994**). "Articulatory timing in English consonant sequences," Ph.D. thesis, University of California, Los Angeles.

Campbell, N. (**1995**). "Loudness, spectral tilt, and perceived prominence in dialogues," 13th International Congress of Phonetic Sciences (Stockholm, Sweden), pp. 676–679.

Cho, T., and Keating, P. (**2001**). "Articulatory strengthening at the onset of prosodic domains in Korean," J. Phonetics **28**, 155–190.

Cho, T., and McQueen, J. M. (**2005**). "Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress," J. Phonetics **33**, 121–157.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper and Row, New York).

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). "Acoustical aspects of contrastive stress in question-answer contexts," J. Acoust. Soc. Am. **77**, 2142–2156.

de Jong, K. (**2004**). "Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration," J. Phonetics **32**, 493–516.

de Jong, K., and Zawaydeh, B. (**2002**). "Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration," J. Phonetics **30**, 53–75.

Diehl, R. L., and Lindblom, B. (**2004**). "Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 101–162.

Eefting, W. (**1991**). "The effect of 'information values' and 'accentuation' on the duration of Dutch words, syllables, and segments," J. Acoust. Soc. Am. **89**, 412–424.

Field, D. J. (**1987**). "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A **4**, 2379–2394.

Fischer-Jørgensen, E. (**1990**). "Intrinsic F_0 in tense and lax vowels with special reference to German," Phonetica **47**, 99–140.

Flege, J. E., and Port, R. F. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang Speech **24**, 125–146.

Fowler, C. A., and Housum, J. (**1987**). "Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction," J. Mem. Lang. **26**, 489–504.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gussenhoven, C. (**1983**). "Focus, mode, and the nucleus," J. Ling. **19**, 377–417.

Hawkins, J. A. (**1992**). "German," in *International Encyclopedia of Linguistics*, edited by W. Bright (Oxford University Press, New York), pp. 61–69.

Heldner, M. (**2003**). "On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish," J. Phonetics **31**, 39–62.

Heldner, M., and Strangert, E. (**2001**). "Temporal effects of focus in Swedish," J. Phonetics **29**, 329–361.

Heldner, M., Strangert, E., and Deschamps, T. (**1999**). "A focus detector using overall intensity and high frequency emphasis," 14th International Congress of Phonetic Sciences (San Francisco, CA).

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3609–3523.

Hunnicutt, S. (**1985**). "Intelligibility vs redundancy—conditions on dependency," Lang Speech **28**, 47–56.

Jakobson, R., Fant, C. G. M., and Halle, M. (**1963**). *Preliminaries to Speech Analysis* (MIT Press, Cambridge).

Jun, S.-A. (**1993**). "The phonetics and phonology of Korean prosody," Ph.D. thesis, Ohio State University.

Keating, P., Cho, T., Fougeron, C., and Hsu, C. (**1999**). "Domain-initial articulatory strengthening in four languages," UCLA Working Papers in Linguistics **97**, 139–156.

Kingston, J., and Diehl, R. L. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Klein, W., Plomp, R., and Pols, L. C. W. (**1970**). "Vowel spectra, vowel spaces and vowel identification," J. Acoust. Soc. Am. **48**, 999–1009.

Knoll, M., and Uther, M. (**2004**). "Motherese and Chinese: Evidence of acoustic changes in speech directed at infants and foreigners," J. Acoust. Soc. Am. **116**, 2522.

Ladd, D. R. (**1980**). "The structure of intonational meaning," Ph.D. thesis, Indiana University.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).

Ladd, D. R., and Cutler, A. (**1983**). "Introduction: models and measurements in the study of prosody," in *Prosody: Models and Measurements*, edited by A. Cutler and D. R. Ladd (Springer-Verlag, Berlin), pp. 39–52.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Lane, H., Tranel, B., and Sisson, C. (**1970**). "Regulation of voice communication by sensory dynamics," J. Acoust. Soc. Am. **47**, 618–624.

Lehiste, I. (**1970**). *Suprasegmentals* (MIT Press, Cambridge).

Lewicki, M. S. (**2002**). "Efficient coding of natural sounds," Nat. Neurosci. **5**, 356–363.

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–187.

Lindblom, B. (**1986**). "Phonetic universals in vowel systems," in *Experimental Phonology*, edited by J. J. Ohala and J. J. Jaeger (Academic, Orlando, FL), pp. 13–44.

Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer, Dordrecht), pp. 403–439.

Lombard, E. (**1911**). "Le signe d'elevation de la voix," Annales Maladies de l'Oreille, Larynx, Nez et Pharynx **37**, 101–119.
Luce, P. (**1986**). "Neighborhoods of words in the mental lexicon," Ph.D. thesis, Indiana University.
Lyberg, B. (**1979**). "Final lengthening—partly a consequence of restrictions on the speed of fundamental frequency change?," J. Phonetics **7**, 187–196.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1308.
Nearey, T. M., and Kiefte, M. (**2003**). "Comparison of several proposed perceptual representations of vowel spectra," 15th International Congress of Phonetic Sciences (Barcelona, Spain), pp. 1005–1008.
Nooteboom, S. G. (**1972**). "Production and perception of vowel duration," Ph.D. thesis, University of Utrecht.
Nooteboom, S. G., and Kruyt, J. G. (**1987**). "Accents, focus distribution, and the perceived distribution of given and new information: An experiment," J. Acoust. Soc. Am. **82**, 1512–1524.
O'Shaughnessy, D. (**1979**). "Linguistic features in fundamental frequency patterns," J. Phonetics **7**, 119–145.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1985**). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," J. Speech Hear. Res. **28**, 96–103.
Picheny, M. A., Durlach, N. I., and Braida, L. D. (**1986**). "Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–446.
Plomp, R., Pols, L. C. W., and Van der Geer, J. P. (**1967**). "Dimensional analysis of vowel spectra," J. Acoust. Soc. Am. **41**, 707–712.
Pollack, I., and Pickett, J. (**1963**). "The intelligibility of excerpts from conversation," Lang Speech **6**, 165–171.
Port, R. F., Al-Ani, S., and Maeda, S. (**1980**). "Temporal compensation and universal phonetics," Phonetica **37**, 235–252.
Redford, M. A., and Diehl, R. L. (**1999**). "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," J. Acoust. Soc. Am. **106**, 1555–1565.
Schroeder, M. R., Atal, B. S., and Hall, J. L. (**1979**). "Objective measure of certain speech signal degradations based on masking properties of human auditory perception," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Öhman (Academic, London), pp. 217–229.
Selkirk, E. O. (**1984**). *Phonology and Syntax* (MIT Press, Cambridge).
Simoncelli, E. P., and Olshausen, B. A. (**2001**). "Natural image statistics and neural representation," Annu. Rev. Neurosci. **24**, 1193–1215.
Sluijter, A. M. C., and van Heuven, V. J. (**1995**). "Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch," Phonetica **52**, 71–89.
Sluijter, A. M. C., and van Heuven, V. J. (**1996a**). "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.
Sluijter, A. M. C., and van Heuven, V. J. (**1996b**). "Acoustic correlates of linguistic stress and accent in Dutch and American English," 4th International Conference of Spoken Language Processing (University of Delaware), pp. 630–633.
Smiljanic, R. (**2004**). *Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian: An Acoustic Study* (Routledge, New York).
Strangert, E., and Heldner, M. (**1998**). "On the amount and domain of focal lengthening in Swedish," 5th International Conference on Spoken Language Processing (Sydney, Australia), pp. 3305–3308.
Syrdal, A. K., and Gopal, H. S. (**1986**). "A perceptual model of vowel recognition based on the auditory representation of American English vowels," J. Acoust. Soc. Am. **79**, 1086–1100.
't Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation* (Cambridge University Press, Cambridge).
Tolhurst, G. C. (**1957**). "Effects of duration and articulation changes on intelligibility, word reception, and listener preference," J. Speech Hear Disord. **22**, 328.
Traunmüller, H., and Eriksson, A. (**2000**). "Acoustic effects of variation in vocal effort by men, women, and children," J. Acoust. Soc. Am. **107**, 3438–3451.
Turk, A. E., and Sawusch, J. R. (**1997**). "The domain of accentual lengthening in American English," J. Phonetics **25**, 25–41.
van Bergem, D. R. (**1993**). "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," Speech Commun. **12**, 1–23.
van Son, R. J. J. H., and Pols, C. W. (**2003**). "Information structure and efficiency in speech production," Eurospeech/Interspeech 2003 (Geneva, Switzerland), pp. 211–214.
Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.
Wright, R. (**2003**). "Factors of lexical competition in vowel articulation," in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge), pp. 75–87.

Comparing the rhythm and melody of speech and music: The case of British English and French

Aniruddh D. Patel,[a] John R. Iversen, and Jason C. Rosenberg
The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, California 92121

(Received 8 September 2005; revised 1 February 2006; accepted 2 February 2006)

For over half a century, musicologists and linguists have suggested that the prosody of a culture's native language is reflected in the rhythms and melodies of its instrumental music. Testing this idea requires quantitative methods for comparing musical and spoken rhythm and melody. This study applies such methods to the speech and music of England and France. The results reveal that music reflects patterns of durational contrast between successive vowels in spoken sentences, as well as patterns of pitch interval variability in speech. The methods presented here are suitable for studying speech-music relations in a broad range of cultures. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2179657]

PACS number(s): 43.70.Fq, 43.71.Es, 43.71.An, 43.75.Cd, 43.70.Kv [BHS] Pages: 3034–3047

I. INTRODUCTION

A. Aims

Humans produce organized rhythmic and melodic patterns in two forms: prosody and music. While these patterns are typically studied by different research communities, their relationship has long interested scholars from both fields. For example, linguists have borrowed musicological concepts in building prosodic theories (Liberman, 1975; Selkirk, 1984), and musicologists have used tools from linguistic theory to describe musical structure (Lerdahl and Jackendoff, 1983). Despite this contact at the theoretical level, there has been remarkably little empirical work comparing rhythmic or melodic structure across domains. There are reasons to believe such work is warranted. One such reason, which motivates the current study, is the claim that a composer's music reflects prosodic patterns in his or her native language. This idea has been voiced repeatedly by scholars over the past half century. For example, the English musicologist Gerald Abraham explored this topic at length (1974), citing as one example Ralph Kirkpatrick's comment on French keyboard music:

"Both Couperin and Rameau, like Fauré and Debussy, are thoroughly conditioned by the nuances and inflections of spoken French. On no Western music has the influence of language been stronger." (p. 83)

Kirkpatrick (a harpsichordist and music scholar) was claiming that French keyboard music sounded like the French language. Similar claims have been made about the instrumental music of other cultures. The linguist Hall, for example, suggested a resemblance between Elgar's music and the intonation of British speech (Hall, 1953). What makes these claims interesting is that they concern *instrumental* music. It might not be surprising if vocal music reflected speech prosody; after all, such music must adapt itself to the rhythmic and melodic properties of a text. In contrast, the notion that speech patterns are mirrored in instrumental music is much more controversial.

While provocative, until recently this idea had not been systematically tested, likely due to a lack of methods for quantifying prosody in a way that could be directly compared to music. Patel and Daniele (2003) sought to overcome this problem (with regard to rhythm) by using a recently-developed measure of temporal patterning in speech, the normalized pairwise variability index or nPVI. The nPVI measures the degree of durational contrast between successive elements in a sequence, and was developed to explore rhythmic differences between "stress-timed" and "syllable-timed" languages (Low, 1998; Low *et al.*, 2000). Empirical work in phonetics has revealed that the nPVI of vowel durations in sentences is significantly higher in stress-timed languages such as British English than in syllable-timed languages such as French (Grabe and Low, 2002; Ramus, 2002). The reason for this is thought to be the greater degree of vowel reduction in the former languages (Dauer, 1983, 1987; Nespor, 1990). Patel and Daniele applied the nPVI to the durations of notes in instrumental classical themes from England and France, and found that English music had a significantly higher nPVI than French music. This earlier work illustrates a cross-cultural approach to comparing prosodic and musical structure which is extended in the current study. This approach is based on determining whether quantitative prosodic differences between languages are reflected in music (cf. Wenk, 1987).

While Patel and Daniele (2003) focused on composers from the turn of the 20th century (a time of musical nationalism), subsequent work showed that their finding generalized to a much broader historical sample of composers from England and France (Huron and Ollen, 2003). Furthermore, it appears that other European cultures with stress-timed languages tend to have higher musical nPVI values than cultures with syllable-time languages (Huron and Ollen, 2003), though interesting exceptions exist (Patel and Daniele, 2003b; Daniele and Patel, 2004). These studies indicate that prosody and instrumental music can be meaningfully compared using quantitative methods. They also raise two key

[a]Author to whom correspondence should be addressed. Tel: 858-626-2085; Fax: 858-626-2099; electronic mail: apatel@nsi.edu

 0001-4966/2006/119(5)/3034/14/$22.50

questions which are the focus of the current study, as detailed in Secs. I A 1 and I A 2 below. As in the previous work, the current study focuses on British English and continental French (henceforth English and French). One of the principal goals, however, is to address issues and develop methods of broad applicability to speech-music research.

1. Are differences in durational contrast a byproduct of variability differences?

Although the nPVI's name refers to a "variability index," it is in fact a measure of durational contrast. This is evident from the nPVI equation,

$$\mathrm{nPVI} = \frac{100}{m-1} \times \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{d_k + d_{k+1}}{2}} \right|, \quad (1)$$

where m is the number of durational elements in a sequence d_k is the duration of the kth element. The nPVI computes the absolute difference between each successive pair of durations in a sequence, normalized by the mean duration of the pair. This converts a sequence of m durations (e.g., vowel durations in a sentence) to a sequence of $m-1$ contrastiveness scores. Each of these scores ranges between 0 (when the two durations are identical) and 2 (for maximum durational contrast, i.e., when one of the durations approaches 0). The mean of these scores, multiplied by 100, yields the nPVI of the sequence. The nPVI is thus a contrastiveness index and is quite distinct from measures of overall variability (such as the standard deviation) which are insensitive to the order of observations in a sequence. Indeed, one cannot compute the nPVI of a given sequence from its standard deviation or vice versa. Nevertheless, at the population level differences in overall variability of two sets of sequences will inevitably drive some degree of nPVI difference between the sets, simply because sequences with greater variability are likely to contain neighbors of greater durational contrast (cf. Sadakata and Desain, submitted).

This point is relevant to the comparison of English and French because there are reasons to expect that vowels in English sentences should exhibit higher overall durational variability than vowels in French sentences. One such reason is that vowel duration in English is substantially modulated by stress and vowel reduction, factors which play much less of a role in modulating vowel duration in French (Delattre, 1966; 1969). Hence it is important to know if linguistic nPVI differences between the languages are simply a by-product of variability differences. A similar question applies to music. Should this be the case, then music may simply reflect differences in linguistic temporal variability, with the nPVI difference being a by-product of such differences.

To examine these issues, a measure of overall variability for each sentence and musical theme is computed in this study in order to examine the relationship between variability and nPVI. Specifically, a Monte Carlo method is used to quantify the likelihood of observing an nPVI difference of a given magnitude between two languages (or two musics) given existing differences in variability.

2. Is speech melody reflected in music?

The original intuition of a link between prosody and instrumental music was not confined to rhythm, but encompassed melody as well (e.g., Hall, 1953). The current study addresses this issue via a quantitative comparison of intonation and musical melody. Earlier comparative work on rhythm had the benefit of an empirical measure which could readily be applied to music (the nPVI). In the case of intonation, no such measure was available. To overcome this problem, this study employs a recent computational model of speech intonation perception known as the "prosogram" (Mertens, 2004a, 2004b). The prosogram converts a sentence's fundamental frequency (Fo) contour into a sequence of discrete tonal segments, producing a representation which is meant to capture the perceived pitch pattern of a speech melody. This representation allows a quantitative comparison of pitch variability in speech and music. Further details on the prosogram and measures of variability are given in the next section.

B. Background

1. Rhythm

Speech rhythm refers to the way languages are organized in time. Linguists have long held that certain languages (such as English and French) have decidedly different rhythms, though the physical basis of this difference has been hard to define. Early ideas that the difference lay in the unit produced isochronously (either stresses or syllables) have not been supported by empirical research (e.g., Roach, 1982, Dauer, 1983). Some linguists have nevertheless retained the "stress-timed" vs "syllable-timed" terminology, likely reflecting an intuition that languages placed in these categories do have salient rhythmic differences (Beckman, 1992). Examples of languages placed in these categories are British English, Dutch, and Thai (stress-timed) vs French, Spanish, and Singapore English (syllable timed) (Grabe and Low, 2002).

Recent years have seen the discovery of systematic temporal differences between stress-timed and syllable-timed languages (e.g. Ramus *et al.*, 1999; Low *et al.*, 2000). These discoveries illustrate the fact that "rhythm" in speech should not be equated with isochrony. That is, while isochrony defines one kind of rhythm, the absence of isochrony is not the same as the absence of rhythm. Rhythm is the systematic temporal and accentual patterning of sound. Languages can have rhythmic differences which have nothing to do with isochrony. For example, Ramus and co-workers (1999) demonstrated that sentences in stress-timed (vs syllable-timed) languages had greater durational variability in "consonantal intervals" (consonants or sequences of abutting consonants, regardless of syllable or word boundaries) and a lower overall percentage of sentence duration devoted to vowels. (They termed these two measures ΔC and $\%V$, respectively). These differences likely reflect phonological factors such as the greater variety of syllable types and the greater degree of vowel reduction in stress-timed languages (Dauer, 1983). Another difference between stress-timed and syllable-timed languages, noted previously, is that the durational contrast

between adjacent vowels in sentences is higher in the former types of languages (as measured by the nPVI), probably also due to the greater degree of vowel reduction in these languages (Low *et al.*, 2000; Grabe and Low, 2002; Ramus, 2002).[1] An interesting point raised by these recent empirical studies is that languages may fall along a rhythmic continuum rather than forming discrete rhythm classes (cf. Nespor, 1990; Grabe and Low, 2002). This "category vs continuum" debate in speech rhythm research has yet to be resolved, and is largely orthogonal to the issues addressed here.

Of the different temporal measures described above, vowel-based measures of rhythm are the most easily transferred to music research. This is because musical notes can roughly be compared to syllables, and vowels form the core of syllables. It therefore seems plausible to compare vowel-based rhythmic measures of speech to note-based rhythmic measures of music. Of the two vowel-based measures discussed above (*%V* and nPVI), the latter can be sensibly applied to music by measuring the durational contrast between successive notes in a sequence. This approach is taken in the current work.

Since the focus here is on English and French, it is worth asking about the robustness of the nPVI difference between these two languages. Significant nPVI differences between English and French have been reported by four published studies, one based on vowel durations in spontaneous speech (Grabe *et al.*, 1999), and three based on vocalic interval durations in read speech (Grabe and Low, 2002; Ramus, 2002; Lee and Todd, 2004). (A vocalic interval is defined as the temporal interval between a vowel onset and the onset of the next consonant in the sentence; a vocalic interval may thus contain more than one vowel and can span a syllable or word boundary, cf. Ramus *et al.*, 1999; Grabe and Low, 2002. The choice of vowels vs vocalic intervals makes little difference when comparing the nPVI of English and French, cf. Secs. II A and II B.) One notable finding is that nPVI values for English and French vary considerably from study to study. For example, Ramus (2002) reported values of 67.0 and 49.3 for English and French, respectively, while Lee and Todd (2004) reported values of 83.9 and 54.3. Both studies measured vocalic intervals in read speech. Possible sources of this discrepancy include differences in speech materials, speech rate, and the criteria for the placement of boundaries between vowels and consonants (cf. Dellwo *et al.*, 2006). Further research is needed to clarify this issue. What one can say confidently, however, is that the finding that English has a significantly higher nPVI than French (within a given study) appears highly robust.[2]

2. Melody

A central issue for comparing melody in speech and music is how to represent speech melodies. Several choices exist. One choice is to use the raw Fo contours of sentences. Another choice is to use sequences of abstract phonological tones, such as high (*H*) and low (*L*) tones in autosegmental-metrical theories of intonation (e.g., Pierrehumbert, 1980; Ladd 1996). This study opts for a representation that is neither as detailed as raw Fo contours nor as abstract as autosegmental-metrical approaches. This is the "prosogram" representation of intonation (Mertens, 2004a, 2004b; cf. d'Alessandro and Mertens, 1995).

The prosogram aims to provide a representation of intonation as perceived by human listeners, and thus follows in the tradition of Fo stylization based on perceptual principles (Rossi, 1971; Rossi, 1978a; Rossi, 1978b; 'tHart *et al.*, 1990). It is based on empirical research which suggests that pitch perception in speech is subject to four perceptual transformations. The first is the segregation of the Fo contour into syllable-sized units due to the rapid spectral and amplitude fluctuations in the speech signal (House, 1990). The second is a threshold for the detection of pitch movement within a syllable (the "glissando threshold"). The third, which applies when pitch movement is detected, is a threshold for detection of a change in the slope of a pitch movement within a syllable (the "differential glissando threshold"). The fourth, which applies when the pitch movement is subliminal, is temporal integration of Fo within a syllable (d'Alessandro and Castellengo, 1994; d'Alessandro and Mertens, 1995). The prosogram instantiates the latter three transformations via an algorithm which operates on the vocalic nuclei of syllables (phonetic segmentation is provided by the user). As a result of these transforms, the original Fo contour of a sentence is converted to a sequence of discrete tonal segments. An example of the model's output is given in Fig. 1, which shows the original Fo contour (top) and the prosogram (bottom) for the English sentence "Having a big car is not something I would recommend in this city." Figure 1 reveals why the prosogram is useful to those interested in comparing speech and music. The representation produced by the prosogram is quite musiclike, consisting mostly of level pitches. (Some syllables are assigned pitch glides if the amount of Fo change is large enough to exceed a perceptual threshold.) On a cognitive level, this is interesting because it implies that the auditory image of speech intonation in a listener's brain has more in common with music than has been generally appreciated. On a practical level, the dominance of level pitches means that intonation patterns in different languages can be compared using tools that can also be applied to music, e.g., statistical measurements of pitch height or pitch interval patterns.[3]

The current study uses the prosogram to examine a simple aspect of the statistical patterning of spoken intonation, namely pitch variability. Specifically, pitch variability in English and French speech is quantified from prosograms and compared to pitch variability in English and French musical themes. Prior studies of pitch variability in the two languages have produced contradictory results. Maidment (1976) computed running Fo from a laryngograph while speakers ($n=16$) read a $2\frac{1}{2}$ min passage of prose. He reported the mean and standard deviation of the Fo contours produced by each speaker. When converted into the coefficient of variation (standard deviation/mean), a significantly higher degree of variability is found for the English than for the French speakers. In contrast, Lee and Todd (2004) reported no significant difference in Fo variability in English and French speech. However, rather than measure raw Fo contours they extracted the Fo at the onset of each vocalic

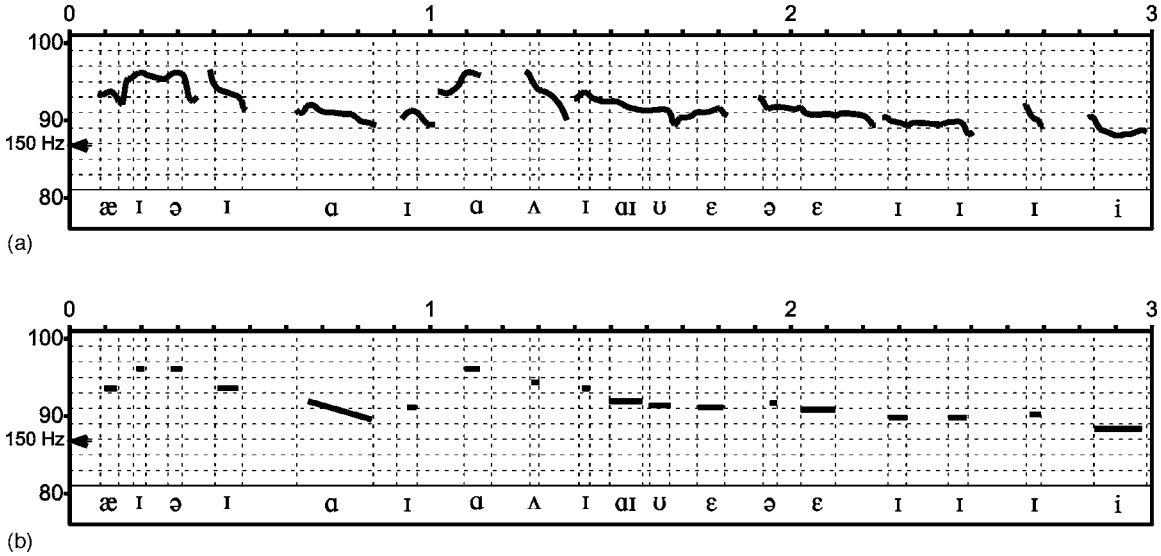


FIG. 1. Illustration of the prosogram, using the British English Sentence "Having a big car is not something I would recommend in this city" as uttered by a female speaker. In both graphs, the horizontal axis along the top shows time in seconds, the vertical axis shows semitones re 1 Hz (an arrow is placed at 150 Hz for reference), and the bottom shows IPA symbols for the vowels in this sentence. The onset and offset of each vowel is indicated by vertical dashed lines above the vowels' IPA symbol. (a) Shows the original Fo contour, while (b) shows the prosogram. In this case, the prosogram has assigned level tones to all vowels save for the vowel in "car," which was assigned a glide. Note that the pitches of the prosogram do not conform to any musical scale.

interval in a sentence and studied the variability of these values across a sentence. (They expressed each value as a semitone distance from the mean vocalic-onset Fo in the utterance, and then computed the standard deviation of this sequence of pitch values.)

These two studies illustrate the fact that pitch variability in speech can be measured in different ways. Whether or not one obtains differences between languages may depend on the method one chooses. For those interested in perception, the prosogram offers a motivated way to study the variability of pitch patterns in speech. Furthermore, it offers two ways to quantify variability, i.e., in terms of pitch height and pitch intervals. The former measures the spread of pitches about a mean pitch (as in Lee and Todd's, 2004 study). The latter measures whether steps between successive pitches tend to be more uniform or more variable in size. Both types of measures were computed in this study in order to compare speech to music.

II. METHODS

A. Corpora

The materials were the same as those in Patel and Daniele (2003). For speech, 20 English and 20 French sentences were taken from the database of Nazzi *et al.* (1998), consisting of four female speakers per language reading five unique sentences each. The sentences had been recorded in a quiet room and digitized at 16 000 Hz. They are short newslike utterances, and have been used in a number of studies of speech rhythm by other researchers (e.g., Nazzi *et al.*, 1998; Ramus *et al.*, 1999; Ramus, 2002). They range from 15 to 20 syllables in length and are approximately 3 s long (see Appendix A for a full list). Table I gives some basic data on sentence characteristics. Sentences contained about 16 vowels on average, most of which were singletons (i.e., a vowel not abutting another vowel). Thus durational computations based on vowels vs vocalic intervals are likely to yield similar results. The original motivation for studying vocalic intervals in studies of speech rhythm was an interest in infant speech perception, under the assumption that infants perform a crude segmentation of the speech stream which only distinguishes between vocalic and nonvocalic (i.e., consonantal) portions (Mehler *et al.*, 1996; Ramus *et al.*, 1999). Since the current work focuses on adult perception of speech, and since vowels are well-established phonological units in language while vocalic intervals are not, this study examines vowels rather than vocalic intervals.

The musical data are themes from turn-of-the 20th century English and French composers, drawn from a musicological sourcebook for instrumental music (*A Dictionary of Musical Themes*, Barlow and Morgenstern, 1983). Themes were analyzed for all English and French composers in the

TABLE I. Some basic statistics on the sentences studied.

	Duration (s) Mean (sd)	Speech rate (syll/s) Mean (sd)	No. Vowels / sentence Mean (sd)	Avg Fo (Hz) Mean (sd)	Total vowels	Singleton vowels
English (n=20)	2.8 (0.2)	5.8 (0.3)	15.7 (1.7)	222.2 (14.1)	314	296
French (n=20)	2.8 (0.2)	6.1 (0.5)	17.3 (1.6)	219.4 (25.7)	346	310

dictionary who were born in the 1800s and died in the 1900s. This era is recognized by musicologists as being a time of musical nationalism, when music is thought to be especially reflective of culture (Grout and Palisca, 2000). It is also not too distant in the past, which is desirable since measurements of speech are based on living speakers and since languages can change phonologically over time. To be included, composers had to have at least five usable themes, i.e., themes that met a number of criteria designed to minimize the influence of language or other external influences on musical structure. For example, themes were excluded if they came from works whose titles suggested a vocal conception or the purposeful evocation of another culture (see Patel and Daniele, 2003 for the full list of criteria). Furthermore, themes were required to have at least 12 notes (to provide a good sample for rhythm measures), and no internal rests, grace notes, or fermatas, which introduced durational uncertainties. These criteria yielded six English composers (137 themes) and ten French composers (181 themes). In reviewing the themes used in the previous study, a few inadvertent errors of inclusion or exclusion were found and corrected, resulting in 136 English and 180 French themes in the current work (see Appendix B for a complete list of composers and themes).

B. Phonetic segmentation

To analyze linguistic nPVI, vowel boundaries were marked in English and French sentences using wide-band speech spectrograms generated with SIGNAL (Engineering Design) running on a modified personal computer (frequency resolution=125 Hz, time resolution=8 ms, one FFT every 3 ms, Hanning window). Vowel onset and offset were defined using standard criteria (Peterson and Lehiste, 1960). Vowel boundaries in this study were marked independently of the boundaries defined by Ramus (2002) for the same set of sentences. Those earlier boundaries, which served as the basis of the nPVI values reported in Patel and Daniele (2003), came from a phonetic segmentation based on a waveform display with interactive playback. In the current study segmentation of the sentences was based on a display showing both the waveform and spectrogram, plus interactive playback. Availability of a spectrogram often resulted in boundary locations which differed from those marked by Ramus.

As noted above, this study focuses on vowels rather than vocalic intervals. However, both quantities were measured and yielded very similar results in the rhythmic analyses. This is not surprising since the great majority of vowels in both languages were singletons (Table I). The results report vowel measurements only (data based on vocalic intervals are available upon request).

C. Duration coding of musical themes

As in earlier work, measurement of duration in musical themes was made directly from music notation. This stands in contrast to the speech measurements, which were (necessarily) based on acoustic signals. Initially this may seem problematic, but as noted in Patel and Daniele (2003) the analysis of acoustic recordings of music raises its own problems, such as which performance of each theme to analyze and how to defend this performance against other available recordings, all of which will differ in the fine nuances of timing. Music notation is thus a reasonable choice since it at least affords an unambiguous record of the composer's intentions. That being said, it is important for future work to study timing patterns in human performances because such performances deviate from the idealized durations of music notation in systematic ways (Repp, 1992; Palmer, 1997). It will be interesting to determine what influence such deviations have on the results of a study like this one.

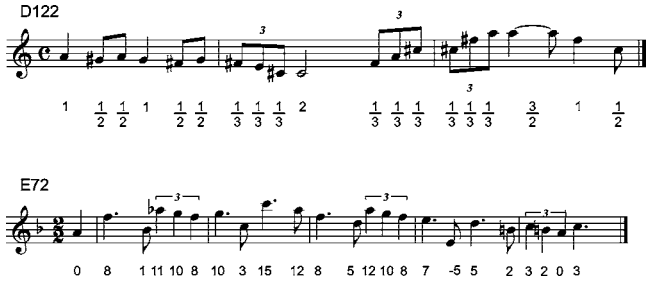


FIG. 2. Examples of duration and pitch coding of musical themes. (D122: Debussy's *Quartet in G minor for Strings, 1st movement, 2nd theme.* E72: Elgar's *Symphony No. 1, in A Flat, Opus 55, 4th movement, 2nd theme.*) D122 illustrates duration coding: the relative duration of each note is shown below the musical staff (see text for details). E72 illustrates pitch coding: each note is assigned a pitch value based on its semitone distance from A4 (440 Hz). The nPVI of note durations in D122 is 42.2. The coefficient of variation (CV) of pitch intervals in E72 is 0.79.

An example of durational coding of a musical theme is shown in Fig. 2 (top). Notes were assigned durations according to the time signature, with the basic beat assigned a duration of 1 (e.g. a quarter note in 4/4 or an eighth note in 3/8), and other notes assigned their relative durations according to standard music notation conventions. Durations were thus quantified in fractions of a beat. Because the nPVI is a relative measure, any scheme which preserves the relative duration of notes will result in the same nPVI for a given sequence. For example, the first note of a theme could always be assigned a duration of 1 and other note durations could be expressed as a fraction or multiple of this value.

D. Computation of rhythmic measures and Monte Carlo analyses

Two rhythmic measures were computed for each sentence and musical theme: the nPVI and the coefficient of variation (CV), the latter defined as the standard deviation divided by the mean. Like the standard deviation, the CV is a measure of overall variability which is insensitive to the order of elements in the sequence. Yet like the nPVI, it is a dimensionless quantity since the same units appear in both the numerator and denominator. It is thus well suited for comparing temporal patterns in speech and music when speech is measured in seconds and music is measured in fractions of a beat. Furthermore, the CV (like the nPVI) is only sensitive to the relative duration of events; scaling these durations up or down by a constant factor does not change its

value. Hence any durational coding scheme for musical themes will produce the same CV as long as the relative durations of events are preserved.

The relationship between nPVI and CV was studied in two ways. First, within each domain (speech or music) linear regressions were performed with CV as the independent variable and nPVI as the dependent variable. This showed the extent to which nPVI was predicted by CV. Next, within each domain a Monte Carlo technique was used to estimate the probability of the observed English-French nPVI difference given existing variability differences. The technique was based on scrambling the order of durations in each sequence, which destroys its temporal structure while retaining its overall variability. For example, if focusing on speech, the sequence of vowel durations in each English and French sentence was randomly scrambled and the nPVI of these scrambled sequences was computed. The difference between the mean nPVI values for the scrambled-English and scrambled-French sentences was recorded. This procedure was repeated 1000 times. The proportion of times that this "scrambled nPVI" difference was equal to or greater than the original nPVI difference was taken as a *p*-value indicating the probability of obtaining an nPVI equal to or greater than the observed nPVI difference between the languages. The same procedure was used for musical themes.

E. Melodic analyses

To quantify pitch patterns in speech, prosograms were computed for all English and French sentences using prosogram version 1.3.6 as instantiated in Praat (Mertens, 2004a; 2004b).[4] As part of the algorithm, an Fo contour was computed for each sentence using the autocorrelation algorithm of Boersma (1993). (Default parameters were used with the exception of frame rate, minimum pitch, and maximum pitch, which were set to 200 Hz, 60 Hz, and 450 Hz, respectively.) Prosogram analysis is based on the vocalic nuclei of syllables. To determine whether a given vowel should be assigned a level tone or a glide, a glide threshold of $0.32/T^2$ semitones/s was used, where T is the duration of a vowel in s. If the rate of pitch change was greater than this threshold, the vowel was assigned a frequency glide (or two abutting glides if the differential glissando threshold was exceeded). The choice of $0.32/T^2$ semitones/s as the glissando threshold is based on perceptual research on the threshold for detecting pitch movement in speech, combined with experiments in which prosogram output is compared to human transcriptions of intonation ('tHart 1976; Mertens, 2004b). Vowels with pitch change below the glide threshold were assigned a level tone equal to their median pitch value. This served as an estimate of the perceived pitch of the syllable, as formerly computed from a time-weighted average of the vowel's Fo contour in earlier versions of the prosogram (e.g., d'Alessandro and Mertens, 1995; Mertens and d'Alessandro, 1995; Mertens *et al.*, 1997).

For maximum comparability to music, only level tones were used in the quantification of pitch variability in speech. Such tones represented 97% of tones assigned to vowels in the current corpus. Occasionally the prosogram did not assign a tonal element to a vowel, e.g., if the intensity of the vowel was too low, the vowel was devoiced (e.g., an unstressed "to" in English being pronounced as an aspirated /t/), or if Praat produced a clearly erroneous Fo value. However, such omissions were rare: 90% of English vowels were assigned level tones and 4% were assigned glides, while 96% of French vowels were assigned level tones and 2% were assigned glides. A typical sentence had about 15 level tones and 1 glide.

Variability in pitch height and of pitch intervals within a sentence was computed via the coefficient of variation (CV). To study pitch height variation, each level tone was assigned a semitone distance from the mean pitch of all level-tones in the sentence, and then the CV of these pitch distances was computed. To study interval variability, adjacent level tones in a sentence were assigned a pitch interval in semitones,

$$\mathrm{st} = 12 \log_2(f2/f1), \tag{2}$$

where $f1$ and $f2$ represent the initial and final tone of the pair, respectively. (Intervals were computed between immediately adjacent level tones only, not between tones separated by a glide.) The CV of these intervals was then quantified. Because the mean appears in the denominator of the CV, measurements of pitch distances and pitch intervals used absolute values in order to avoid cases where mean distance size or mean interval size was equal to or near 0, which would yield mathematically ill-defined CVs. The choice of semitones as units of measurement is based on recent research on the perceptual scaling of intonation (Nolan, 2003). Earlier work by Hermes and Van Gestel (1991) had suggested ERBs should be used in measuring pitch distances in speech. Since the CV is dimensionless, one could measure pitch in speech and music in different units (ERBs vs semitones) and still compare pitch variability across domains using the CV as a common metric. The precise choice of units for speech is unlikely to influence the results reported here.

To quantify melodic patterns in music, musical themes were coded as sequences of pitch values where each value represented a given pitch's semitone distance from A440 (Fig. 2, bottom). This permitted straightforward computation of each tone's semitone distance from the mean pitch of the sequence and of pitch interval patterns (the latter simply being the first-order difference of the pitch values). Measures of pitch height and interval variability were then computed in precisely the same manner as for speech. (Note that the choice of A440 as a referent pitch makes no difference to the measures of variability computed here. Any scheme which preserves the relative position of tones along a semitone scale would yield the same results. For example, one could code each pitch in a musical theme as its distance in semitones from the lowest pitch of the theme, or from the mean pitch of the theme.)

III. RESULTS

A. Rhythm

Table II shows nPVI and CV values for speech and music. Reported *p*-values in this and following tables were computed using the Mann-Whitney U-test, except for *p*-values

TABLE II. nPVI and CV for speech and music (mean and s.e.). The rightmost column gives the probability that the observed nPVI difference is due to the difference in CV.

	English nPVI	French nPVI	p	English CV	French CV	p	$p(\Delta\text{nPVI}\vert\Delta\text{CV})$
Speech (vowels)	55.0 (3.0)	35.9 (1.8)	<0.001	0.55 (0.03)	0.36 (0.02)	<0.001	0.01
Music (notes)	47.1 (1.8)	40.2 (1.9)	<0.01	0.61 (0.02)	0.58 (0.02)	0.34	0.001

associated with Monte Carlo analyses, which were computed as described in Sec. II D. Table II reveals that English and French sentences show a highly significant difference in durational contrastiveness (nPVI) as well as in durational variability (CV). English and French music, on the other hand, show a significant difference in contrastiveness but not in variability.

Regressions of CV on nPVI in each domain are shown in Fig. 3. The linear regressions reveal that within each domain, higher CV is predictive of higher nPVI, though the relationship appears to be stronger in speech than in music (see Fig. 3 captions for regression slopes, r^2 values, and p values).

To assess whether CV differences between the two languages are responsible for linguistic nPVI differences, a Monte Carlo analysis was conducted as described in Sec. II D. The result of this analysis is shown in Fig. 4, which plots the distribution of nPVI differences between English and French speech when the order of vowel durations in each sentence is scrambled and the nPVI difference between the two languages is computed (1000 iterations). The actual nPVI difference (19.1 points) is shown by an arrow. Based on this frequency distribution, the probability of an nPVI difference of 19.1 points or greater is quite small ($p=0.01$). This value is listed in the right-most column of Table II as $p(\Delta\text{nPVI}\vert\Delta\text{CV})$, i.e., the probability of the observed difference in nPVI given the observed difference in variability. A similar analysis was conducted for music, with the resulting p-value being 0.001. Thus it is highly unlikely that variability differences account for nPVI differences in either domain.

B. Melody

Table III shows the results of pitch variability measurements for speech and music. Table III reveals that English and French speech do not differ in the variability of pitch height within utterances, but do show a significant difference in pitch interval size variability, with French having lower

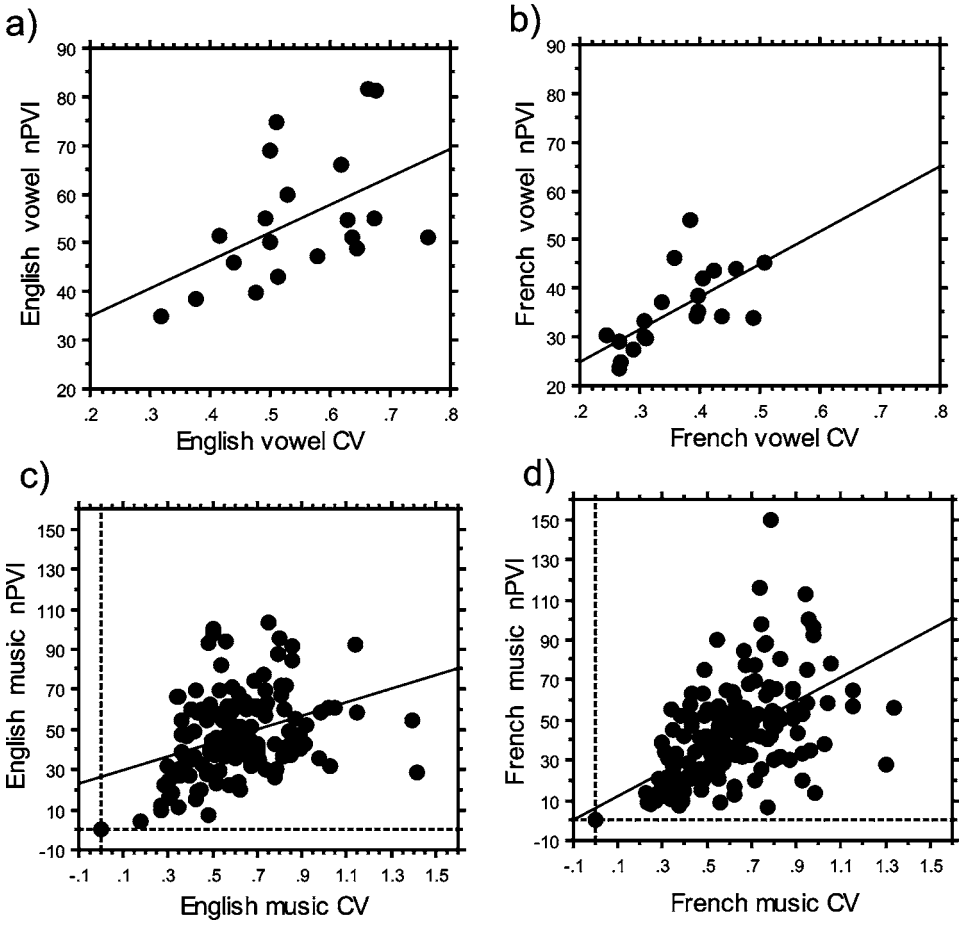


FIG. 3. The relationship between CV and nPVI for speech (a, b) and music (c, d). For speech each dot represents one sentence; for music each dot represents one theme. The best fitting regression line for each panel is also shown. English speech: nPVI=23.5 +57.3×CV, $\mathbf{r^2=0.34}$, df=18, $\mathbf{p=0.03}$; French speech: nPVI=11.7 +66.5×CV, $\mathbf{r^2=0.43}$, df=18, $\mathbf{p<0.01}$; English music: nPVI=26.6 +33.9×CV, $\mathbf{r^2=0.14}$, df=134, $\mathbf{p<0.001}$; French music: nPVI=6.2 +59.1×CV, $\mathbf{r^2=0.36}$, df=178, $\mathbf{p<0.001}$. For the musical data, hatched lines show the lower possible limit of the nPVI and CV at 0 on each axis: the axes range into negative numbers for display purposes only, so that the points at (0,0) can be clearly seen. Themes with a score of 0 for nPVI and CV have notes of a single duration. There were two such English themes and eight such French themes.

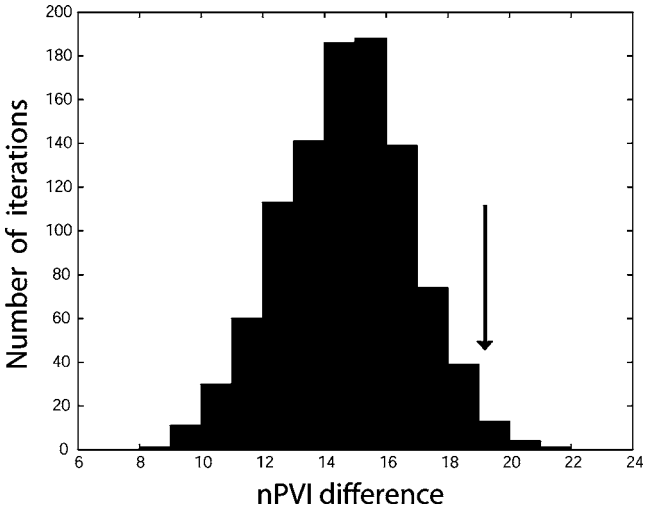


FIG. 4. Result of Monte Carlo analysis for English vs French speech. The actual nPVI difference between the two languages in this study (19.1 points) is shown by an arrow. See text for details.

interval variability than English. With regard to the latter point, it is interesting to note that the average absolute interval size is virtually identical in the two languages (2.1 vs 2.2 st for French vs English, respectively), yet French speech shows significantly lower interval variability than English. In other words, as the voice moves from one vowel to the next the size of the pitch change is more uniform in French than in English speech.

Turning to music, a similar picture emerges: differences in pitch height variability are not significant, but French music has significantly lower pitch interval variability than English music. In other words, as the melody moves from one note to the next the size of the pitch change is more uniform in French than in English music. Once again, this difference exists despite the fact that the average absolute interval size is nearly identical (2.7 vs 2.6 st for French and English music, respectively). Figure 5 shows the data for pitch interval variability in speech and music. Figure 5 shows that the linguistic difference in pitch interval variability between English and French speech is much more pronounced than is the musical difference, a finding reminiscent of earlier work

TABLE III. Pitch variability in speech and music, measured in terms of pitch height (the CV of pitch distances from the mean pitch of a sequence) or pitch intervals (the CV of pitch interval size within a sequence). Mean and s.e. are shown.

	English Pitch CV	French Pitch CV	p
Speech			
Pitch height	0.71 (0.04)	0.75 (0.04)	0.32
Pitch intervals	0.88 (0.05)	0.68 (0.03)	<0.01
Music			
Pitch height	0.69 (0.01)	0.71 (0.01)	0.14
Pitch intervals	0.76 (0.02)	0.71 (0.02)	0.03

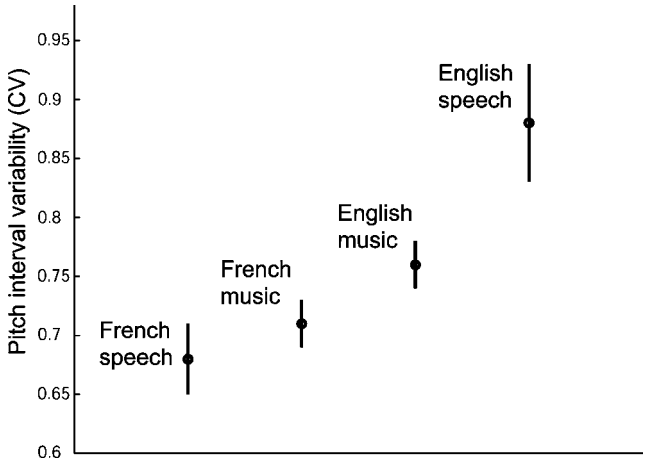


FIG. 5. Pitch interval variability in English and French speech and music. Pitch interval variability is defined as the CV of absolute interval size between pitches in a sequence. Error bars show standard errors.

on the nPVI. This should not be surprising, since music is an artistic endeavor with substantial intracultural variation and (unlike speech) no *a priori* reason to follow melodic norms. What is remarkable is that despite this variation, a significant cultural difference emerges in the same direction as the linguistic difference.

C. Rhythm and melody combined

Having analyzed rhythm and melody independently, it is interesting to combine the results in a single graph showing English and French speech and music in two-dimensional space with rhythm and melody on orthogonal axes. This representation can be referred to as an RM-space plot. Figure 6 shows such a plot, with nPVI being the measure of rhythm and melodic interval variability (MIV) being the measure of melody. MIV is defined as $100 \times$CV of pitch interval variability. Scaling the CV by 100 serves to put MIV in the same general range of values as nPVI. One aspect of this figure deserves immediate comment. Recall from Sec. I B 1 that the value of a given language's nPVI can differ widely from one study to the next, likely due to differences in corpora, speech

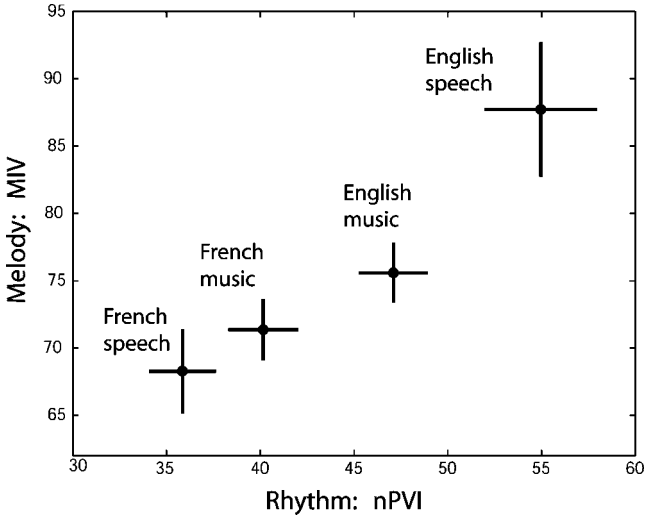


FIG. 6. Rhythm-melody (RM) space for speech and music. Axes are nPVI and MIV. Error bars show standard errors.

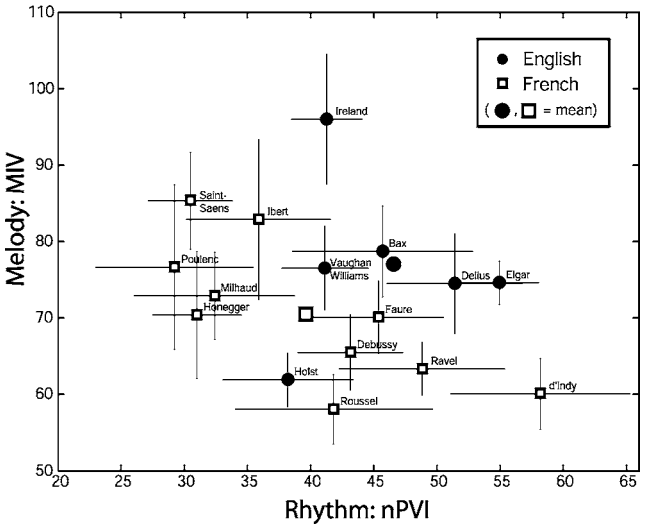


FIG. 7. nPVI and MIV values for individual composers. Error bars show standard errors. Note the almost complete separation of English and French composers in RM space, despite large overlap between the nationalities along either single dimension.

rate, and phonetic segmentation criteria. The same may be true of MIV values. Thus the position of a given language in RM space will likely vary from study to study. What is relevant is the *distance* between languages *within* a given study, where corpora and other criteria are more tightly controlled. The same point applies to music.

Using RM space, the "prosodic distance" (pd) between two languages can be defined as the Euclidean distance between the points representing the mean nPVI and MIV of the languages. For English and French speech (Es, Fs), this distance is

$$\mathrm{pd(Es,Fs)} = \sqrt{(\mathrm{nPVI_{Es}} - \mathrm{nPVI_{Fs}})^2 + (\mathrm{MIV_{Es}} - \mathrm{MIV_{Fs}})^2}. \quad (3)$$

In the current data the distance between English and French speech is 27.7 RM units. Applying the same equation to the musical data (i.e., replacing Es and Fs with Em and Fm), yields a distance between English and French music of 8.5 RM units. Thus the musical distance is about 30% of the linguistic difference.

Another aspect of Fig. 6 worth noting is that a line connecting English and French speech in RM space would lie at a very similar angle to a line connecting English and French music. In fact, if one defines vectors between the two languages and the two musics, using standard trigonometric formulas one can show that the angle between these vectors is only 14.2°. Thus a move from French to English speech in RM space involves going in a very similar direction as a move from French to English music.

Focusing now on the musical data, it is interesting to examine the position of individual composers in RM space, as shown in Fig. 7. Figure 7 reveals that English and French composers occupy distinct regions of RM space, despite large variation along any single dimension (Holst is the one exception, and is discussed later). This suggests that the *joint* properties of melody and rhythm, not either one alone, are involved in defining national characteristics of music.

IV. DISCUSSION

A. Aspects of speech prosody reflected in music

New tools from phonetics permit the confirmation of an old intuition shared by musicologists and linguists, namely that the instrumental music of a culture can reflect the prosody of its native language. These tools (the nPVI and the prosogram) are noteworthy because they allow quantitative comparisons of rhythm and melody in speech and music.

An exploration of the relationship between durational contrastiveness (nPVI) and durational variability reveals that nPVI differences between English and French speech are not a by-product of variability differences, even though the two languages do have significant differences in the variability of vowel durations in sentences. Variability differences also cannot account for musical nPVI differences between English and French (in fact, the two musics do not show a significant variability difference). It is thus clear that music reflects durational contrastiveness in speech, not durational variability. This finding is interesting in light of claims by linguists that part of the characteristic rhythm of English is a tendency for full and reduced vowels to alternate in spoken sentences (e.g., Bolinger, 1981). It appears that this tendency (or its absence) may be reflected in music.

Turning to melody, measures of pitch variability reveal that a specific aspect of speech melody is reflected in music, namely the variability of interval size between successive pitches in an utterance. English speech shows greater interval variability than French speech, even though the average pitch interval size in the two languages is nearly identical. This same pattern is reflected in music. Initially it may seem odd that pitch *intervals* in speech are reflected in music. While the human perceptual system is quite sensitive to interval patterns in music (where for example a melody can be recognized in transposition as long as its interval pattern is preserved), music features well-defined interval categories such as the minor second and perfect fifth while speech does not. Might the perceptual system attend to interval patterns in speech despite the lack of stable interval structure? Recent theoretical and empirical work in intonational phonology suggests that spoken pitch intervals may in fact be important in the mental representation of intonation, even if they do not adhere to fixed frequency ratios (Dilley, 2005). If this is the case, then one can understand why the perceptual system might attend to interval patterns in speech as part of learning the native language. It remains to be explained, though, why English speech should have a greater degree of interval variability than French. One idea is that British English may have three phonologically distinct pitch levels in its intonation system, while French may only have two (cf. Willems, 1982; Ladd and Morton, 1997; Jun and Fougeron, 2000; Jun, 2005). A compelling explanation, however, awaits future research.

While the current study focuses on just two cultures, it is worth noting that the techniques presented here are quite general. They can be applied to any culture where a sufficient sample of spoken and musical patterns can be collected. It would be quite interesting, for example, to use them to study relations between speech and instrumental music in

tone languages. Thai and Mandarin may be a good choice for such a comparison, since the languages have very different nPVI values (Grabe and Low, 2002) and different tone patterns, and since both languages come from cultures with well-developed instrumental music traditions.

B. A possible mechanism

By what route might speech patterns find their way into music? One oft-heard proposal is that composers borrow tunes from folk music, and that these tunes bear the stamp of linguistic prosody because they were written with words. This might be termed the "indirect route" from speech to music. The current study proposes a different hypothesis based on the idea of a "direct route" between the two domains. One advantage of the direct-route hypothesis is that it can account for the reflection of speech in music not thought to be particularly influenced by folk music (e.g., much of Elgar's and Debussy's work, cf. Grout and Palisca, 2000). The direct-route hypothesis centers on the notion of statistical learning of prosodic patterns in the native language. Statistical learning refers to tracking patterns in the environment and acquiring implicit knowledge of their statistical properties, without any direct feedback. Statistical learning of prosodic patterns in one's native language likely begins early in development. Research on language development has shown that infants are adept at statistical learning of phonetic/syllabic patterns in speech and of pitch patterns in nonlinguistic tone sequences (Saffran *et al.*, 1996; 1999) Thus it seems plausible that statistical learning of rhythmic and tonal patterns in speech would also begin in infancy, especially since infants are known to be quite sensitive to the prosodic patterns of language (Jusczyk, 1997; Nazzi *et al.*, 1998; Ramus 2002b).

Statistical learning of tone patterns need not be confined to infancy. Adult listeners show sensitivity to the distribution of different pitches and to interval patterns (Oram and Cuddy, 1995; Saffran *et al.*, 1999; Krumhansl, 2000; Krumhansl *et al.*, 2000). Importantly, statistical learning in music can occur with atonal or culturally unfamiliar materials, meaning that it is not confined to tone patterns that follow familiar musical conventions. Synthesizing these findings with the comments in the preceding paragraph leads to the hypothesis that statistical learning of the prosodic patterns of speech creates implicit knowledge of rhythmic and melodic patterns in language, which can in turn influence the creation of rhythmic and tonal patterns in music. Importantly, there is no claim that this influence is obligatory. Rather, linguistic patterns are seen as one resource available to a composer (either consciously or subconsciously) when they set out to compose music that is "of their culture" (cf. Patel and Daniele, 2003). Future work should consider what sort of evidence would support a causal link of this kind between speech and music.

C. Rhythm-melody relations: A domain for future investigation

Figure 7 suggested that joint properties of melody and rhythm may be particularly important for distinguishing between the music of different nations. While this study has examined rhythm and melody independently, it would be worth examining relations between rhythmic and melodic patterns in future work. Such relations may distinguish between languages and be reflected in music. One motivation for pursuing this issue is research in music perception suggesting that listeners are sensitive to such relations, e.g., the temporal alignment (or misalignment) of peaks in pitch and in duration (Jones, 1987; 1993). A second motivation is research on prosody which has revealed language-specific relations in the form of stable patterns of alignment of pitch peaks and valleys relative to the segmental string (Arvaniti *et al.*, 1998; Ladd *et al.*, 1999; Atterer and Ladd, 2004). This suggests that part of the characteristic "music" of a language is the temporal alignment between rhythmic and melodic patterns.

The key practical issue for future studies of rhythm-melody relations is how to represent pitch and temporal patterns in a manner that facilitates cross-domain comparison. Figure 8 illustrates one idea based on a vowel-based analysis of pitch and timing in speech. Panel (a) shows the pitch pattern of a French sentence ("Les mères sortent de plus en plus rapidement de la maternité") as a sequence of vowel

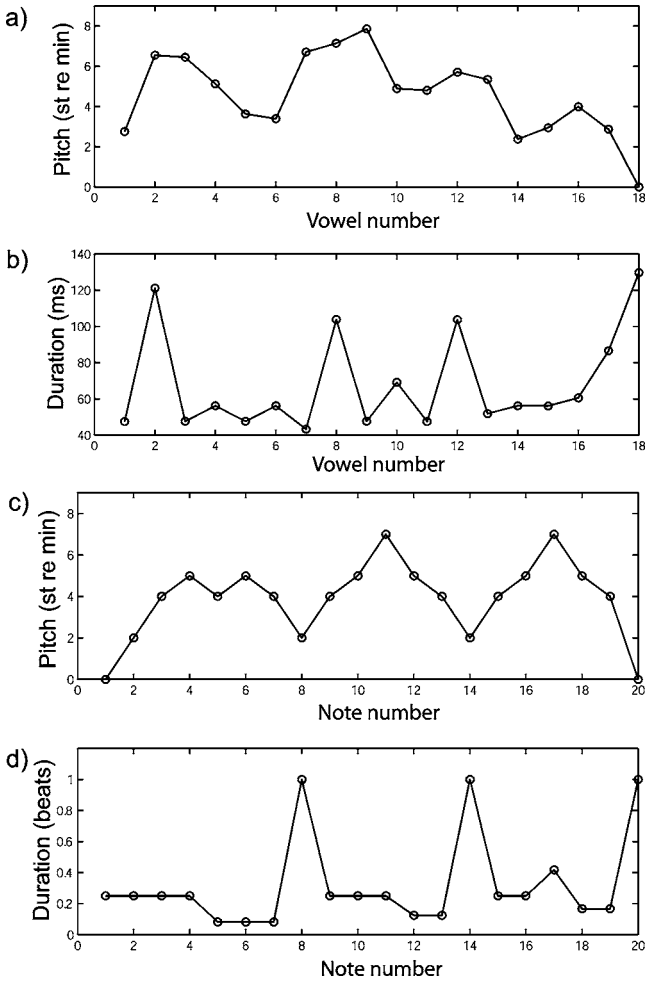


FIG. 8. (a, b) Pitch and duration patterns for the French sentence "Les mères sortent de plus en plus rapidement de la maternité" as uttered by a female speaker. (c, d) Pitch and duration patterns for a French musical theme from Debussy's *Les Parfums de la Nuit*. See text for details.

pitches, with each number representing the semitone distance of each vowel from the lowest vowel pitch in the sequence. Panel (b) shows the corresponding sequence of vowel durations (in ms). Panel (c) shows the pitch pattern of a French musical theme (4th theme of the 2nd movement of *Les Parfums de la Nuit*, theme D57 in Barlow and Morgenstern, 1983), using the same convention as panel (a), while panel (d) shows the duration pattern of the musical theme in fractions of a beat. The pitches in panel (a) are taken from the prosogram of the French sentence, but a reasonable estimate of these pitches can be made by simply taking the median Fo of each vowel (cf. Sec. IV D below).

When viewed in this way, it is easy to see how equivalent measurements of rhythm-melody relations can be made in speech and music. For example, one can ask whether there is a different pattern of alignment between pitch and duration peaks in language A than in language B, and if this difference is reflected in music. (This particular example is of interest to comparisons of British English and French, since phonologists have long noted that French sentences tend to be organized into phrases with pitch peaks and durational lengthening on the last syllable of each nonfinal phrase, e.g., Delattre, 1963; Jun and Fougeron, 2000; Di Cristo, 1998). Of course, one need not always look from speech to music; the direction of analysis can be reversed. In this case, one would first examine music for culturally-distinctive rhythm-melody relations, and then ask if these relations are reflected in speech. For example, the British composer Gustav Holst is unusual in being located in the "French" region of RM space in Fig. 6. Yet intuitively there is nothing French sounding about Holst's music (Indeed, the tune of "I Vow to Thee my Country," based on a theme from the *Jupiter* movement of *The Planets*, has often been suggested as an alternative national anthem for Britain). While Holst's anomalous position in RM space could simply be a sampling issue (i.e., it could change if more themes were added) it could also be that some as-yet unidentified relationship between rhythm and melody identifies his music as distinctively English. If so, it would be quite interesting to use this relationship to define a third dimension in RM space to see if it also separates English and French speech.

D. A possible application of RM space to quantifying non-native prosody

Learning to speak a second language with fluency requires acquisition of both segmental and prosodic patterns. Difficulty with L2 prosody is recognized as a challenge for language learners, particularly when L1 and L2 are rhythmically or melodically quite distinct (Pike, 1945; Chela-Flores, 1994). Yet there are very few quantitative methods for assessing non-native prosody. Such methods could prove useful for providing quantitative feedback to language learners in computer-based accent training programs (i.e., when practicing without the benefit of feedback from a human teacher), as well as for basic research on prosodic acquisition. RM space may have some useful qualities in this regard. As discussed in Sec. III C, it is possible to compute the prosodic distance between any two points in RM space [see Fig. 6 and Eq. (3)]. This means RM space could be used to quantify a speaker's prosodic distance from a target language which he or she is trying to acquire. For example, imagine that French and English represent $L1$ and $L2$ for a hypothetical language learner X. Further imagine that X is asked to read a set of sentences in $L2$, and nPVI and MIV measurements are made of these sentences based on the techniques used in the current study. X's mean nPVI and MIV values would define a point in RM space, whose prosodic distance from native English values, pd(X,Es), could be quantified using Eq. (3). This number could serve as the basis for a quantitative prosodic score indicating how close a given speaker was to native $L2$ prosody.

Should such an approach be taken, it will be important to determine how prosodic distance relates to perceptual judgments of foreign accent. For example, there may be a nonlinear relationship between these quantities. Only empirical research can resolve this issue. Fortunately, research relating quantitative measures of prosody to judgments of non-native accent has recently begun, motivated by the desire to test the perceptual relevance of different proposed rhythm measures. For example, White and Mattys (2005b) have studied the rhythm of non-native English as spoken by native Dutch and Spanish speakers. They examined a number of rhythm measures (such as ΔC, $\%V$, nPVI, etc.) and found that the best predictor of native-accent rating was a measure based on the coefficient of variation of vowel duration in sentences This measure, "VarcoV" (cf. Dellwo, 2006) showed an inverse linear relationship with degree of perceived foreign accent.[5] The methods of White and Mattys could be adapted to study the relationship between foreign accent judgments and measures of prosodic distance in RM space.

One notable feature of RM space is that it can easily be generalized to higher dimensions. Since RM space represents duration and pitch, an obvious candidate for an additional dimension is amplitude. In fact, a vowel-based amplitude measure has been shown to differentiate English and French speech. As part of a study comparing the variability of syllabic prominence in the two languages, Lee and Todd (2004) measured the intensity (rms) variation among vowels in a sentence, computing a value ΔI for each sentence, representing the standard deviation of the intensity values. ΔI was significantly larger for English than for French. Inspired by this work, the current study measured the RMS of each vowel in the current corpus (after first normalizing each sentence to 1 V rms). The CV of vowel rms was then computed within each sentence. Consistent with Lee and Todd's (2004) findings, English had a significantly higher vowel rms variability than French (mean and s.e. were 0.53 (0.02) and 0.40 (0.02) for English and French respectively, $p < 0.001$). Thus English and French are distinct in a three-dimensional RM space with nPVI, MIV, and rms variation as orthogonal axes. Prosodic distances can be computed in $3d$ RM space via a generalization of Eq. (3) to three dimensions, and can be used to measure a non-native speaker's prosodic distance from a target language in three dimensions.

Whether using two or three dimensions, the major practical challenge in using RM space is phonetic segmentation of sentences, a time-consuming endeavor when done by hu-

mans. This problem can be alleviated by having speakers read a fixed set of sentences whose texts are known. In this case vowel boundaries can be marked in speech signals using the procedure of forced alignment, in which a speech recognizer is given the list of segments in the order in which they appear in the signal. Dellwo (personal communication) has used this technique with the HTK speech recognition software and found very high agreement with human labeling for speech at a normal rate.

Once vowel boundaries are placed within utterances, then nPVI and MIV measurements can be made based on existing algorithms. It is even possible to compute an accurate estimate of MIV without computing prosograms, by simply assigning each vowel a level pitch based on its median Fo and computing intervals from these pitches. (To illustrate the accuracy of this approach, the resulting mean and s.e. of MIV for English and French in the current corpus were 87 (4) and 68 (3), respectively, compared to 88 (5) and 68 (3) for prosogram-based measures, cf. Table III.) In this case, all that is needed to create RM-space plots such as Fig. 6 are sentences with vowel boundaries marked and Fo contours extracted, together with software that can compute the median Fo of each vowel based on this information. The remaining nPVI and MIV computations can be done via simple equations in a spreadsheet.

V. CONCLUSION

The rhythms and melodies of speech and instrumental music can be quantitatively compared using tools from modern phonetics. Using these tools, an investigation of language and music from England and France confirms the intuition that music reflects the prosody of a composer's native language. The approaches developed here can be applied to the study of language-music relations in other cultures, and may prove useful in quantifying non-native prosody.

ACKNOWLEDGMENTS

We thank Franck Ramus for providing the English and French sentences (audio files and phonetic transcriptions), Piet Mertens for help with prosograms, Peter Ladefoged for help with IPA transcription of the English sentences, Laura Dilley and D. Robert Ladd for helpful comments on this manuscript, and Philip Ball for drawing our attention to "I vow to thee my country." This work was supported by the Neurosciences Research Foundation as part of its research program on music and the brain at The Neurosciences Institute, where A.D.P. is the Esther J. Burnham Fellow and J.R.I. is the Karp Foundation Fellow, and by a grant from the H.A. and Mary K. Chapman Charitable Trust.

APPENDIX A: ENGLISH AND FRENCH SENTENCES

A hurricane was announced this afternoon on the TV.

My grandparent's neighbor's the most charming person I know.

Much more money will be needed to make this project succeed.

The local train left the station more than 5 minutes ago.

The committee will meet this afternoon for a special debate.

The parents quietly crossed the dark room and approached the boy's bed.

This supermarket had to close due to economic problems.

In this famous coffee shop you will eat the best donuts in town.

This rugby season promises to be a very exciting one.

Science has acquired an important place in western society.

The last concert given at the opera was a tremendous success.

In this case, the easiest solution seems to appeal to the court.

Having a big car is not something I would recommend in this city.

They didn't hear the good news until last week on their visit to their friends.

Finding a job is difficult in the present economic climate.

The library is open every day from 8 a.m. to 6 p.m.

The government is planning a reform of the education program.

This year's Chinese delegation was not nearly as impressive as last year's.

The city council has decided to renovate the Medieval center.

No welcome speech will be delivered without the press offices' agreement.

Les parents se sont approchés de l'enfant sans faire de bruit.

Cette boulangerie fabrique les meilleurs gâteaux de tout le quartier.

La femme du pharmacien va bientôt sortir faire son marché.

Les voisins de mes grandparents sont des personnes très agréables.

Il faudra beaucoup plus d'argent pour mener à bien ce projet.

Le magasin est ouvert sans interruption toute la journée.

Les mères sortent de plus en plus rapidement de la maternité.

L'été sera idyllique sur la côte méditerranéenne.

Ils ont appris l'évènement au journal télévisé de huit heures.

La nouvelle saison théâtrale promet d'être des plus intéressante.

Un tableau de très grande valeur a été récemment dérobé.

Le plus rapide est encore le recours auprès de la direction.

Les récents événements ont bouleversé l'opinion internationale.

Le train express est arrivé en gare il y a maintenant plus de 5 minutes.

La reconstruction de la ville a commencé après la mort du roi.

L'alcool est toujours la cause d'un nombre important d'accidents de la route.

Aucune dérogation ne pourra être obtenue sans l'avis du conseil.

Les banques ferment particulièrement tôt le vendredi soir.

Trouver un emploi est difficile dans le contexte économique actuel.

Le ministère de la culture a augmenté le nombre de ces subventions.

APPENDIX B: COMPOSERS AND MUSICAL THEMES

Code numbers are those used in Barlow and Morgenstern (1983).

English: **Bax** b508, b509, b510, b511, b515, b517, b518, b519, b520. **Delius** d189, d191, d192, d193, d194, d195, d196, d197, d198, d199, d200, d201, d202, d205, d208, d214, d215, d216, d219. **Elgar** e3, e4, e7, e8, e13, e14, e15, e16, e17, e18, e19, e20, e21, e23, e27, e28, e30, e31, e33, e34, e35, e51, e52, e53, e56, e58, e60, e61, e62, e63, e64, e66, e67, e68, e70, e71, e72, e73a, e73b, e73c, e73d, e73f, e73h, e73i, e73j. **Holst** h798, h799, h801, h803, h804, h805,

h806, h807, h810, h811, h813, h814, h817, h818, h819, h820. **Ireland** i95, i97, i98, i102, i104, i105, i109, i110, i111, i112, i113. **Vaughan Williams** v4, v5, v6, v7, v8, v12, v13, v14, v17, v18, v19, v20, v21, v22, v23, v24, v26, v27, v28, v29, v30, v31, v32, v33, v34, v35, v37, v38, v39, v40, v41, v42, v43, v44, v45, v49.

French: **Debussy** d13, d14, d20, d21, d42, d43, d55, d57, d58, d62, d70, d71, d74, d77, d78, d80, d83, d85, d86, d87, d88, d90, d97, d98, d100, d105, d107, d108, d109, d113, d116, d117, d118, d122, d123, d124, d125, d126, d127, d129, d132, d134, d135, d138, d139, d140. **Fauré** f60, f61, f62, f63, f72, f75, f76, f76d, f77, f78, f79, f80, f84, f85, f87, f89, f91, f92, f93, f94, f95, f97, f98, f101, f102, f103, f104, f105. **Honegger** h830, h832, h833, h834, h836, h842, h843, h844. **Ibert** i1, i3, i4, i6, i8, i13, i14, i24, i26, i27. **D'Indy** i31, i33, i40, i41, i42, i44, i47, i48. **Milhaud** m382, m383, m384, m386, m387, m394, m395. **Poulenc** p170, p171, p176, p177, p178. **Ravel** r124, r128, r129, r130, r132, r133, r147, r148, r150, r151, r152, r153, r154, r155, r156, r183, r184, r186. **Roussel** r407, r409, r410, r411, r412, r416, r417, r419, r420, r422, r423. **Saint-Saëns** s18, s20, s21, s22, s26, s31, s32, s33, s34, s35, s36, s40, s42, s49, s50, s66, s69, s77, s79, s89, s92, s98, s99, s100, s102, s103, s104, s105, s106, s107, s108, s109, s110, s112, s114, s127, s129, s133, s134.

[1]In the case of Thai and other languages with phonemic vowel length contrast, a high nPVI value may be driven by these length contrasts in addition to (or rather than) vowel reduction.

[2]It should be noted that all published nPVI studies comparing English and French have focused on speakers of the standard dialect of each language. Research on dialectal variation in English speech rhythm suggests that within-language variation in nPVI is smaller than the nPVI difference between English and French (E. Ferragne, unpublished data; White and Mattys, 2005a), but dialectical rhythm studies in both languages are needed to establish whether within-language variation is smaller than between-language variation.

[3]As with the better-known autosegmental-metrical approach, the prosogram is an abstraction of an Fo curve. The current study prefers the prosogram to the autosegmental-metrical abstraction because it is explicitly concerned with the psychoacoustics of intonation perception, and thus seems better suited to comparing speech and music as patterns of perceived pitches. That being said, it should be noted that the prosogram is based on research on native listeners of intonation languages, and hence its applicability to pitch patterns in tone languages is uncertain.

[4]The prosogram is freely available from http://bach.arts.kuleuven.be/pmertens/prosogram/, and runs under Praat, which is freely available from http://www.fon.hum.uva.nl/praat/

[5]It should be noted that VarcoV was only slightly better than nPVI as a predictor of accent ratings in this study. Furthermore, a concern about this research is that segmental and suprasegmental cues are probably both contributing to foreign accent ratings. Thus some method is needed to remove interspeaker variability in segmental cues, such as speech resynthesis. In fact, White and Mattys (personal communication) are pursuing such an approach.

Abraham, G. (**1974**). *The Tradition of Western Music* (University of California Press, Berkeley), Chap. 4, pp. 61–83.

Arvaniti, A., Ladd, D. R., and Mennen, I. (**1998**). "Stability of tonal alignment: The case of Greek prenuclear accents," J. Phonetics **26**, 3–25.

Atterer, M. and Ladd, D. R. (**2004**). "On the phonetics and phonology of 'segmental anchoring' of Fo: Evidence from German," J. Phonetics **32**, 177–197.

Barlow, H. and Morgenstern, S. (**1983**). *A Dictionary of Musical Themes*, revised ed. (Faber and Faber, London).

Beckman, M. (**1992**). "Evidence for speech rhythm across languages," in *Speech Perception, Production, and Linguistic Structure*, edited by Y. Tohkura *et al.* (IOS, Tokyo), pp. 457–463.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," in Proceedings of the Institute Of Phonetic Sciences, University of Amsterdam, Vol. **17**, pp. 97–110.

Bolinger, D. (**1981**). *Two Kinds of Vowels, Two Kinds of Rhythm* (Indiana University Linguistics Club, Bloomington).

Chela-Flores, B. (**1994**). "On the acquisition of English rhythm: Theoretical and practical issues," IRAL **32**, 232–242.

d'Alessandro, C. and Castellengo, M. (**1994**). "The pitch of short-duration vibrato tones," J. Acoust. Soc. Am. **95**, 1617–1630.

d'Alessandro, C. and Mertens, P. (**1995**) "Automatic pitch contour stylization using a model of tonal perception," Comput. Speech Lang. **9**, 257–288.

Daniele, J. R. and Patel, A. D. (**2004**). "The interplay of linguistic and historical influences on musical rhythm in different cultures," in Proceedings of the 8th International Conference on Music Perception and Cognition, Evanston, IL, pp. 759–762.

Dauer, R. M. (**1983**). "Stress-timing and syllable-timing reanalyzed," J. Phonetics **11**, 51–62.

Dauer, R. M. (**1987**). "Phonetic and phonological components of language rhythm," in Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn, Vol. **5**, pp. 447–450.

Delattre, P. (**1963**). "Comparing the prosodic features of English, German, Spanish and French," IRAL **1**, 193–210.

Delattre, P. (**1966**). "A comparison of syllable length conditioning among languages," IRAL **4**, 183–198.

Delattre, P. (**1969**). "An acoustic and articulatory study of vowel reduction in four languages," IRAL **7**, 295–325.

Dellwo, V. (**2006**). "Rhythm and Speech Rate: A variation coefficient for deltaC," in *Language and Language-processing. Proceedings of the 38th linguistic Colloquium, Piliscsaba 2003*, edited by P. Karnowski and I. Szigeti(Peter Lang, Frankfurt am Main), pp. 231–241

Dellwo, V., Steiner, I., Aschenberner, B., Dankovičová, J., and Wagner, P. (**2004**). "The BonnTempo-Corpus and BonnTempo-Tools: A database for the study of speech rhythm and rate," in Proceedings of the 8th ICSLP, Jeju Island, Korea.

Di Cristo, A. (**1998**). "Intonation in French," in *Intonation Systems: A Survey of Twenty Languages*, edited by D. Hirst and A. Di Cristo (Cambridge University Press, Cambridge), pp. 195–218.

Dilley, L. (**2005**). "The phonetics and phonology of tonal systems," Ph.D. dissertation, MIT.

Grabe, E., Post, B., and Watson, I. (**1999**). "The acquisition of rhythmic patterns in English and French," in Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, pp. 1201–1204.

Grabe, E. and Low, E. L. (**2002**). "Durational variability in speech and the rhythm class hypothesis," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, Berlin), pp. 515–546.

Grout, D. J. and Palisca, C. V. (**2000**). *A History of Western Music* 6th ed. (Norton, New York).

Hall, R. A., Jr. (**1953**). "Elgar and the intonation of British English," The Gramophone, June 1953:6-7, reprinted in *Intonation: Selected Readings*, edited by D. Bolinger (Penguin, Harmondsworth), pp. 282–285.

Hermes, D. and van Gestel, J. C. (**1991**). "The frequency scale of speech intonation," J. Acoust. Soc. Am. **90**, 97–102.

House, D. (**1990**) *Tonal Perception in Speech* (Lund University Press, Lund).

Huron, D. and Ollen, J. (**2003**). "Agogic contrast in French and English themes: Further support for Patel and Daniele," Music Percept. **21**, 267–271.

Jones, M. R. (**1987**). "Dynamic pattern structure in music: recent theory and research," Percept. Psychophys. **41**, 621–634.

Jones, M. R. (**1993**). "Dynamics of musical patterns: How do melody and rhythm fit together?" in *Psychology and Music: The Understanding of Melody and Rhythm*, edited by T. J. Tighe and W. J. Dowling (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 67–92.

Jusczyk, P. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, MA).

Jun, S.-A. (**2005**). "Prosodic Typology," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S-A. Jun (Oxford University Press, Oxford), pp. 430–458.

Jun, S-A. and Fougeron, C. (**2000**) "A Phonological Model of French Intonation," in *Intonation: Analysis, Modeling and Technology*, edited by A. Botinis (Kluwer Academic, Dordrecht), pp. 209–242.

Krumhansl, C. (**2000**). "Tonality induction: A statistical approach applied

cross-culturally," Music Percept. **17**, 461–479.
Krumhansl, C., Toivanen, P., Eerola, T., Toiviainen, P., Järvinen, T., and Louhivuori, J. (**2000**). "Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks," Cognition **76**, 13–58.
Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).
Ladd, D. R. and Morton, R. (**1997**). "The perception of intonational emphasis: Continuous or categorical?," J. Phonetics **25**, 313–342.
Ladd, D. R., Faulkner, D., Faulkner, H., and Schepman, A. (**1999**). "Constant 'segmental anchoring' of Fo movements under changes in speech rate," J. Acoust. Soc. Am. **106**, 1543–1554.
Lee, C. S. and Todd, N. P. McA. (**2004**). "Toward an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora," Cognition **93**, 225–254.
Lerdahl, F. and Jackendoff, R. (**1983**). *A Generative Theory of Tonal Music* (MIT Press, Cambridge, MA).
Liberman, M. (**1975**). "The intonational system of English," Ph.D. thesis, MIT.
Low, E. L. (**1998**). "Prosodic prominence in Singapore English," Ph.D. thesis, University of Cambridge.
Low, E. L., Grabe, E., and Nolan, F. (**2000**). "Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English," Lang Speech **43**, 377–401.
Maidment, J. (**1976**). "Voice fundamental frequency characteristics as language differentiators," Speech and Hearing, Work in Progress, Univ. College London, pp. 75–93.
Mehler, J., Dupuox, E., Nazzi, T., and Dehaene-Lambertz, D. (**1996**). "Coping with linguistic diversity: The infant's viewpoint," in *Signal to Syntax*, edited by J. L. Morgan and D. Demuth (Lawrence Erlbaum, Mahwah, NJ), pp. 101–116.
Mertens, P. (**2004a**). "The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model," in Proceedings of Speech Prosody 2004, Nara, Japan, pp. 23–26.
Mertens, P. (**2004b**). "Un outil pour la transcription de la prosodie dans les corpus oraux," Traitement Automatique des Langues **45**, 109–130.
Mertens, P. and d'Alessandro, C. (**1995**). "Pitch contour stylization using a tonal perception model," in Proceedings of the 13th International Congress on Phonetic Sciences, Stockholm, pp. 228–231.
Mertens, P., Beaugendre, F., and d'Alessandro, C. (**1997**). "Comparing approaches to pitch contour stylization for speech synthesis," in *Progress in Speech Synthesis*, edited by J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg (Springer Verlag, New York), pp. 347–363.
Nazzi, T., Bertoncini, J., and Mehler, J. (**1998**). "Language discrimination in newborns: Toward an understanding of the role of rhythm," J. Exp. Psychol. Hum. Percept. Perform. **24**, 756–777.
Nespor, M. (**1990**). "On the rhythm parameter in phonology." in *Logical Issues in Language Acquisition*, edited by I. Rocca (Foris, Dordrecht), pp. 157–175.
Nolan, F. (**2003**). "Intonational equivalence: An experimental evaluation of pitch scales," in Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, pp. 771–774.
Oram, N. and Cuddy, L. L. (**1995**). "Responsiveness of Western adults to pitch distributional information in melodic sequences," Psychol. Res. **57**, 103–118.
Palmer, C. (**1997**). "Music performance," Annu. Rev. Psychol. **48**, 115–138.
Patel, A. D. and Daniele, J. R. (**2003**). "An empirical comparison of rhythm in language and music," Cognition **87**, B35–B45.
Patel, A. D. and Daniele, J. R. (**2003b**). "Stress-timed vs syllable-timed music? A comment on Huron and Ollen," Music Percept. **21**, 273–276.
Peterson, G. E. and Lehiste, I. (**1960**). "Duration of Syllabic Nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Pierrehumbert, J. (**1980**). "The phonetics and phonology of English intonation," Ph.D. dissertation, MIT.
Pike, K. N. (**1945**). *The Intonation of American English* (University of Michigan, Ann Arbor).
Ramus, F., Nespor, M., and Mehler, J. (**1999**). "Correlates of linguistic rhythm in the speech signal," Cognition **73**, 265–292.
Ramus, F. (**2002**). "Acoustic correlates of linguistic rhythm: Perspectives," in Proceedings of Speech Prosody, Aix-en-Provence, pp. 115–120.
Ramus, F. (**2002b**). "Language discrimination by newborns: teasing apart phonotactic, rhythmic, and intonational cues," Annual Review of Language Acquisition **2**, 85–115.
Repp, B. H. (**1992**). "Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's 'Träumerei'," J. Acoust. Soc. Am. **92**, 2546–2568.
Roach, P. (**1982**). "On the distinction between 'stress-timed' and 'syllable-timed' languages," in *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour of F.R. Palmer*, edited by D. Crystal (Edward Arnold, London), pp. 73–79.
Rossi, M. (**1971**). "Le seuil de glissando ou seuil de perception des variations tonales pour la parole," Phonetica **23**, 1–33.
Rossi, M. (**1978a**). "La perception des glissandos descendants dans les contours prosodiques," Phonetica **35**, 11–40.
Rossi, M. (**1978b**). "Interactions of intensity glides and frequency glissandos," Lang Speech **21**, 384–396.
Sadakata, M. and Desain, P., "Comparing rhythmic structure in popular music and speech" (submitted).
Saffran, J. R., Aslin, R. N., and Newport, E. L. (**1996**). "Statistical learning by 8-month old infants," Science **274**, 1926–1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., and Newport, E. L. (**1999**). "Statistical learning of tone sequences by human infants and adults," Cognition **70**, 27–52.
Selkirk, E. O. (**1984**). *Phonology and Syntax: The Relation Between Sound and Structure* (MIT Press, Cambridge, MA).
'tHart, J. (**1976**). "Psychoacoustic backgrounds of pitch contour stylization." I. P. O. Annual Progress Report **11**, 11–19.
'tHart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge).
Wenk, B. J. (**1987**). "Just in time: On speech rhythms in music," Linguistics **25**, 969–981.
White, L. S. and Mattys, S. L. (**2005a**). "Calibrating rhythm: A phonetic study of British Dialects." in Proceedings of the Vth Conference on UK Language Variation and Change, Aberdeen, Scotland.
White, L. S. and Mattys, S. L. (**2005b**). "How far does first language rhythm influence second language rhythm?," in Proceedings of Phonetics and Phonology in Iberia, Barcelona, Spain.
Willems, N. (**1982**). *English Intonation from a Dutch Point of View* (Foris, Dordrecht).

Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei

Matthew Aylett[a)]
Rhetorical Systems Ltd. UK
Department of Linguistics and English Language, University of Edinburgh, United Kingdom

Alice Turk[b)]
Department of Linguistics and English Language, University of Edinburgh, United Kingdom

(Received 24 November 2004; revised 19 February 2006; accepted 21 February 2006)

The language redundancy of a syllable, measured by its predictability given its context and inherent frequency, has been shown to have a strong inverse relationship with syllabic duration. This relationship is predicted by the smooth signal redundancy hypothesis, which proposes that robust communication in a noisy environment can be achieved with an inverse relationship between language redundancy and the predictability given acoustic observations (acoustic redundancy). A general version of the hypothesis predicts similar relationships between the spectral characteristics of speech and language redundancy. However, investigating this claim is hampered by difficulties in measuring the spectral characteristics of speech within large conversational corpora, and difficulties in forming models of acoustic redundancy based on these spectral characteristics. This paper addresses these difficulties by testing the smooth signal redundancy hypothesis with a very high-quality corpus collected for speech synthesis, and presents both durational and spectral data from vowel nuclei on a vowel-by-vowel basis. Results confirm the duration/ language redundancy results achieved in previous work, and show a significant relationship between language redundancy factors and the first two formants, although these results vary considerably by vowel. In general, however, vowels show increased centralization with increased language redundancy. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188331]

PACS number(s): 43.70.−h, 43.71.−k, 43.71.Es, 43.70.Bk [RLD] Pages: 3048–3058

I. INTRODUCTION

Much of speech research has been devoted to the explanation and description of within-speaker variation in the production of the same phoneme in different contexts. One fairly consistent observation made by researchers going back to the 1960's (Bolinger, 1963; Lieberman, 1963; Sharp, 1960) is that the more predictable a section of speech, either because of context or inherent frequency, the shorter, or more reduced, it tends to be.

This predictability due to language structure can be termed language redundancy. Its correlation with duration is predicted by the hypothesis that the greater the language redundancy of a word, syllable or phoneme, the lower its acoustic redundancy is likely to be, where acoustic redundancy refers to its likelihood of recognition based on acoustic properties alone. On this view, duration reflects acoustic redundancy since longer segments are arguably more salient and recognizable, all other things being equal. Recent work has confirmed these earlier findings for duration (e.g., Aylett, 2000; Bybee, 2000; Bell *et al.*, 2003; Wright, 2003; Munson and Soloman, 2004).

Spectral characteristics (e.g., formant frequencies in the central portion of a vowel, among other things) can also be considered as measures of acoustic redundancy, since these are known to relate to vowel distinctiveness and thus probability of recognition. However, although spectral differences have clearly been shown to occur due to differences in speaking style (Picheny *et al.*, 1986; Lindblom, 1990; Moon and Lindblom, 1994; Bradlow *et al.*, 1996), prosodic variation (Summers, 1987; van Bergem, 1993) and also lexical neighborhood density (Wright, 2003; Munson and Soloman, 2004), evidence showing a language redundancy relationship with spectral measurements has been more elusive. Wright (2003), Munson and Soloman (2004), and Bybee (2000) all showed a relationship between word frequency and spectral measures, with vowel centralization more prevalent in more frequent words. Because these studies used a relatively small set of laboratory data, it is difficult to know whether these findings extend to a more general context. Relevant corpus studies include Aylett (2000, 2001), which show only a weak correlation between language redundancy and vocalic spectral change, and Jurafsky, Bell, and others (Jurafsky *et al.*, 2001, 2003; Bell *et al.*, 2003), which show a more robust result, but for vowel spectral reduction of function words only.

The relationship between language redundancy and acoustic redundancy has important implications for theories of speech production and perception because it offers a framework for predicting variation in articulation and, more crucially, a framework for explaining this variation in terms of speech perception. The idea of relating production to perception is by no means new. Lindblom (1990) in his H&H

[a)]Electronic mail: matthewa@cogsci.ed.ac.uk
[b)]Electronic mail: turk@ling.ed.ac.uk

 0001-4966/2006/119(5)/3048/11/$22.50

theory explicitly presents the idea that articulation varies in order to reduce effort while at the same time producing a signal which shows sufficient acoustic distinctiveness for the listener to correctly identify the linguistic content of the message. Language redundancy offers a means of quantitatively modeling differences in acoustic distinctiveness because the more predictable a section of speech, given its context, the less acoustic distinctiveness (or acoustic redundancy) is required for the listener to successfully identify the contents. Thus, the relationship between language redundancy and acoustic redundancy should implicitly mirror Lindblom's findings. In addition, by offering a quantitative framework for measuring what is sufficiently acoustically distinct, it allows the investigation of these phenomena across large corpora and thus many more contexts, speakers, and speech styles than is possible in an experimental setting. Indeed, studies such as Jurafsky (2001), van Son and Pols (2003), Aylett (2000), and Aylett and Turk (2004) have all tested the relationship between language and acoustic redundancy for natural speech corpora. These authors share Lindblom's view that the more predictable a section of speech the more articulatory reduction/centralization we expect to see (Moon and Lindblom, 1994; implicit in Lindblom, 1990). This view is referred to as the probabilistic reduction hypothesis (Jurafsky, 2001), speech efficiency (van Son and Pols, 2003), and is included as a fundamental element of Aylett and Turk's (2004) smooth signal redundancy hypothesis.

The probabilistic reduction hypothesis (Jurafsky *et al.*, 2001) states that word forms are reduced when they have higher probability, and is thus more general than theories concerned only with word frequency (e.g., Zipf, 1949) and predictability (Fowler and Housum, 1987) because, "The probability of a word is conditioned on many aspects of its context, including neighboring words, syntactic and lexical structure, semantic expectations, and discourse factors" (Jurafsky *et al.*, 2001, p. 229). van Son and Pols (1999) support the relationship between language redundancy and acoustic characteristics with phonemic data, and propose that the underlying reason for this language/acoustic redundancy relationship is to make speech an efficient communication channel. Aylett and Turk's (2004) smooth signal redundancy hypothesis also proposes that word forms are reduced when they are associated with a higher probability of recognition, and agrees with van Son and Pols' claim that this relationship between language redundancy and acoustic characteristics is governed by the drive to make the speech signal more efficient. However, it makes a wider claim, first that this efficiency is gained in order to make speech robust in a noisy environment, and second that acoustic differences are linguistically implemented through prosodic prominence structure. In contrast to this view, the probabilistic reduction hypothesis argues that "probabilistic relations between words are represented in the mind of the speaker" (Jurafsky *et al.*, 2001), and thus relate directly to their acoustic correlates without the intermediary of prosodic prominence structure.

In this paper an analysis of a large speech corpus (approximately 500 000 syllables) was carried out with the following goals: (1) to test for the expected relationship between language redundancy and syllabic duration; (2) to determine the relationship between language redundancy and measures of the first and second formants; and (3) to determine the degree to which prosodic prominence and language redundancy predict the same variation in each of these measures, as predicted by the smooth signal redundancy hypothesis.

A. The smooth signal redundancy hypothesis

The smooth signal redundancy hypothesis claims that the acoustic consequences of differences in redundancy can be explained functionally within an information theoretical framework, by the drive for speakers to achieve robust information transfer in a potentially noisy environment while conserving effort. These pressures encourage speakers to produce utterances whose elements have similar probabilities of recognition, that is, utterances with a smooth signal redundancy profile. In Aylett and Turk (2004) we presented evidence showing that phrase-medial syllables with high language redundancy (i.e., they are highly predictable from lexical, syntactic, semantic, and pragmatic factors) were shorter than less predictable elements.

To take a concrete example, in the utterance "I'm going to the beach," the word "to" is more likely to occur than the word "beach" since "to" is a relatively high-frequency closed class word, and often occurs immediately after the word "going," whereas "beach" is a lower frequency word that is not predictable from its preceding context. Consequently, the amount of acoustic information both in terms of duration and in terms of spectral distinctiveness is lower in "to" than in "beach," which contains a full vowel, and is likely to bear phrasal stress. If we regard the overall probability of recognizing the lexical item as the probability of recognition given a language model multiplied by the probability of recognition given an acoustic model, we can see that the overall signal probability is smoothed by this inverse relationship. This smoothness adds robustness to communication because information content is spread evenly across the signal.

Because prosodic prominence appears to largely shadow redundancy effects (e.g., the highly language redundant word *to* is unstressed in the above example while the less predictable *beach* bears a nuclear pitch accent (phrasal stress), and because it is arguably unrealistic for articulation to adjust itself to reflect minor changes in redundancy, the smooth signal redundancy hypothesis proposes that language redundancy is implemented through prosodic prominence structure, which itself has its own language-specific conventions and rules. Aylett and Turk's (2004) present supporting evidence of a large shared influence of language redundancy and prosodic prominence structure on duration, with a small, expected, unique influence of prosodic prominence structure. In addition, a small unexpected unique influence of language redundancy factors was observed.

In this paper, we use spectral data to test a more general version of the smooth signal redundancy hypothesis that claims that any acoustic feature used to recognize the identity of a syllable could be used to smooth signal redundancy by covarying inversely with language redundancy. The measures used to test this version of the hypothesis were $F1/F2$

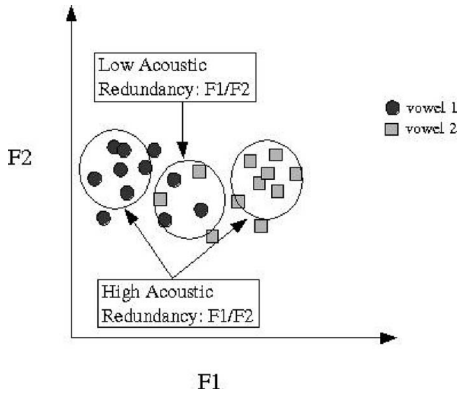


FIG. 1. Illustration of acoustic redundancy differences in an $F1/F2$ vowel space.

formant values at the temporal midpoint of vowel nuclei. These measures were chosen because relationships between vowel centralization (as measured by $F1/F2$) and intelligibility are well documented (see Sec. I B), and like syllable duration, variation in $F1/F2$ values relates straightforwardly to acoustic redundancy. For example, the probability of identifying a particular vowel is higher if its formant values are unambiguously associated with a single vowel category, and not simultaneously associated with multiple vowel categories (see Fig. 1).

B. The vowel space

The $F1/F2$ formant values for vowels can be measured by reference to the acoustic vowel space, a two-dimensional space defined by vocalic $F1/F2$ formant values. The smaller the vowel space the more bunched and centralized the vowels; the larger the vowel space the more distinctive and peripheral the vowels. It has been shown that expanded vowel spaces (associated with hyperspeech) are correlated with speech intelligibility, and thus increased acoustic redundancy (Bradlow *et al.*, 1996; Bond and Moore, 1994). In contrast, casual or reduced speech (hypospeech) is characterized by centralization (Picheny *et al.*, 1986; Lindblom, 1990; Moon and Lindblom, 1994; Bradlow *et al.*, 1996), a reduced vowel space, and arguably lower acoustic redundancy. A reasonable prediction would therefore be, that as language redundancy increases, average formant values would shift towards the neutral, centralized position, leading to lower acoustic redundancy and, smoother signal redundancy overall. For example, in the case of /i/ (as in "lease"), as language redundancy increases, $F1$ would be expected to rise and $F2$ to fall while for /ɑ/ (as in "father") $F1$ would be expected to fall and $F2$ to rise.

Different specific predictions for formants of vowels with different height and backness specifications raise the issue of how to compare centralization results across vowels. Previous work has often measured the size of the vowel space and argued that if the size increases the vowels are less reduced (e.g., Bradlow *et al.*, 1996; Wright, 2003; Munson and Soloman, 2004). This method works quite well for a constrained set of laboratory stimuli, but has the disadvantage of masking the way results vary by vowel and by formant. Similar problems occur using distance or "clarity" measurements, whereby all vowels are related to a model of a citation "clear" version of the vowel space, using a value that reflects how distinctive or clear each vowel token is given this model. The advantage of this approach is that a single comparable measurement can then be used for analysis. In previous work, a number of alternative vowel space models were applied to a single dataset in an attempt to produce a clarity score that would show a systematic relationship with language redundancy. These included a citation vowel space model to measure degrees of hyperarticulation (Aylett, 2000), hidden Markov models (HMMs) (Aylett, 2001), as well as a simple Euclidean distance to the center of the vowel space in terms of mean $F1/F2$ (1) across all vowels Aylett (2000) and (2) from a speaker's citation speech (prepublished versions of Aylett and Turk, 2004). Results for all techniques were disappointing, arguably because any modeling technique introduces articulatory and perceptual assumptions about degree of distinctiveness or clarity that obscure spectral effects that are subtle by nature and potentially vary across vowels.

Having said this, the importance of modeling the vowel space should not be discounted, for much progress has been made in developing perceptual scales and by-vowel normalization schemes (see Rosner and Pickering, 1994, for a review). However, for analyses relating language redundancy to spectral measures, the advantage of a single model-derived measure applied to all vowels is heavily outweighed by the potential noise such a model might introduce. For this reason, absolute $F1/F2$ measurements in hertz were used in the current study. By doing so, theoretical assumptions concerning the relationship between human vowel perception and production were intentionally avoided. By avoiding the pitfalls of these previous studies, the current study was able to show (1) that raw $F1/F2$ values do vary significantly with language redundancy factors; (2) that the direction of this change is one that can be interpreted as reduction, and (3) the extent of the overlap between redundancy and prosodic factors in terms of predictive power.

II. METHODOLOGY

A very large corpus of citation speech (the Rhetorical Corpus detailed in Sec. II A) was used to test the smooth signal redundancy hypothesis. This corpus was chosen because its recording quality and carefully articulated speech allowed reliable automatic formant tracking and segmentation. Each syllable of the corpus was scored on the basis of (1) a simple language redundancy model based on syllable n-gram probability; and (2) a traditional prosodic model (detailed in Sec. II B), and was represented in terms of its acoustic redundancy, i.e., in terms of (1) its $F1/F2$ values at the temporal midpoint of its vocalic nucleus; and (2) in terms of its duration. ANOVAs with posthoc tests, as well as regression analyses, were applied to investigate the relationship between language redundancy and prosodic prominence and these two types of acoustic redundancy measures. (See Sec. II D for details of the statistical analyses.)

A. Materials

1. The Rhetorical Corpus

The Rhetorical Corpus was collected by Rhetorical Systems Ltd. as a database for speech synthesis. It contains data from eight general American speakers, three female and five male, with approximately 50 000 syllables recorded for each speaker. The speakers were all professional voice talents. The material was read in a recording studio environment and contained sentences of varying lengths taken from different genres (e.g., travel directions, financial news). Disfluencies and mispronunciations were avoided by requiring speakers to repeat errorful utterances. The resulting corpus, as compared to spontaneous speech corpora such as the Map Task (Anderson *et al.*, 1991) or switchboard (Godfrey *et al.*, 1992), lends itself to spectral analysis due to the high quality of the recording, the relatively slow rate of speech (lower than typical spontaneous speech), the quality of the speakers who were chosen for their ability to produce clear natural citation speech, and the filtering out of disfluency.

2. Vowels

Results presented in Sec. III are for a limited set of possible vowels within our corpus: /ɑ/ (father), /æ/ (tap), /ɛ/ (less), /i/ (lease), /u/ (goose). /ə, ɪ, ʊ, ʌ/ were avoided since these vowels were often too short for accurate automatic spectral analysis. In addition, diphthongs were avoided due to their moving spectral targets which would require more than a single measurement. And finally, the /ɒ/ vowel (as in "lawn") was avoided because of the extensive pronunciation variation of this vowel within the general American accent (Wells, 1982).

3. Segmentation and spectral analysis

As part of rhetorical systems' synthesis process, the corpus was phonemically segmented using a mixture of proprietal automatic segmentation and hand correction. Results are similar, although superior, to automatic segmentation using HTK (Young *et al.* (1996)), which was used as a basis for word medial segmentation in Aylett and Turk (2004). Using this segmentation, log duration in milliseconds was calculated for each syllable in the corpus, and $F1$ and $F2$ values (Hz) at the temporal midpoint of each vowel were estimated using the ESPS formant tracker (Talkin, 1987).

B. Redundancy and prosodic models

1. The acoustic redundancy model

In these analyses, vowel nucleus first and second formant values were taken as measures of a syllable's spectral quality, and centralization in terms of $F1/F2$ was interpreted as reduced acoustic redundancy. Due to the possibility of centralization behavior differing markedly between vowels, $F1/F2$ results are presented separately for each vowel. In addition, syllable duration was used as another measure of acoustic redundancy, and allowed for comparison with results in the literature, e.g., Aylett and Turk (2004).

TABLE I. Contribution from the original n-gram log likelihoods to wide and narrow factors. Values for the wide and narrow factors were produced by normalizing unigram, bigram, and trigram values using mean and standard deviation values and by multiplying them by the factor analysis components.

	Factor analysis component			
	Wide	Narrow	Mean	Standard deviation
Unigram	0.315	1.090	−6.924	1.956
Bigram	0.421	−0.314	−5.480	2.931
Trigram	0.407	−0.520	−5.063	3.411

2. The language redundancy model

A wide variety of possible measurements can be used to represent language redundancy. Examples in previous work have included word frequency, syllabic trigram models, accessibility (Aylett, 2000), joint probability, and conditional probability (Bell *et al.*, 2003), and Bayesian probability given phonemic identity (van Son and Pols, 1999). All of these measurements tended to support the hypothesis that there is a relationship between language redundancy and acoustic realization.

Syllabic probabilities were determined over unigram, bigram, and trigram contexts, i.e., the likelihood of a syllable without context, given the previous syllable, and given the previous two syllables. A log transformation was applied to these likelihoods in order to relate the values more closely to information content (Pierce, 1961) and to normalize the distributions. In order to calculate these likelihoods, the CMU language toolkit (Clarkson and Rosenfeld, 1997) was used to compute n-gram statistics based on the syllabic transcriptions of 187 million words found in news resources on the Internet. Transcriptions were automatically generated using the Rhetorical Systems speech synthesizer.

To make analyses of variance possible, and to help determine the direction of language redundancy effects, factor analysis was carried out on these likelihoods. Factor analysis has the effect of reducing the number of variables and, uses a linear transformation to create new factors which have no correlation with each other. The analysis produced two factors which we called wide context and narrow context measures of redundancy, where wide context was roughly an average of all three likelihoods, and narrow context was the unigram likelihood with bigram and trigram likelihoods subtracted (see Table I).

The resulting factors had the following normalized distributions: wide factor mean 0.0, median 0.132, s.d. 1.00; narrow factor mean 0.0, median 0.1130, s.d. 1.00. These distributions explained 96.8% of the variance in unigram/bigram/trigram likelihoods. Data points were then grouped into three redundancy groups on the basis of these distributions, for use in analyses of variance where language redundancy grouping was the independent variable of interest. The groups were constructed as follows:

(1) High language redundancy group: Data points for which both narrow and wide redundancy factors were higher than the median value (+, +);

TABLE II. Example syllables from each language redundancy group (in capitals). Example syllables are shown along with the two preceding syllables used to compute the target syllable's language redundancy.

Language Redundancy		
High	Medium	Low
to town **OF**	*pause* from **O**maha	flag is **NAILED**
alternative **IN**	tanza**NI**a	unper **TURBED**
fathers such **A**	medi**NA**	radioactive gas **GUSHED**
andy **TO**	*pause* in **PO**tentially	alba**TROSS**
kind of **THE**	flights from **MA**laysia	lab lo**CUSTS**

(2) Medium language redundancy group: Data points for which either (1) narrow redundancy was higher than the median and wide redundancy was lower than the median or (2) narrow redundancy was lower than the median and wide redundancy was higher than the median. (+,−) or (−,+); and
(3) Low language redundancy group: Data points for which both narrow and wide redundancy factors were lower than the median value (−,−).

See Table II for an example of syllables that fell into each of these three groups.

3. The prosodic model

The smooth redundancy hypothesis argues that an important function of prosodic prominence is as a linguistic means for increasing smooth signal redundancy (see Aylett, 2000 and Aylett and Turk, 2004 for supporting durational data). The prosodic model used to test this hypothesis in the current study is a simple one that consists of three levels of prosodic prominence and three levels of boundary. It is identical to the one Aylett and Turk (2004) used for their automatically coded data, apart from the full/reduced vowel distinction that does not apply to the current, full-vowel only, dataset. It can be summarized as follows:

Prominence:
- None
- Primary lexical stress
- High probability of having a phrasal stress (primary lexical stress+open class)

Boundaries:
- No following prosodic boundary
- Followed by a word boundary
- High probability of a following phrase boundary (followed by a pause >100 ms)

For practical reasons phrasal stress and phrasal boundaries were not hand-labeled, but were instead predicted on the basis of primary lexical stress on open-class words (for phrasal stress), and on the basis of following pauses (for phrasal boundaries). It should be noted that the occurrence of phrasal stress has probably been overpredicted, and the occurrence of phrasal boundaries has probably been underpredicted.

In addition, due to practical limitations, several known distinctions were not taken into account. These include accent type, boundary tone type, phrase type (e.g., full vs intermediate), and distinctions between unstressed syllables and those that could bear secondary lexical stress (e.g., /i/ in "city" and /i/ in "reply"). Despite these obvious limitations, this prosodic model represents a reasonable number of factors known to affect duration and segmental phonetic characteristics, and as such, is a rough approximation of current prosodic phonological theory. For our purposes, it made it possible to automatically code the half million syllables in the Rhetorical Corpus in terms of prosodic factors.

C. Data summary

To summarize, each syllable in our corpus was coded for language redundancy, prosodic prominence, and position relative to prosodic boundaries, according to the two models presented above. Syllable durations were determined on the basis of segmentation already present in the Rhetorical Corpus, and spectral measures ($F1/F2$) at the temporal midpoint of each vowel were determined using the ESPS formant tracker (Talkin, 1987).

D. Data analyses

1. Control for prosodic boundary location and speaker sex

The weak form of the smooth signal redundancy hypothesis accepts that the expected language/acoustic redundancy relationship may be confounded by prosodic boundary markers, which, among other things, act to increase the duration of phrase-final syllables. In order to address this concern, in the data presented here boundaries are controlled by discarding any syllables (1) before a phrase boundary; and (2) with no word boundary following them, i.e., all syllables are discarded not followed by a break index of 1. This approach differs from both Aylett (2000) and Aylett and Turk (2004), where only syllables from monosyllabic words were used. This change was necessary because of differences between the two corpora. The Map Task Corpus (Anderson *et al.*, 1991) used in the earlier work had a high preponderance of monosyllabic words (around 75%), whereas the Rhetorical Corpus had a much higher incidence of polysyllabic words (only 40% were monosyllabic). Whereas the monosyllabic words in the Map Task Corpus arguably provided a fair representation of the overall behavior of data within the corpus, this was not the case for the Rhetorical Corpus. In addition, given the requirement of by-vowel analysis, where data sparsity becomes a significant problem, it was decided to relax the boundary control requirements.

As well as controlling for prosodic boundaries, spectral measure results for male and female speakers were analyzed separately. An alternative approach would have been to manipulate the underlying values using a speaker normalization scheme (e.g., Nearey, 1992). However, there are problems with these techniques (see Adank, 2003 for a review) which would be further compounded by artifacts caused by formant tracking as high female $f0$ can lead to significant formant tracking errors. Separate by-sex spectral analyses have pre-

vented any normalization errors, and have made it possible to evaluate results which may have been compromised by formant tracking errors.

2. Statistical analyses

The smooth signal redundancy hypothesis predicts a strong inverse relationship between language redundancy, and acoustic redundancy, and a positive relationship between prosodic prominence and acoustic redundancy. In addition, it predicts that the predictive power of the language redundancy model and the prosodic prominence model will be substantially shared, while acknowledging that the conventions and rules of English prosodic prominence may yield a unique prediction of variation due to prosodic prominence over and above the shared effects. The significance and direction of these relationships for our two types of dependent variables, syllable duration and vowel $F1/F2$ frequency, were addressed by way of a univariate ANOVA followed by *posthoc* t-tests with Bonferroni correction. The size of the relationships and the shared proportion of the two models predictive power were analyzed by way of multiple linear regression together with a likelihood ratio test (Neter *et al.*, 1990).

a. ANOVA and posthoc t-tests. The prosodic variables in our study were categorical and therefore ideal for an ANOVA model with either syllabic duration or $F1/F2$ formant value as a dependent variable. Continuous language redundancy values were coded in terms of the categorical redundancy grouping variable described in Sec. II B 2 to make ANOVA possible. All *posthoc* t-tests were carried out for effects that were significant at $p<0.05$ in the ANOVA. There is some debate concerning when a Bonferroni correction should be made in these circumstances. A conservative approach to the data was taken here, whereby the correction was used for all tests carried out within the same analysis. To give a concrete example, a comparison of three redundancy groups across five vowels would result in a correction factor of 15, which would then be multiplied by the p value generated by each t-test. Thus, with a correction factor of 15, if a *posthoc* t-test returned a significance of 0.01, the p value would be recomputed as a nonsignificant 0.15.

b. Linear regression and likelihood ratio tests. The main goal of the linear regressions was to show how much the prosodic and redundancy models could predict the dependent variable and to what extent the two models overlapped. To this end, multiple regression analyses with fixed factor ordering were carried out with syllabic duration and $F1/F2$ formant values as dependent variables, and with language redundancy and prosodic factors as independent variables. The result of the regression analysis revealed the size of the relationship between dependent and independent variables in terms of predicted variance. Likelihood ratio tests were then performed to reveal the unique contributions of each type of independent variable to the regression model. This was accomplished by carrying out the linear regression (1) with all factors; (2) without language redundancy factors; and (3) without prosodic factors. The extent the r-squared value decreased when factors were removed was used to represent the unique contribution of these factors. We calculated the shared contribution of the independent factors by subtracting all unique contributions from the predicted variance of the entire model.

Because $F1/F2$ analyses were conducted on a by-vowel basis, a table is included which shows the prosodic prominence and redundancy characteristics for each vowel type (see Sec. III A 2). In some cases a dependency exists between vowel identity and the variance of these factors. For example, for the /ɛ/ vowel (as in "set"), in our data 90% of all tokens were lexically stressed and 84% were open class. Such homogeneity must be taken into account when interpreting the relative contribution of prosodic and redundancy models to predicting the dependent variables by vowel.

III. RESULTS

A. Syllabic duration

1. Do the results for syllabic duration found for spontaneous speech apply to the citation speech in this database?

In Aylett and Turk (2004), the evidence supporting the smooth redundancy hypothesis was taken from a spontaneous spoken dialogue. It was by no means clear that the citation speech in the Rhetorical Corpus would show the same relationship. It could be argued that the natural patterns of redundancy would persist, despite the unnatural nature of the recording environment. Alternatively, significant differences might be expected between the duration results in the Rhetorical Corpus and those for a spontaneous speech corpus such as the Map Task Corpus. To investigate this question the analysis carried out on the Rhetorical Corpus was matched as closely as possible to the Map Task Corpus analysis reported in Aylett and Turk (2004).

Univariate ANOVAs testing effects of (1) language redundancy group; and (2) prominence group both showed highly significant effects (language redundancy group—high, medium, low, $F(2,313\,941)=67\,726$, $p<0.001$; prominence Group—no stress, lexical stress, phrasal stress, $F(2,313\,941)=72\,118$, $p<0.001$). Figure 2 shows the mean values of each group; all were significantly different from each other (*posthoc* t-tests with Bonferroni correction, all with $p<0.001$). Syllables were shorter for higher values of language redundancy and, conversely, syllables were longer for higher levels of prosodic prominence.

Linear regression analysis, with boundaries controlled, found that regression models using these factors predicted a significant proportion of the Rhetorical Corpus data ($r=0.7219, r^2=0.5211-52\%$ of the variance). The independent contribution of the redundancy and prosodic model and the shared predictive power of both models is shown in Fig. 3. Close to 20% of the predictive power was shared. These results were similar to the Map Task Corpus results reported in Aylett and Turk, 2004, where a comparable proportion of predictive power was shared. The results support the view that the smooth signal redundancy hypothesis holds for this citation speech corpus, and raise the possibility that other smooth signal redundancy effects might be found for this corpus, e.g., in the spectral domain. Because spectral effects were to be analyzed on a by-vowel basis, we analyzed the

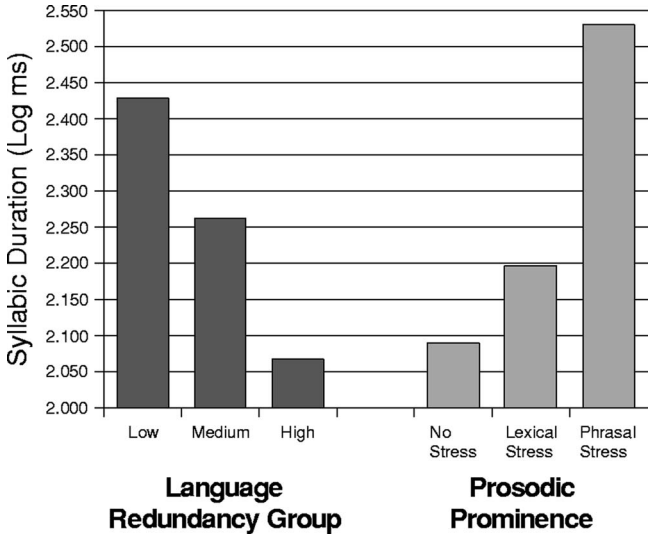


FIG. 2. Syllabic duration means (log ms), presented by redundancy group and by prosodic prominence, with prosodic boundaries controlled as specified in Sec. II D. All differences in means are significant. As a reference: 2.00 log ms=100 ms, 2.20 log ms=158.5 ms, 2.40 log ms=251.2 ms, 2.60 log ms=398.1 ms.

syllabic duration results on a by-vowel basis as well, in order to see if smooth signal redundancy effects were more reliable for some vowels as compared to others.

2. By-vowel durational evidence for smooth signal redundancy

Durational results analyzed separately for each vowel showed general support for the smooth signal redundancy hypothesis, both in terms of the direction of the language redundancy effects, and in terms of the shared predictive power of language redundancy and prosodic prominence. Figure 4 shows that vowels in the low language redundancy group were generally longer than vowels in the high redundancy group. However, there were differences across vowels in the relationship between the durations associated with the medium language redundancy group vs those associated with the high redundancy group. For example, for /i/ there was a smooth decrease across redundancy groups, but for /u/ the result was not as easy to interpret given that durations associated with the medium redundancy group were actually shorter than durations in the high redundancy group. All differences were significant at $p<0.001$ (*posthoc* t-test with Bonferroni correction).

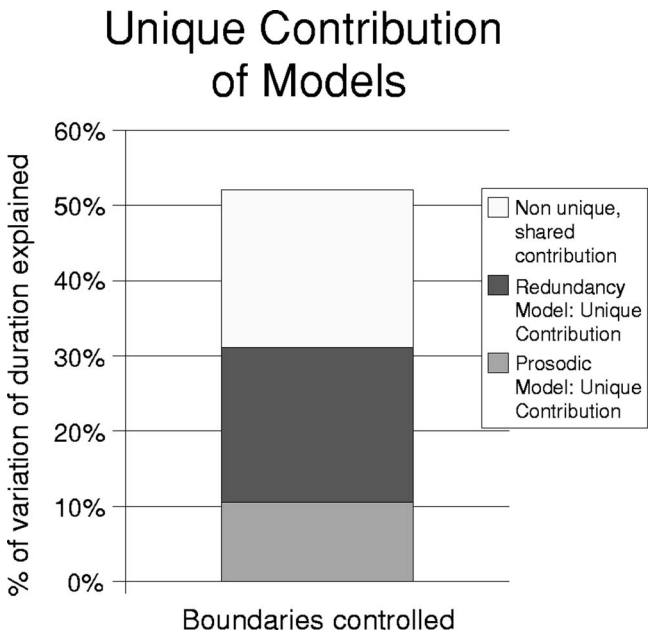


FIG. 3. Unique contributions of prosodic prominence and redundancy factors to the linear regression model.

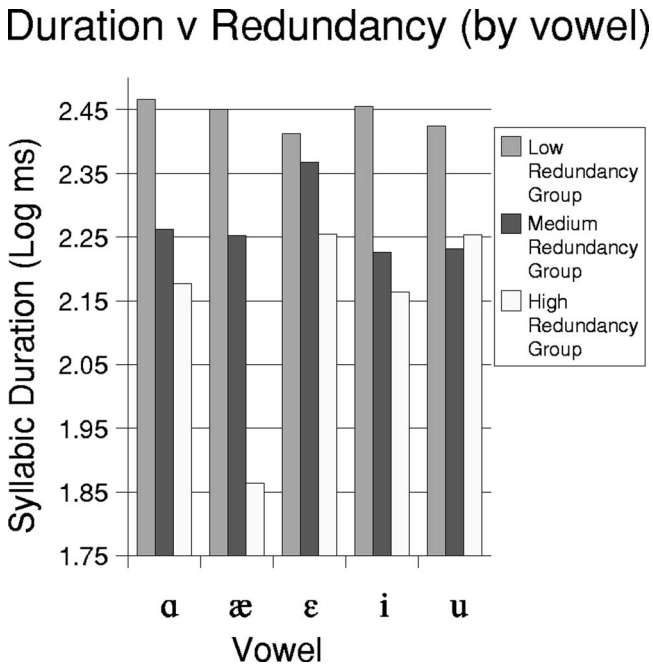


FIG. 4. Syllabic duration means (log ms), by redundancy group and by vowel. All differences in means are significant. As a reference: 2.00 log ms=100 ms, 2.20 log ms=158.5 ms, 2.40 log ms=251.2 ms, 2.60 log ms=398.1 ms.

Figure 5 shows how much of the variance in syllabic duration (log ms) for each vowel was predicted by full regression models containing both prosodic prominence and language redundancy factors. The combined regression/prosody model significantly predicted ($p<0.001$) between 18%–57% of the syllabic duration variation across vowels. Much of the predictive power was shared between language redundancy and prosodic prominence, although the extent of this shared contribution varied for different vowels. Some vowels, such as /i/, behave similarly to the pooled results, while others, for example /ɛ/, seem to exhibit a very different relationship between the prosodic model and the redundancy factors. One complication is that neither the prosodic factors nor the redundancy factors have similar variances across different vowels (see Table III). For example, /i/ is a common vowel which is often used in unstressed contexts (such as the last syllable of the word "spongy"), as well as in stressed syllables. In contrast, /ɛ/ has lower variance and a low mean in the wide redundancy category which suggests this vowel rarely if ever occurs in a very predictable syllable in any context. In addition, with boundaries controlled, nearly all the /ɛ/ tokens are stressed and in open-class words. This lack of variance available to both prominence and redundancy

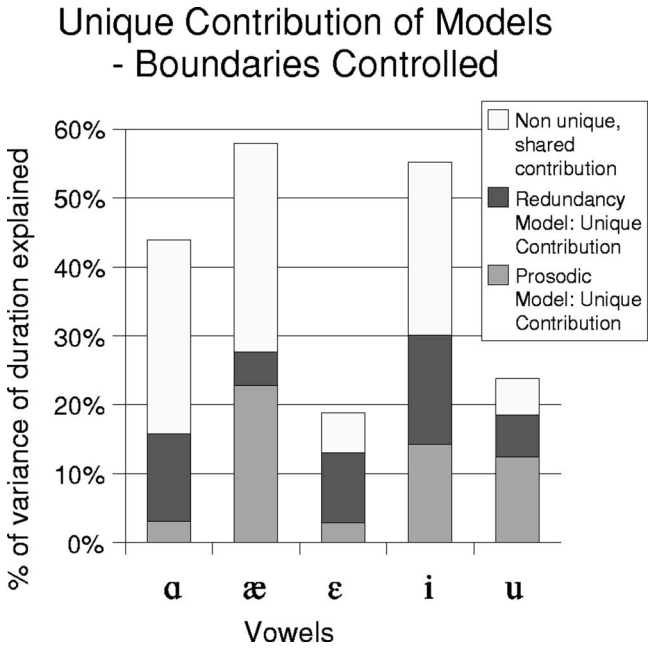


FIG. 5. Unique contributions of prosodic prominence and redundancy factors to linear regression models for each vowel.

models can thus explain the low overall performance of the models in the /ɛ/ context (less than 20% of total variance explained).

B. Spectral results

1. Spectral results: t-tests

Our hypothesis as far as $F1/F2$ effects are concerned was that vowels should show increased centralization with increased language redundancy, and that the direction of $F1/F2$ differences would be different for each vowel type (See Table IV for predicted effects for each analyzed vowel). Separate data for males vs females were analyzed for each vowel, with prosodic boundaries controlled (see Sec. II D 1).

Results show that language redundancy did indeed have a significant effect on $F1/F2$, but the significance and the direction of the effect varied by vowel. The extent to which our predictions were met can be seen in Table V, in which effects that clearly run counter to our predictions are shown in bold face. Overall, the results were reasonably consistent with a simple centralization hypothesis except for /u/, where $F1$ drops and /ɑ/ for female speakers, where $F2$ rises. Note that cases where no difference in formant value was predicted were not evaluated in this way, on the assumption that small differences could easily be accommodated by the centralization model.

TABLE IV. $F1/F2$ predictions for each vowel as acoustic redundancy drops due to centralization.

	$F1/F2$ Prediction from a centralization model of acoustic redundancy	
Vowel	$F1$ change	$F2$ change
ɑ	minus	plus
æ	minus	same
ɛ	minus	same
i	plus	minus
u	plus	plus

Figures 6 and 7 show these changes across the vowel space presented separately for male and female speakers. Groups 1 (high language redundancy) and 3 (low language redundancy) differed in a direction consistent with the hypothesis that language redundancy correlates with centralization, with /i/ showing a particularly good match to prediction. That is, average values for group 1 tended to be more central than average values for group 3. However, for many of the vowels, the difference between groups is reflected primarily in differences in $F1$, with little contribution from $F2$. Group 2, formed from a less homogeneous mixture of high/low and low/high items (see Sec. II B 2) did not show the same relationship with groups 1 and 3 across vowels.

2. Spectral results: Linear regression

Linear regression analysis showed that the redundancy and prosodic models could together predict between 0.5%–7.9% of variation in $F1$ ($p<0.001$), and 0.2%–5.9% of the variation in $F2$ ($p<0.001$), as shown in Table VI. The smaller percentage of explained variance for female speakers may be due in large part to poor performance of the formant tracker for these speakers. Given this concern about formant tracker reliability for the female speakers, relationships between prosodic and redundancy models were compared for male speakers only. Figure 8 shows that for many of the vowel formants, redundancy factors and prosodic prominence factors predicted a large proportion of shared variance. But, in contrast to results for syllable duration, this pattern did not hold for all cases: the shared contribution was quite small or nonexistent for e.g., /u/**–f1** and /æ/**–f2**. In addition, the unique contributions of redundancy and prosodic prominence factors seemed to vary considerably across vow-

TABLE III. Variation in prosodic prominence and language redundancy by vowel (prosodic boundaries controlled).

	Prominence		Wide redundancy		Narrow redundancy	
	Lexical stress	High prob phrasal stress	Mean	SD	Mean	SD
a	91%	63%	−0.58	0.98	−0.19	1.01
æ	95%	54%	−0.85	0.78	−0.24	0.87
ɛ	92%	76%	−0.53	0.89	−0.33	0.92
i	52%	37%	0.4	0.94	0.02	0.89
u	89%	54%	−0.03	0.83	0.46	0.98

TABLE V. Changes in $F1/F2$ mean values from low to high redundancy.

$F1/F2$ change from low to high redundancy prediction/result low redundancy group $F1/F2$ - high redundancy group $F1/F2$(*Hz*)		
Male Speakers (n=3)		
Vowel	$F1$ change	$F2$ change
ɑ	minus/−42.0	plus/+22.2
æ	minus/−21.0	same/+15.1
ɛ	minus/−72.2	same/−36.3
i	plus/ 8.4	minus/−48.8
u	**plus/−18.43**	plus/ 31.72
Female speakers (n=5)		
Vowel	$F1$ change	$F2$ change
ɑ	minus/−188.6	plus/−13.2
æ	minus/−44.1	same/ 37.65
ɛ	minus/−119.06	same/ 12.93
i	plus/ 8.76	minus/−37.0
u	**plus/−5.0**	plus/ 82.2

els: /u/**−f1** had a strong redundancy component orthogonal to the prosodic components of the model, whereas /i/ had a very low unique contribution from the redundancy model. However, in all cases except /ɛ/**−f2** and /u/**−f2**, both redundancy and prosodic prominence factors showed significant unique contributions to the regression model ($p<0.001$).

IV. CONCLUSIONS

The primary aim of this paper was to test whether the predictions of the probabilistic reduction hypothesis, speech efficiency, and the smooth signal redundancy hypothesis could be generalized to spectral dimensions. General support for these views was found in the form of significant effects of language redundancy on $F1/F2$ formant values: vowels with higher language redundancy and less prominence showed more centralization. Evidence for the smooth signal redundancy hypothesis in particular was found in terms of considerable overlap in effects of redundancy and prosodic prominence, supporting the view that prosodic prominence is a functional structure for achieving smooth signal redundancy and thus making the speech signal more robust in a noisy environment.

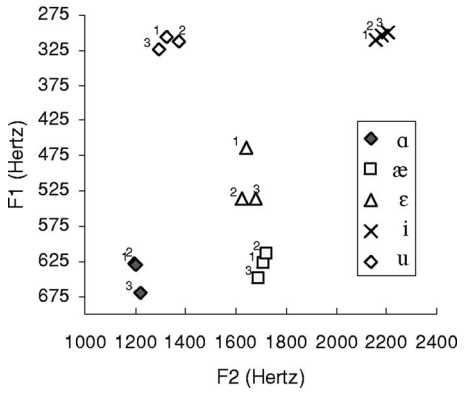


FIG. 6. Male spectral results by vowel (n=3). 1=high language redundancy, 2=medium language redundancy, 3=low language redundancy. All differences between 1 & 2, 2 & 3, 1 & 3 are significant ($p<0.005$) for both $F1$ and $F2$ (unless noted) on the basis of a *posthoc* t-test with Bonferroni correction. Nonsignificant group differences: /ɑ/ $F1$ 1 & 2, ɑ F2 1 & 2 & 3, /ɛ/ $F1$ 2 & 3, /ɛ/ $F2$ 1 & 2.

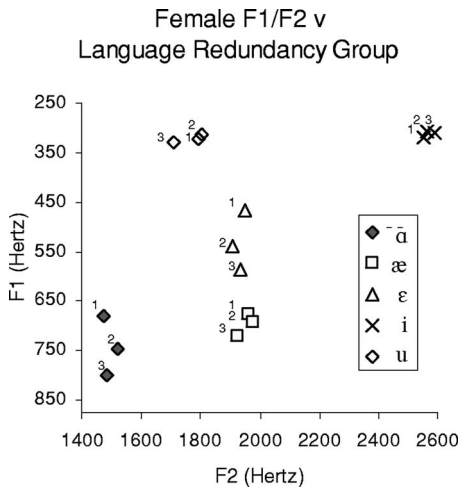


FIG. 7. Female spectral results by vowel (n=5). 1=high language redundancy, 2=medium language redundancy, 3=low language redundancy. All differences between 1 & 2, 2 & 3, 1 & 3 are significant ($p<0.005$) for both $F1$ and $F2$ (unless noted) on the basis of a *posthoc* t-test with Bonferroni correction. Nonsignificant group differences: æ/ $F1$ 1 & 2, /æ/ $F2$ 1 & 2, /i/ F1 2 & 3, /u/ $F2$ 1 & 2.

Nevertheless, these general effects were complicated by considerable by-vowel variation, some of which may be ascribed to inherent differences in the range of prosodic and language redundancy variation for these vowels in the lexicon. This was touched upon in Sec. III A 2, where differences in the redundancy profiles between /ɛ/ and /i/ in the Rhetorical Corpus were highlighted. However, limitations of our prosodic, redundancy, and acoustic models could also contribute to this variation between vowels. Indeed, many possible improvements could be made to the prosodic model. For example, the prosodic model used here does not include secondary lexical stress, degrees of variation in phrasal stress, accent type differentiation based on $f0$ contours, nor any interaction between prominence and prosodic boundaries. Refining the model along these lines might well lead to

TABLE VI. $F1/F2$ linear regression results for combined redundancy and prosodic prominence models, split by vowel and by speaker sex.

Male F1/F2	$F1$		$F2$	
	r	r^2	r	r^2
ɑ	0.261	0.068	0.113	0.013
æ	0.215	0.046	0.190	0.036
ɛ	0.281	0.079	0.081	0.007
i	0.214	0.046	0.243	0.059
u	0.228	0.052	0.205	0.042
Female F1/F2	$F1$		$F2$	
	r	r^2	r	r^2
ɑ	0.228	0.052	0.071	0.005
æ	0.071	0.005	0.189	0.036
ɛ	0.266	0.071	0.040	0.002
i	0.115	0.013	0.223	0.050
u	0.075	0.006	0.096	0.009

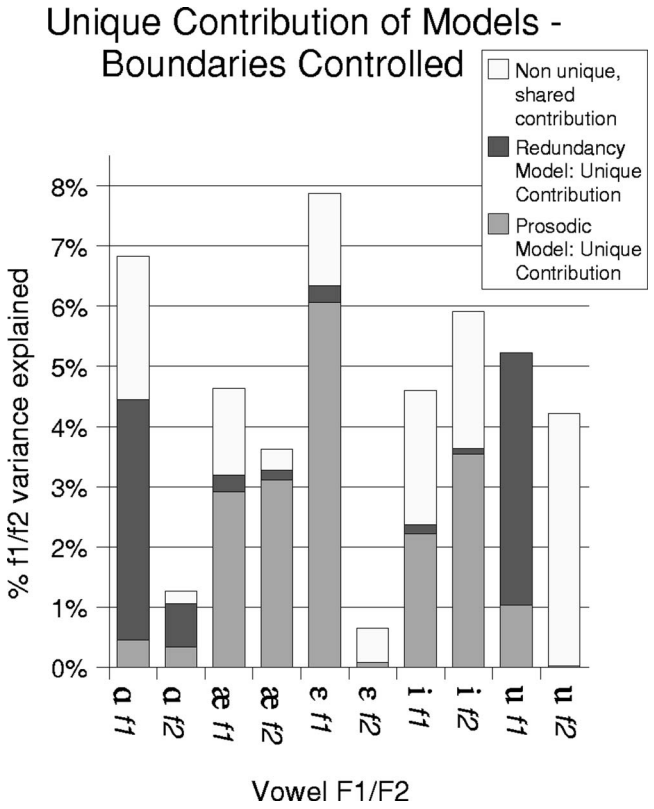


FIG. 8. The relationship between the predictive power of redundancy and prosodic prominence linear regression models, for male speakers only.

a greater overlap between the prosodic and the redundancy models. And, similarly, the redundancy model could be improved with additional specifications such as word frequency phone sequence likelihood, a neighborhood density metric, the probability of the syntactic category given a probabilistic syntax, and, at a discourse level, factors such as whether a referent has been mentioned previously. Using factor analysis as described in Sec. II B 2, any number of factors could be distilled to a smaller orthogonal set and added to the redundancy model.

Nevertheless, it may be unwise to complicate either the redundancy or the prosodic model until limitations in our acoustic modeling are addressed. In particular, syllables in the current study were represented by single $F1$ and $F2$ values taken at the temporal midpoint of each syllable in spite of the fact that formant trajectories are known to be dynamic, conditioned by the identity of surrounding consonants as well as the vowel. In fact, coarticulation with consonants, which may cause vowel centralization, may also add information concerning the surrounding segments. For example, nasalization in a vowel means it is easier to predict the subsequent segment is nasal, thus increasing acoustic redundancy. Such coarticulation, even while centralizing the middle portion of the vowel, can also add to acoustic redundancy by adding dynamic cues to the vowels' identity. Such cues at the onset and offset of a vowel have been shown to have a fundamental effect on speech perception (e.g., Strange, 1989).

In addition, the characterization of the vowel space we have used could also be improved. $F1/F2$ values could be modified by different types of vowel space normalization. There is a strong argument for the use of a perceptual frequency scaling metric, such as bark or jnd's, which we avoided in this paper to aid transparency. Finally, a more complex model than centralization may do much better at modeling this change in acoustic redundancy, for example, one that takes into account vowel distinctiveness at different points in the vowel space.

Finally, it is far from clear which underlying perceptual model is most appropriate for building a model of acoustic redundancy. Unlike duration alone, where longer segments are generally agreed to be more intelligible, little is known about how factors such as $F1/F2$, amplitude, $f0$ change, and spectral tilt combine both with dynamic and static cues to increase or decrease intelligibility. Investigating these factors and their relationship to perception is an ongoing task, but a language redundancy framework can clearly offer a solid quantitative approach for their examination.

Adank, P. (**2003**). "Vowel Normalization: A perceptual study of Dutch Vowels" Ph.D. thesis, University of Nijmegan.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty-Sneddon, G. M., Garrod, S., Isard, S., Kowtko, J. C., McAllister, J. M., Miller, J. E., Sotillo, C. F., Thompson, H. S., and Weinert, R. (**1991**). "The HCRC Map Task Corpus," Lang Speech **34**(4), 351–366.

Aylett, M. P. (**2000**). "Stochastic Suprasegmentals: Relationships between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech" (http://www.cogsci.ed.ac.uk/matthewa/thesis_sum.html). Ph.D. thesis, University of Edinburgh.

Aylett, M. (**2001**). "Modeling care of articulation with hmms is dangerous," in Proceedings Eurospeech 01, pp. 1491–1494.

Aylett, M., and Turk, A. (**2004**). "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech," Lang Speech **47**, 31–56.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., and Gildea, D. (**2003**). "Effects of disfluencies, predictability, and utterance position on word form variation in english conversation," J. Acoust. Soc. Am. **113**, 1001–1024.

Bolinger, D. (**1963**). "Length, vowel, juncture," Linguistics **1**, 5–29.

Bond, Z., and Moore, T. (**1994**). "A note on the acoustic-phonetic characteristics of inadvertently clear speech," Speech Commun. **14**, 325–337.

Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (**1996**). "Intelligibility of normal speech I. Global and fine-grained acoustic-phonetic talker characteristics," Speech Commun. **20**, 255–272.

Bybee, J. (**2000**). "Lexicalization of sound change and alternating environments," in *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, edited by M. Broe and J. Pierrehumbert (Cambridge University Press, Cambridge).

Clarkson, P., and Rosenfeld, R. (**1997**). "Statistical language modeling using the CMU-Cambridge toolkit," in Proceedings of Eurospeech 97, pp. 2707–2710.

Fowler, C. A., and Housum, J. (**1987**). "Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction," J. Mem. Lang. **26**, 489–504.

Godfrey, J., Holliman, E., and McDaniel, J. (**1992**). "SWITCHBOARD: Telephone speech corpus for research and development," in Proceedings of ICASSP-92 San Francisco, pp. 517–520.

Jurafsky, D., Bell, A., and Girand, C. (**2003**). "The role of the lemma in form variation," in *Laboratory Phonology VII*, edited by N. Warner and C. Gussenhoven (Mouton de Gruyter, Berlin).

Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. (**2001**). "Probabilistic relations between words: Evidence from reduction in lexical production," in *Frequency and the Emergence of Linguistic Structure*, edited by J. Bybee and P. Hopper (Benjamins, Amsterdam), pp. 229–254.

Lieberman, P. (**1963**). "Some effects of semantic and grammatical context on the production and perception of speech," Lang Speech **6**, 172–187.

Lindblom, B. (**1990**). "Explaining phonetic variation: a sketch of the H & H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hard-

castle and A. Marchal (Kluwer Academic, Dordrecht, The Netherlands), pp. 403–439.
Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.
Munson, B., and Soloman, N. (**2004**). "The influence of phonological neighborhood density on vowel articulation," J. Speech Lang. Hear. Res. **47**, 1048–1058.
Nearey, T. (**1992**). "Applications of generalized linear modeling to vowel data," in Proceedings ICSLP 92, pp. 583–587.
Neter, J., Wasserman, W., and Kutner, M. H. (**1990**). *Applied Linear Statistical Models: Regression, Analysis of Variance and Experimental Design*, 3rd ed. (Irwin, Boston).
Picheny, M., Durlach, N., and Braida, L. (**1986**). "Speaking clearly for the hard of hearing. II Acoustic characteristics of clear and conversational speech," J. Speech Hear. Res. **29**, 434–445.
Pierce, J. R. (**1961**). *Symbols, Signals and Noise: The Nature and Process of Communication* (Harper, New York).
Rosner, B., and Pickering, J. (**1994**). *Vowel Perception and Production*. (Oxford Science Publications, Oxford).
Sharp, A. E. (**1960**). "The analysis of stress and juncture in English," *Transactions of the Philological Society*, pp. 104–135.
Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.
Summers, W. V. (**1987**). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," J. Acoust. Soc. Am. **82**, 847–863.
Talkin, D. (**1987**). "Speech formant trajectory estimation using dynamic programming with modulated transition costs," J. Acoust. Soc. Am. **82**, Suppl. 1, S55.
van Bergem, D. R. (**1993**). "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," Speech Commun. **12**, 1–23.
van Son, R. J. J. H., and Pols, L. C. W. (**1999**). "An acoustic profile of consonant reduction," Speech Commun. **28**, 125–140.
van Son, R. J. J. H., and Pols, L. C. W. (**2003**). "How efficient is speech?," Institute of Phonetic Sciences, University of Amsterdam, Proceedings, **25**, pp. 171–184.
Wells, J. C. (**1982**). *Accents of English* (Cambridge University Press, Cambridge), Vol. **3**.
Wright, R. (**2003**). "Factors of lexical competition in vowel articulation," in *Papers in Laboratory Phonology VI: Phonetic Interpretation*, edited by J. Local, R. Ogden, and R. Temple (Cambridge University Press, Cambridge).
Young, S., Jansen, J., Odell, J., Ollason, D., and Woodland, P. (**1996**). *The HTK Book*, Entropic. Version 2.00.
Zipf, G. K. (**1949**). *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Reading, MA).

Cue weighting in auditory categorization: Implications for first and second language acquisition[a)]

Lori L. Holt[b)]
Department of Psychology and the Center for the Neural Basis of Cognition, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, Pennsylvania 15213

Andrew J. Lotto
Center for Perceptual Systems University of Texas at Austin, Austin, Texas 78712

(Received 22 April 2005; revised 14 February 2006; accepted 28 February 2006)

The ability to integrate and weight information across dimensions is central to perception and is particularly important for speech categorization. The present experiments investigate cue weighting by training participants to categorize sounds drawn from a two-dimensional acoustic space defined by the center frequency (CF) and modulation frequency (MF) of frequency-modulated sine waves. These dimensions were psychophysically matched to be equally discriminable and, in the first experiment, were equally informative for accurate categorization. Nevertheless, listeners' category responses reflected a bias for use of CF. This bias remained even when the informativeness of CF was decreased by shifting distributions to create more overlap in CF. A reversal of weighting (MF over CF) was obtained when distribution variance was increased for CF. These results demonstrate that even when equally informative and discriminable, acoustic cues are not necessarily equally weighted in categorization; listeners exhibit biases when integrating multiple acoustic dimensions. Moreover, changes in weighting strategies can be affected by changes in input distribution parameters. This methodology provides potential insights into acquisition of speech sound categories, particularly second language categories. One implication is that ineffective cue weighting strategies for phonetic categories may be alleviated by manipulating variance of uninformative dimensions in training stimuli. [DOI: 10.1121/1.2188377]

PACS number(s): 43.71.An, 43.66.Ba, 43.71.Hw [ALF] Pages: 3059–3071

I. INTRODUCTION

Outside of the acoustic researcher's laboratory, few sound categories are distinguished by a single acoustic dimension. In the natural world, auditory categories more typically are defined along multiple dimensions. Integration of information across acoustic dimensions, then, must be a central characteristic of auditory processing.

Speech categories provide an excellent illustration of both the complexity with which multiple acoustic dimensions define auditory categories and of the adeptness of the auditory system at integrating information across multiple dimensions. For speech categories, rarely is a single acoustic dimension necessary or sufficient to define category membership; this is the essence of the classic "lack of invariance" issue (Liberman, 1996; Liberman *et al.*, 1967). For example, there are many acoustic dimensions that contribute to voicing, as in the difference between English /ba/ and /pa/; Lisker (1986) cataloged as many as 16 acoustic dimensions that may characterize the English voicing distinction.

Nevertheless, for speech and other auditory categories the existence of multiple acoustic dimensions does not imply their perceptual equivalence. Simply put, some acoustic dimensions play a greater role in determining the perceptual identity of a sound than do others. The fact that acoustic dimensions need not contribute equivalently to category identity has been referred to as *cue weighting*. Acoustic dimensions appear to be perceptually weighted in the sense that some are strongly correlated to categorization responses whereas others, although present, weakly determine perceived category membership. As an example, both spectral and temporal acoustic cues differentiate English tense and lax vowels like /i/ versus /I/. Adult native American-English listeners, however, rely much more on the spectral dimension (formant frequency) than the temporal dimension (vowel duration) in categorizing /i/ and /I/ (e.g., Hillenbrand *et al.*, 2000). Likewise, Francis and colleagues (2000) have demonstrated that although burst cues and formant transition cues co-vary for American English stop consonant categories, listeners rely on the formant cue significantly more than the burst cue in categorization responses (see also, Walley and Carrell, 1983). Even with explicit training with feedback to use the burst cue, the listeners continue to rely on the formant transition (Francis *et al.*, 2000).

What determines the relative weighting of different sources of acoustic information? An adaptive listener would weight dimensions based on experience over time with the acoustic environment. For example, we might be able to pre-

[a)]Portions of this work were presented at the 148th Meeting of the Acoustical Society of America, San Diego, CA, November, 2004 and at the 28th Mid-Winter Meeting of the Association for Research in Otolaryngology, New Orleans, LA, February, 2005.
[b)]Electronic mail: lholt@andrew.cmu.edu

dict weighting functions for speech perception if we knew how acoustic dimensions co-varied with phonetic contrasts in a listener's experience. In principle, the regularity of information along different dimensions could be fully characterized if it were possible to make exhaustive acoustic measurements of the full range of speech productions for a language, an idiolect, an accent, a speaker, or group of speakers. In practice, this has been approached much more modestly by examining regularities in the realization of speech productions of small samples of speakers (e.g., Lisker and Abramson, 1964; Lotto *et al.*, 2004). Nevertheless, even with this more limited approach, it is possible to estimate the distributions of phonetic categories across a subset of acoustic dimensions. Adaptive listeners will tune their weighting of the dimensions based on the characteristics of these distributions to maximize accuracy. However, there are a number of constraints related to sensory processing, cognitive processing, and previous experience that may prevent listeners from achieving ideal performance on any particular categorization task.

Examining this at a finer grain, cue weighting can be thought of as a function of at least four variables. The first two relate to the distributional characteristics of acoustic information and are language specific. First, acoustic dimensions vary in their *informativeness* for category identity; that is, there is a difference in the distinctiveness of the distributions for competing categories along each dimension. For example, voice onset time (VOT) is a robust acoustic signal for voicing across languages. Lisker and Abramson and others (Keating, 1984; Lisker and Abramson, 1964) have examined the distribution of VOT across speech categories in several languages. In American English, for example, VOT values produced by speakers are quite reliably differentiated across speech categories. This is to say that the distributions of the voicing categories do not overlap much along the VOT dimension. As a result, VOT is highly informative for the American English voicing distinction. Because of this informativeness, VOT is likely to be a heavily weighted cue in the perception of voicing categories by American English listeners. Note, as well, that implicit in this description is the hypothesis that listeners are sensitive, not just to the absolute value on an acoustic dimension for a particular stimulus presentation, but also to the distributional characteristics of that cue as it occurs across speech instances within a language.

A second parameter of auditory distributions that could affect perceptual cue weighting is distribution *variance*. Information is carried by variance in the signal and the auditory system seems to be especially sensitive to dimensions that are varying. Studies of nonspeech pattern analysis have established that components with greater relative variance receive greater perceptual weight (Lutfi, 1993; Lutfi and Doherty, 1994). From these findings, one may predict that dimensions on which speech distributions vary the most will receive higher perceptual weights. But in categorization tasks, such as phonetic categorization, the relationship between variance and cue weighting is likely to be more complex. The relationship between within-category and between-category variance is probably an important determinant of perceptual weight. Whereas large overall variance (disregarding category) is indicative of an informative acoustic dimension, large within-category variance can decrease the informativeness of a dimension by creating distribution overlap (in the same way that variance can decrease d' in signal-detection theory). In addition, increased within-category variance across a dimension could mean that the dimension is not robustly related to the category or that there is a great deal of noise in the transmission of this dimension. In either case, the reliability of the dimension is questionable and, correspondingly, the dimension may be weighted less.

It is important to note that each of these distributional aspects of cue weighting must be considered to be language specific, dialect specific, and perhaps even speaker specific. Although the same acoustic dimensions may exist across languages or dialects, the phonetic distributions are clearly not equivalent. For example, across the voiced-voiceless distinction in American English stops, aspiration is correlated with voicing. Voiceless consonants (with longer VOTs) are typically aspirated whereas voiced consonants are much less likely to be produced with aspiration (e.g., Stevens, 1998). Aspiration is thus likely to be a relatively strong cue to voicing in American English. Aspiration also exists for Hindi stop consonant categories, but it does not carry weight as a cue for voicing because it is strongly correlated with an orthogonal category dimension (aspirated versus unaspirated, Benguerel and Bhatia, 1980). As another example, consider the relationship of vowel length to syllable-final consonant voicing. Adult native listeners of languages that do not have syllable-final obstruents or that fail to make a vowel-length distinction as a function of the voicing of the final consonant do not weight vowel duration as a cue to the final-consonant's category identity as much as listeners of languages that make vowel-length distinctions as a function of final-consonant voicing (Crowther and Mann, 1992, 1994; Flege and Wang, 1989). Listeners' weighting of acoustic dimensions, then, arises from experience with the regularities of acoustic realization in the native patterns of speech production, although the learning mechanisms remain, as yet, relatively poorly understood. At a somewhat finer level, perceptual weighting may have speaker-specific characteristics. Weighting of acoustic dimensions across individual speakers may vary somewhat and listeners' familiarity with this variation may contribute to voice recognition.

Two other variables that can influence cue weighting can be thought to be non-distributional and not language specific or speaker specific. First, the basic auditory representation of an acoustic dimension can influence its perceptual weight. Some acoustic dimensions are more robustly encoded by the auditory system than others. For example, there may be discontinuities in auditory processing along particular acoustic dimensions such that equal physical steps do not produce equivalent changes in percept (e.g., Kuhl and Miller, 1975; Stevens, 1989). VOT, again, serves as a fine example because its perception is thought to be influenced by a general auditory discontinuity in processing the onsets of simultaneous acoustic events (Kuhl and Miller, 1975; Pisoni, 1977). The acoustic dimension distinguishing VOT can be thought of as a relative onset asynchrony between two component frequencies, one higher and one lower in frequency. The

lower-frequency component may either lead or lag the higher-frequency component, or it may occur simultaneously with it. Research from human adult, infant, and animal behavior as well as electrophysiology from these populations has suggested that the mammalian auditory system is poor at resolving onset asynchrony differences less than about 20 ms, whether they are speech or nonspeech stimuli (e.g., Jusczyk *et al.*, 1980; Kuhl and Miller, 1975; Pisoni, 1977; Simos and Molfese, 1997; Simos *et al.*, 1998: Sinex *et al.*, 1991; Steinschneider *et al.*, 1999). Sinex and colleagues, for example, relate this auditory discontinuity to changes in the variance of the neural representation at the auditory nerve; there appears to be a relatively more robust neural code in the representation of stimuli with onset asynchronies greater than about 20 ms (Sinex and McDonald, 1989; Sinex *et al.*, 1991). It has been suggested that languages may exploit this discontinuity in auditory processing by developing voicing categories that straddle this sensitivity and thus exaggerate the perceived difference between categories (e.g., Pisoni, 1977). Recent research has indicated that in addition to the perceptual benefit of exaggerated discriminability, such placement may also facilitate category learning (Holt *et al.*, 2004). By these means, the relatively quantal representation for VOT (and consequent high discriminability of the category distinction) may contribute to a stronger cue weight. Of course, the opposite circumstances may also influence perceptual weighting; acoustic dimensions may be less heavily weighted as a consequence of a weak auditory representation, for example, if they are more susceptible to noise or are represented with more variability by the auditory system than other acoustic dimensions. Moreover, if the auditory system does not have sufficient resolving power or if there is too much noise in the representation of an acoustic dimension then the covariance between the dimension and the category label may be underestimated by the perceiver. In general, more discriminable acoustic dimensions will tend to be more informative for a listener and, correspondingly, should be expected to have a greater perceptual weight.

Cue weighting may also be influenced by task. The acoustic dimensions heavily weighted for phonetic categorization may be much less informative in making a lexical decision, identifying a talker, evaluating the emotional content of speech, or making other perceptual decisions. The same acoustic dimension present in the same signal may be heavily weighted for one perceptual task, but less heavily weighted for a different task. Likewise, changes in the task that affect the auditory representation (for example, the addition of low-frequency noise that selectively masks a particular acoustic dimension) or the informativeness of a cue (for example, manipulation of the distribution characteristics on an acoustic dimension in the laboratory) would also be expected to bring about shifts in listeners' perceptual cue weighting. This expectation of adaptive plasticity is particularly important for our understanding of how to alter cue weighting in aid of acquiring new categories, for example in the acquisition of second-language (L2) phonetic categories. The present experiments exploit these possibilities.

To summarize the above, not all available information is equivalent in perceptual processing and the differential contribution of acoustic dimensions to categorization arises as a result of experience with regularities in the input, the robustness or variability of the perceptual coding of an acoustic dimension, and its informativeness to category identity as a function of task. Cue weighting is a quantitative description of how auditory information (and information from other modalities as well) is integrated in perceptual categorization. An appreciation of perceptual cue weighting leads to the perspective that speech categorization is not just a matter of detecting available auditory cues along various acoustic dimensions, but also applying some weighting function that is, at least in part, dependent on experience with phonetic distributions.

This hypothesis has important implications for language acquisition. There has been considerable interest in the perceptual difficulties of L2 learners in acquiring second-language speech categories. One underlying cause of these difficulties may be that well-established native-language weighting functions are inappropriate for the L2. The language-specific nature of cue weighting may create perceptual difficulties in integrating acoustic information for nonnative speech categories. Understanding the mechanisms of cue weighting and investigating the means by which it may be possible to change these functions are therefore of considerable importance for understanding L2 category acquisition. As an example, take the classic case of English syllable-initial /l/ versus /r/ perception by native Japanese listeners (Miyawaki *et al.*, 1975). For native American English listeners, the starting frequency of the third formant (F3) carries most of the perceptual weight, but Japanese listeners appear to heavily weight F2 starting frequency in categorizing English /l/ and /r/ (Iverson *et al.*, 2003; Yamada and Tohkura, 1990). This weighting is nonoptimal as F2 is not varied contrastively for syllable-initial liquids by English speakers (Lotto *et al.*, 2004). An understanding of the basic processes by which acoustic dimensions are perceptually weighted and determination of the means by which these weights may be altered therefore are likely to have straightforward applications to understanding this classic L2 learning problem, and others.

Cue weighting also appears to be important for a full understanding of first language (L1) acquisition. Although infants begin to become perceptually tuned to the characteristics of L1 already in the first year (Jusczyk, 1997; Kuhl *et al.*, 1992; Werker and Tees, 1984), this process ultimately takes significant time to develop to adultlike perception (e.g., Morrongiello *et al.*, 1984; Nittrouer *et al.*, 1998; Nittrouer *et al.*, 2000; Nittrouer, 2004; Parnell and Amerman, 1978). Specifically, children appear to apply different weightings to acoustic dimensions in perceiving L1 speech categories (Hazan and Barrett, 2000; Mayo and Turk, 2004; Nittrouer, 2004). For example, in languages in which adults weight vowel duration as a strong cue in categorization of following final stop consonants, children tend to favor the consonant-influenced vowel-formant transitions over vowel duration (Nittrouer, 2004).

The purpose of the present experiments is to address issues of cue weighting in an effort to begin to understand the means by which the auditory system weights acoustic

dimensions and, also, to investigate methods by which listeners' cue weighting might be shifted. This latter aim is of particular importance in understanding the link between cue weighting and speech category acquisition in L1 and L2. It is presumed that cue weighting functions are determined in large part by the characteristics of distributions of experienced sounds. However, uncovering the role of experience in shaping speech categorization is quite difficult because it is impossible to know the precise history of speech experience that a listener brings to the laboratory. Therefore, as a starting point, we have taken the approach of investigating novel nonspeech auditory category learning. This allows us to maintain total control over listeners' histories of experience with the acoustic exemplars, thus ensuring that we have a precise characterization of the distributional characteristics. We can then describe the mapping between the informativeness of each dimension and the correlation of that dimension with categorization responses (i.e., its weighting). In the experiments that follow, we examine how changing the characteristics of these acoustic distributions or the manner of presentation can affect weighting functions. In each experiment, listeners are trained with feedback to categorize nonspeech sounds as coming from one of two predefined distributions. These microgenetic experiments were designed to examine auditory category formation with application to phonetic category acquisition. In fact, the characteristics of the distributions that were used were very similar to previous studies examining speech sound categorization in human adults, infants and non-human animals (Grieser and Kuhl, 1989; Kluender *et al.*, 1998; Kuhl, 1991).

II. EXPERIMENT 1

In Experiment 1, listeners were trained to categorize unfamiliar sounds drawn from two distributions situated in a two-dimensional acoustic space. The two acoustic dimensions were equated for perceptual discriminability in pilot psychophysical experiments and possessed no known *a priori* relationship. Moreover, the input distributions were created so that the two dimensions were equally informative for the categorization task. The questions under investigation were whether listeners would integrate information from both dimensions in making category decisions and, if so, whether the resulting cue weighting functions were predictable from the (in this case equal) informativeness of the dimensions.

A. Methods

1. Participants

Fourteen volunteers recruited from Carnegie Mellon University, Pittsburgh, PA, participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The experiment used methods from auditory category-learning experiments (e.g., Holt *et al.*, 2004; Mirman *et al.*, 2004) whereby distributions of sounds drawn from a two-dimensional acoustic space were sampled in stimulus presentation. Listeners labeled the sound and received feedback to learn to assign category labels during training. In a subsequent test, novel stimuli drawn from the same two-dimensional acoustic space were presented to assess the relative weighting of the two acoustic dimensions as a consequence of categorization training. This design allows response patterns to be directly compared to distribution characteristics (e.g., means, variances, overlap).

The nonspeech stimuli used in all four experiments were frequency-modulated tones. The two-dimensional acoustic space from which stimuli were sampled was defined by center, or carrier, frequency (CF) and modulation frequency (MF). Within this acoustic space, two distributions corresponding to the to-be-learned auditory categories were created; they are illustrated in Fig. 1(a) as open symbols. Forty-eight unique stimulus exemplars defined each of the categories; each category exemplar is represented by an open symbol in Fig. 1(a), with circles marking one category and diamonds the other. Each stimulus was created from a sine wave tone with a particular CF modulated with a depth of 100 Hz at the corresponding MF. For example, if the CF was 760 Hz and the MF was 203 Hz, the tone was modulated from 710 to 810 Hz at a rate of 203 Hz. Each stimulus was 300 ms long and was sampled at 10 kHz with 16-bit resolution. Stimuli were created using the Cool Edit (Syntrillium, Inc., Phoenix, AZ) software.

It was desirable that sampling of the distributions along the two acoustic dimensions produce perceptual changes that were roughly equivalent. To this end, estimates of the just-noticeable-difference along the CF and MF dimensions were determined by pilot psychophysical studies in which one dimension (CF or MF) was varied while the other was held constant at an intermediate value. Based on informal listening, these tests began with minimum steps of 15 Hz along the CF dimension and 3 Hz along the MF dimension. Young, normal-hearing listeners ($N=20$) recruited from the same population as those participating in the reported categorization studies responded to pairs of stimuli varying in either CF or MF in an AX discrimination test with stimulus pairs separated by 1, 2, 3, 4, or 5 steps. This series of pilot discrimination studies revealed that the CF and MF dimensions were approximately equally discriminable (70% accuracy) when step size was adjusted to be 30 Hz along the CF dimension and 18 Hz along the MF dimension. The results of these studies also confirmed that discrimination was flat across each of the dimensions, indicating that there were no auditory discontinuities in discriminability across the region of the two-dimensional stimulus space under observation.

For Experiment 1, the positioning of the input distributions in the acoustic space resulted in the CF and MF acoustic cues being equally informative for the categorization task. Distribution overlap was equal on both dimensions. One way to quantify "informativeness" is to calculate the relative increase in accuracy over chance performance that an ideal observer could attain by using the dimension. For the current study, we used a simple criterion-bound model in which the ideal observer would place a criterion at the optimal (in terms of accuracy) position along the dimension. All stimuli on one side of the criterion are designated as category "A" and all on the other side are designated category "B." Any

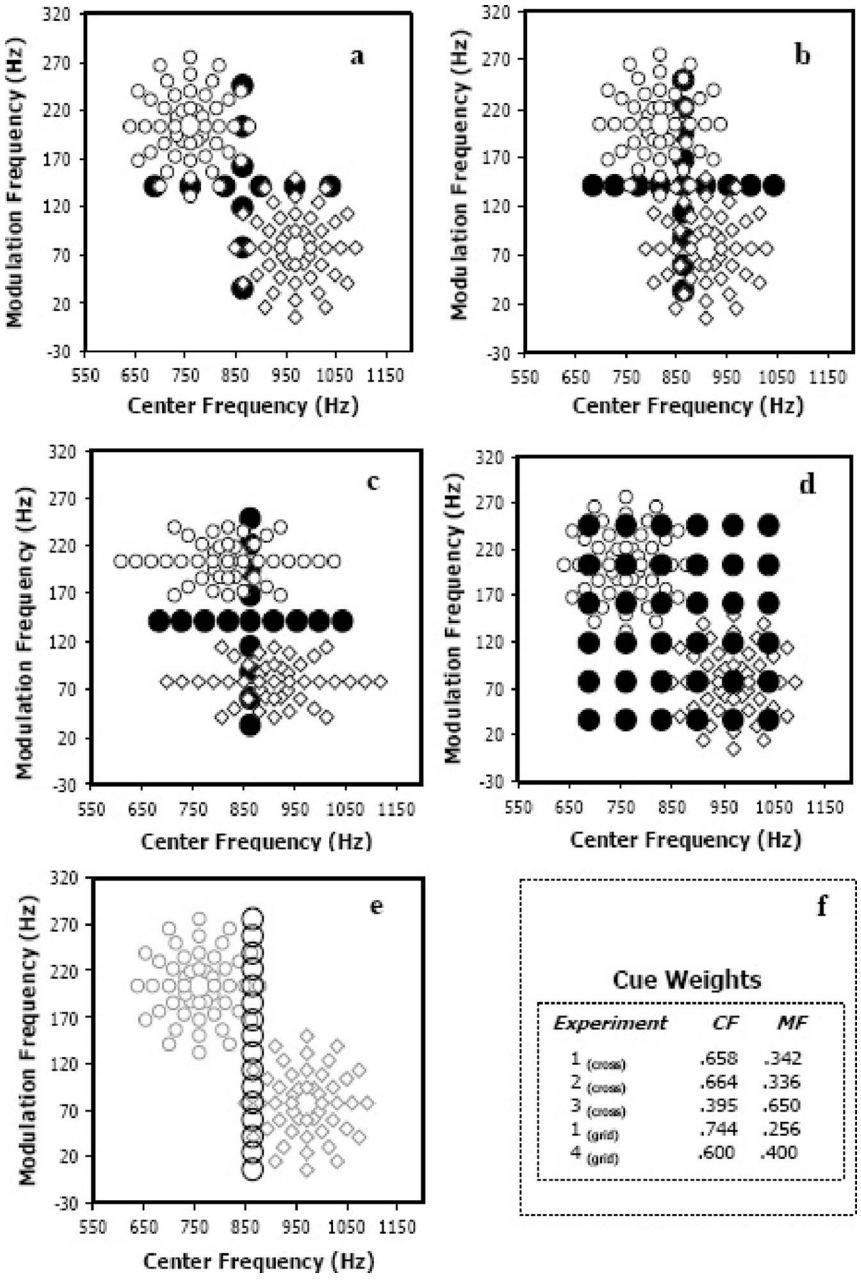


FIG. 1. Schematic representations of the stimulus input distributions in the center frequency (CF) by modulation frequency (MF) acoustic space. Open circles and open diamonds illustrate the individual stimuli from the two categories, respectively. The large filled circles show novel stimuli withheld from training from which cue weights were calculated. (a) Stimulus distributions for Experiment 1. (b) Stimulus distributions for Experiment 2. (c) Stimulus distributions for Experiment 3. (d) Stimulus distributions for Experiment 4. (e) Training distributions for Experiment 4 are shown as small light symbols for reference; they are identical to those of Experiment 1. The large open circles illustrate the stimuli presented during passive listening prior to categorization training. (f) Summary of the relative cue weights for the CF and MF dimensions across experiments.

stimuli landing on the criterion are considered "A" 50% of the time. Unlike more sophisticated ideal observer models, there is no noise in the encoding of the value on the dimension and the response is not probabilistic. With this decision model, one can calculate the informativeness of a dimension as

$$I_{\text{dimension}} = (\%\,\text{Correct Ideal Observer} - \text{Chance})/(\text{Perfect Performance} - \text{Chance}). \quad (1)$$

Equation (1) gives informativeness as the increase in accuracy over chance due to information in a dimension relative to the total amount of possible improvement. This function is equal to 1 when perfect performance can be achieved using that dimension and 0 when the dimension provides no information above chance. For Experiment 1, there are two possible responses and the prior probabilities of each response are equal; therefore, perfect performance is 100% and chance is 50%. By using either acoustic dimension alone, a perceiver could attain 95.8% correct with an optimal linear boundary [as can be seen in Fig. 1(a), the two categories are nearly perfectly linearly separable along either acoustic dimension with optimal boundaries at CF=865 Hz and MF =140 Hz]. Since the use of either acoustic dimension yields the same optimal percent correct, these acoustic dimensions are equally informative to the categorization task $I_{\text{CF}}=I_{\text{MF}}(95.8-50/100-50)=0.916$. However, integrating information across the two acoustic dimensions is more informative to the categorization response than either dimension individually; the categorization task can be performed perfectly if an ideal observer integrates across both acoustic dimensions (i.e., $I_{\text{CF+MF}}=1.0$).

In addition to the 96 stimulus exemplars (open symbols) comprising the categorization training stimuli, 12 novel test

stimuli (filled circles) were created to probe listeners' use of the CF and MF dimensions following training. These novel stimuli were reserved from categorization training to be used in generalization tests. They varied orthogonally in the acoustic space such that the value along one acoustic dimension was held constant at an intermediate value (CF = 865 Hz or MF = 140 Hz) while the values along the other acoustic dimension varied.[1]

Acoustic presentation was controlled by TDT System II hardware (Tucker-Davis Technologies, Alachua, FL). Stimuli were converted from digital to analog, low-pass filtered at 4.8 kHz, amplified and presented diotically over linear headphones (Beyer DT-150) at approximately 70 dB SPL(A).

3. Procedure

a. Categorization training. Listeners were tested individually in sound-attenuating booths. On each trial, listeners heard a single stimulus, pressed one of two unlabeled buttons to record a categorization response and received feedback as to the correct response via illumination of a light above the correct response button.

Listeners were instructed that the task was to determine whether each sound belonged to the left-most or the right-most response button. The experimenter also informed listeners that, although they would need-to guess at first, they should attempt to use the feedback given on each trial to guide later responses. Participants were encouraged to attempt to get as many responses correct as possible. Feedback assignment followed the symbols of Fig. 1; stimuli represented as open circles were assigned to one category (one response button) whereas those illustrated as open diamonds were assigned to the opposite category (and button). Left versus right button assignment to stimulus distributions was counterbalanced across participants.

Each block of training included a single presentation of each of the 96 training stimuli. There were ten blocks of training overall and the order of stimulus presentation was randomized within each block. After each block, listeners were able to take a brief break.

b. Test. Following training, listeners completed five more blocks of the categorization task. Within a block, stimulus presentation was randomized. These blocks were identical to the previous training blocks, except that in addition to the original 96 training stimuli, novel generalization stimuli were introduced. On these trials, listeners did not receive informative feedback; lights above all of the buttons lit following the response. This approach was chosen so that listeners' categorization of the generalization stimuli could be observed without encouraging participants to learn labels for the generalization stimuli across the course of testing. Listeners continued to receive feedback on trials for which familiar training stimuli were presented to assure that they continued to assign category labels in a manner that respected the distribution characteristics and feedback assignments present in training. This model of feedback assignment and mixing of novel and familiar stimuli is identical to paradigms successfully used in studying nonhuman animal category learning (see, e.g., Kluender *et al.*, 1998). Continuing to provide feedback for familiar stimuli, but not for novel stimuli, provides a safeguard against responses that drift away from learned categories as a consequence of the distribution differences that novel generalization stimuli introduce in the input. It also encourages participants to remain engaged in the task.

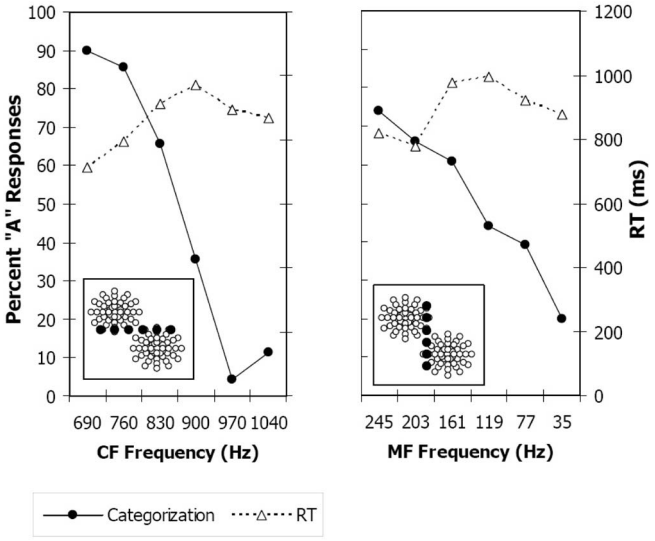


FIG. 2. Experiment 1 categorization responses and reaction time for generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

Participants completed both training and testing in a single experimental session lasting approximately 2 h.

B. Results and conclusions

The percent of "A" categorization responses was calculated whereby category "A" was arbitrarily assigned to be the category in the upper left corner of the two-dimensional acoustic space, indicated by open circles in Fig. 1(a). Overall, listeners performed fairly well in learning to categorize these unfamiliar sounds. Across the ten categorization training blocks, listeners' mean accuracy (i.e., responses consistent with feedback assignments) was $M = 87.25\%$, $SE = 0.016$. The orthogonal arrangement of the novel stimuli presented in the test blocks made it possible to assess categorization across one cue while the other was held constant. Figure 2 illustrates listeners' categorization responses and reaction times to the 12 generalization stimuli. The categorization functions averaged across listeners' responses to these novel stimuli hint that CF may have been more effective in cueing category identity because listeners were more adept at categorizing along the CF dimension than the MF dimension. An analysis of variance (ANOVA) of the percent "A" categorization responses as a function of the two orthogonal acoustic cues reveals that there was a statistically reliable Dimension × Stimulus Step interaction for both categorization response, $F(5,65) = 8.45$, $p < 0.0001$, $\eta_p^2 = 0.394$, and the reaction times, $F(5,65) = 2.77$, $p = 0.025$, $\eta_p^2 = 0.175$.

Cue weights were computed for each subject as the correlation between dimension values and percentage "A" responses across generalization stimuli. The absolute values of the correlation coefficients were normalized to sum to one.[2] These relative weights confirm the impression from the cat-

egorization functions of the dominance of the CF cue. The average weights were 0.658 for the CF cue and 0.342 for MF. The CF weight was significantly different from 0.5, meaning that listeners did not weight the two cues equivalently as may be expected given the match on informativeness.[3]

Two major conclusions may be made from the present study. First, listeners were able to learn to categorize novel acoustic stimuli effectively and did integrate information across the two dimensions (weights were above 0 for both dimensions). As such, this paradigm provides a useful means for examining cue weighting for categorization in a two-dimensional acoustic space. Second, despite the fact that the discriminability of the exemplars was psychophysically matched across the acoustic dimensions and despite the equivalent informativeness of the acoustic dimensions to the categorization task, listeners weighted CF more heavily than MF.

Of the 14 subjects, only one weighted MF higher than CF. Two participants used CF almost exclusively whereas the majority relied on both cues to some extent but weighted CF higher than MF. The reason for the dominance of CF is not immediately clear; we suggest some possibilities in the Sec. VI. However, it provides us with an opportunity to examine how changes in the training distribution characteristics may encourage listeners to shift their weighting functions.

III. EXPERIMENT 2

The aim of Experiment 2 was to determine whether listeners would make greater use of the MF cue in categorization under circumstances in which CF was rendered less informative. To accomplish this, the input distributions employed for Experiment 1 were shifted toward one another along the CF dimension, which resulted in greater overlap of the distributions on this dimension.

A. Methods

1. Participants

Fifteen volunteers recruited from Carnegie Mellon University participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The two-dimensional stimulus space constructed for Experiment 1 was used again in Experiment 2, but the input distributions defining the two categories were moved closer to one another along the CF dimension. The centroids of the training distributions in Experiment 1 had CF values of 760 and 970 Hz (differentiated by seven stimulus steps or 210 Hz), respectively. In Experiment 2, these values were 820 and 910 Hz (three stimulus steps, 90 Hz). As a result of this stimulus manipulation, the category input distributions overlapped more along the CF dimension than the MF dimension. In other words, the CF dimension became less informative. By the informativeness metric introduced in Experiment 1, $I_{CF}=0.5$ (75% correct with optimal boundary). All other distribution characteristics were identical to those of Experiment 1. Therefore, the informativeness of the MF dimension was unchanged from Experiment 1 ($I_{MF}=0.916$).

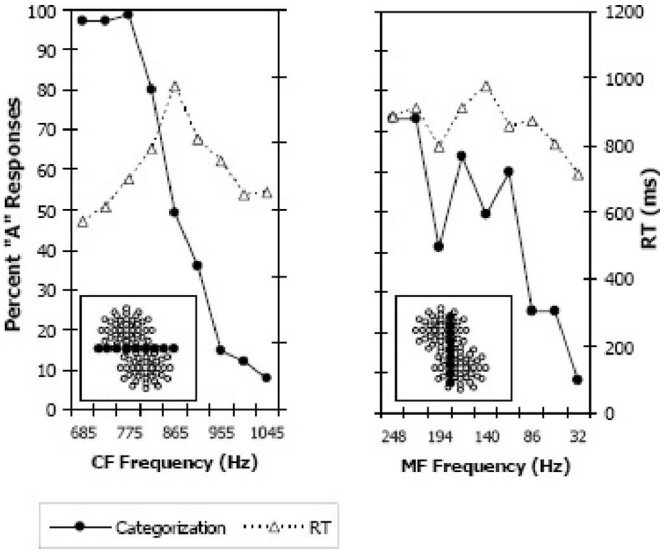


FIG. 3. Experiment 2 categorization responses and reaction time for generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

The new input distributions are illustrated as open symbols in Fig. 1(b). The generalization stimuli, which were identical to the stimuli from Experiment 1, are illustrated as filled circles in Fig. 1(b). To provide a finer sampling of categorization across the two cues, several additional stimuli were added to the novel stimulus set of Experiment 2; whereas there were 12 novel stimuli in Experiment 1, there were 17 in Experiment 2.

3. Procedure

The apparatus and procedures were identical to those used in Experiment 1.

B. Results and conclusions

Overall, listeners learned the categories very well, exhibiting a mean of 90.42% correct, $SE=1.46$, across the ten training blocks. Figure 3 illustrates listeners' categorization responses to the 17 generalization stimuli. Again, listeners' responses to the novel stimuli were more categorical along the CF dimension than the MF dimension. Supporting this conclusion, there was a significant Dimension × Stimulus Step interaction for both categorization responses, $F(8,112)=31.83$, $p<0.0001$, $\eta_p^2=0.695$ and reaction times, $F(1,112)=2.818$, $p=0.007$, $\eta_p^2=0.168$.

The average cue weights for the two dimensions were 0.664 and 0.336 for CF and MF, respectively. The relative weighting for CF was statistically equivalent to that obtained for Experiment 1, $t(27)=0.10$, $p=0.92$. Despite the reduced informativeness of CF for the task, listeners continued to rely on it in categorization judgments. Only two of the 15 subjects-gave greater weighting to the more informative cue (MF) and even in these cases the weights for CF and MF were approximately equal. Shifting the informativeness of the CF cue by moving the distribution means closer therefore was not effective in altering listeners' cue weighting.

IV. EXPERIMENT 3

The continued reliance on CF in Experiment 2 is surprising given that the observed pattern of cue weighting was a relatively inefficient weighting function for the training distributions. An obvious question is whether listeners' weighting functions are malleable at all. The goal of Experiment 3 was to present a categorization task that penalized a CF-dominant weighting function more than did the tasks of the previous two experiments. To accomplish this, we changed the variance of the training distributions in addition to changing the position of the centroids, as in Experiment 2.

A. Methods

1. Participants

Fourteen volunteers recruited from Carnegie Mellon University participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The training distributions for Experiment 3 are shown in Fig. 1(c). The distribution centroids were identical to those of Experiment 2. In Experiments 1 and 2 the distributions spanned nine stimulus steps along each dimension. In Experiment 3, the variability along the CF dimension was increased (distributions spanned 15 stimulus steps) whereas the variability along the MF dimension was reduced (distributions spanned approximately five stimulus steps). The combined changes rendered MF a maximally informative cue for the categorization ($I_{MF}=1.0$) and CF a marginally informative cue ($I_{CF}=0.442$). The training distributions were composed of 86 unique stimuli. The generalization stimuli are illustrated as filled circles in Fig. 1(c) and are identical to those of Experiment 2.

3. Procedure

The procedure and apparatus were identical to previous experiments.

B. Results and conclusions

Across the ten training blocks, listeners exhibited a mean categorization accuracy of $M=78.1\%$, $SE=3.38$. Figure 4 illustrates listeners' categorization of the generalization stimuli by acoustic dimension. Unlike the previous experiments, there was no statistically significant dimension (CF versus MF) by stimulus step interaction in listeners' categorization responses, $F(8,104)=1.36$, $p=0.222$, $\eta_p^2=0.095$, or reaction times, $F(8,104)=1.73$, $p=0.10$, $\eta_p^2=0.117$. There was, however, a significant difference in overall reaction time, $F(1,13)=4.69$, $p=0.05$, $\eta_p^2=0.265$, with stimuli varying along the MF dimension categorized somewhat faster ($M=667.63$, $SE=43.01$) than stimuli varying along the CF dimension ($M=738.21$, $SE=68.5$).

The changes in distribution variances were effective in changing the preferred weighting of the dimensions. The average weight for CF was 0.395 with a corresponding MF weight of 0.65. The change in CF weight from Experiment 2 was significant, $t(27)=4.41$, $p<0.0005$. Eleven of the 14 subjects weighted MF greater than CF.

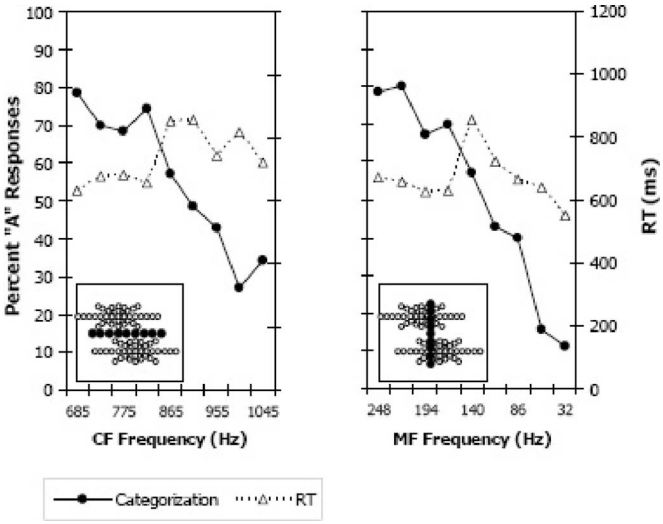


FIG. 4. Experiment 3 categorization responses and reaction time of categorization responses to novel generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

The results of Experiment 3 make it clear that weighting functions *can* be changed by the characteristics of the training distributions during a short learning session. Of theoretical and practical interest is the determination of which distributional characteristics can effect change in weighting functions. Distribution variance appears to be a better predictor of relative weighting than are measures of central tendency or our measure of informativeness. Whereas the shift in distribution centroids in Experiment 2 resulted in no change in weighting functions despite a large change in informativeness of the CF cue, the manipulations of variability in Experiment 3 led to a reversal of relative weighting with a smaller change in informativeness of the cues.

From the classic perspective that variability is the essence of information in perception, the lower weighting of the cue with greater within-category variability appears to be counterintuitive. For example, in experiments on sample discrimination in which listeners are asked to listen and respond to tones in a target frequency region, irrelevant tones in distant non-target frequency regions have little effect if they are fixed in frequency. However, when these distracter tones are allowed to vary in frequency, listeners' performance on the target task suffers dramatically (Lutfi, 1992; Neff and Odgaard, 2004). This effect can be thought of as a capture of selective attention by the variability of the irrelevant tones or it can be conceptualized as a greater weighting of highly variable components in the task, as in Lutfi's (1993) CoRE model. In the current experiment, the listeners seemed to be sensitive to increased variability across the CF dimension, as it led to a significant change in weighting functions. However, it was the lower-variance MF dimension that received the greater perceptual weight.

The details of sampling discrimination experiments and the current categorization experiments are too different to make direct comparisons that may illuminate the reason for

these different outcomes. However, the idea that the auditory system would be particularly sensitive to highly variable dimensions, *ceteris paribus*, is intuitively appealing. One may predict that exposure to variance along a dimension, with no attendant feedback, will result in this dimension becoming more heavily weighted. In Experiment 4 we investigated whether an acoustic dimension becomes more heavily weighted perceptually following exposure to its range of variability.

V. EXPERIMENT 4

Prior to categorization training in Experiment 4, listeners were exposed (without feedback) to stimuli varying along the MF dimension while CF was held constant at an intermediate value. It is predicted that this exposure independent of feedback for category membership will lead to a greater weighting of the MF dimension.

A. Methods

1. Participants

Fourteen volunteers recruited from Carnegie Mellon University participated for course credit or a small payment. All listeners reported normal hearing.

2. Stimuli

The distributions defining the categories were identical to those of Experiment 1. In addition, a set of generalization stimuli was created to more finely assess cue weighting across the acoustic space. The grid of filled symbols illustrated in Fig. 1(d) shows the position of these stimuli in acoustic space. This grid of novel stimuli, along with the cross of novel stimuli shown in Fig. 1(a), was presented to listeners in both Experiments 1 and 4.

In addition, another set of stimuli was created especially for Experiment 4. These stimuli were created as described for the previous experiments. They possessed a constant, intermediate CF frequency of 865 Hz and varied in MF from 5 to 275 Hz in 18-Hz steps. They are illustrated in Fig. 1(e) by the large open circle symbols.

3. Procedure

The apparatus and procedure were identical to those of Experiment 1. However, Experiment 4 participants completed a brief passive-listening segment before entering into categorization training. This segment lasted approximately 15 min. In that time, listeners heard the 16 stimuli illustrated in Fig. 1(e) as large open symbols 20 times each in random order. Listeners were instructed to simply listen to the sounds. Immediately after this passive exposure, listeners completed categorization training and generalization tests identical to those of Experiment 1. Thus, Experiments 1 and 4 differed only in the precategorization-training exposure.

B. Results and conclusions

A comparison of categorization accuracy during training across Experiments 1 and 4 reveals that the pretraining exposure to variability along the MF dimension had a significant influence on accuracy, $t(13)=2.19$, $p<0.05$. Across the ten training blocks, Experiment 4 listeners were significantly more accurate ($M=92.74\%$, $SE=0.02$) than Experiment 1 listeners ($M=87.25\%$, $SE=0.02$). Figure 5 illustrates listeners' categorization of the generalization stimuli cross stimuli [see Fig. 1(a)] by acoustic dimension.

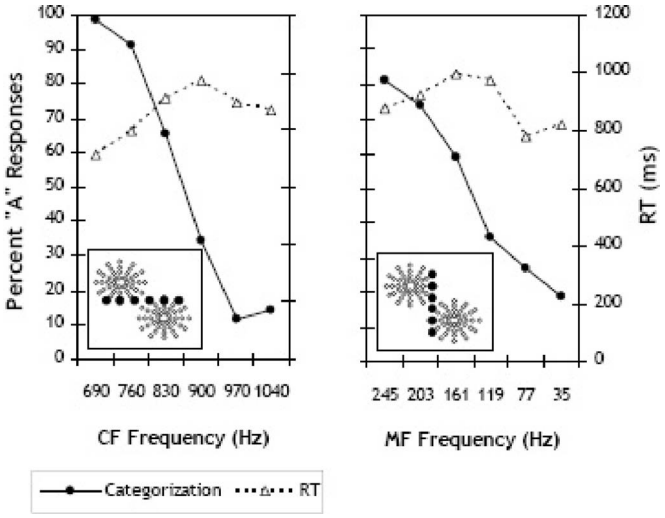


FIG. 5. Experiment 4 categorization responses and reaction time of categorization responses to novel generalization stimuli. Filled circles in the inset indicate the stimuli for which responses are plotted.

Calculations of cue weights also demonstrated that the passive exposure had an effect on categorization. Recall that Experiments 1 and 4 listeners also heard the grid of stimuli illustrated in Fig. 1(d). A contour plot illustrating listeners' categorization responses to the grid of novel generalization stimuli is shown in Fig. 6. When relative cue weights are calculated from responses to this grid of stimuli, Experiment 1 listeners exhibited a greater reliance on the CF dimension (0.744) than they did on MF (0.256). As predicted, the rela-

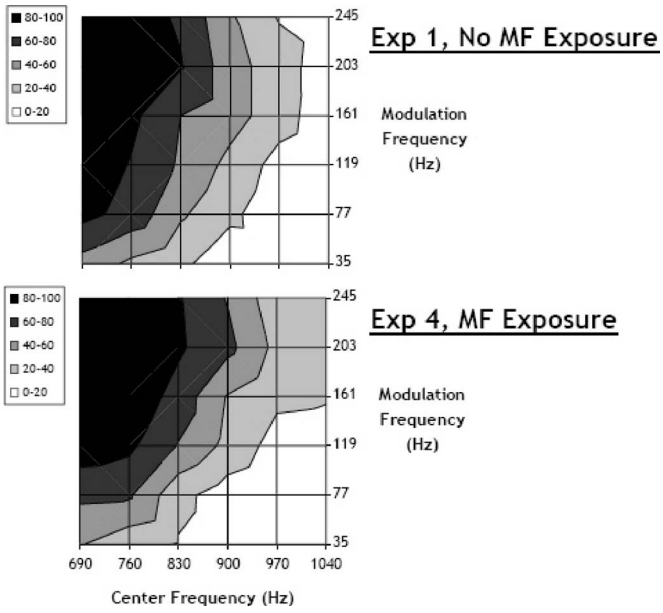


FIG. 6. Experiment 1 (no exposure to MF dimension) and Experiment 4 (exposure to MF dimension) categorization responses to generalization grid stimuli, as shown in Fig. 1(d), plotted as percent "A" category responses.

tive cue weight for CF was significantly lower for Experiment 4 (0.600) than for Experiment 1 (0.744), [$t(26)=1.75$, $p<0.05$, one tailed]. That is, preexposure to variability along the MF dimension with no feedback led listeners to rely upon MF in categorization responses moderately more than listeners who did not experience the preexposure (0.256 versus 0.400). These results suggest that the listeners were sensitive to the range of the MF dimension present in preexposure and treated it as a potential source of information. Note that this was a very conservative estimate of the effect of exposure to cue variability on cue weighting because the dependent variable was the categorization response to generalization stimuli presented at the end of categorization training. Still, a very short period of preexposure had an influence on cue weighting. It is quite possible that the categorization responses at the beginning of categorization training were more heavily weighted toward MF but that the preference for CF rebounded during training. The design of the current experiment makes it difficult to test this explicitly. However, the current evidence is suggestive of a role for unsupervised exposure with cue variability in determining weighting functions.

VI. GENERAL DISCUSSION

Although multiple cues define most auditory categories, including speech categories, how the auditory system integrates multiple acoustic dimensions in perceptual categorization is not yet well understood. The present work investigated cue weighting in auditory categorization, approaching the issue from the perspective that the perceptual system does not merely perceive acoustic dimensions and apply them as evidence for a particular category, but rather weights these dimensions as a function of characteristics of distributions of experienced sounds.

In the Introduction, we listed four variables that may affect the weighting functions for a categorization task. We summarize here the results of the experiments as they relate to these factors. The first of these potential factors was the informativeness of the cue for categorization. If the listener is weighting cues optimally then relative weightings should be predictable from relative informativeness (at least in terms of the rank ordering of the cues). The data from our experiments do not reveal such a direct relationship. In Experiment 1, CF received a much greater weighting despite the fact that the two cues were equal on our informativeness metric. CF continued to receive a higher weighting even when it was decreased in its informativeness to the categorization task in Experiment 2 by shifting the centroids of the distributions closer together on the CF dimension. We refer to a cue receiving a higher weight than dictated by its informativeness for the current task as having greater *salience* (cue *preference* or *bias* are other possible terms). Christensen and Humes (1996) also examined the categorization of nonspeech sounds differing on several dimensions and found that one of the cues tended to be weighted more despite equal informativeness for the task. However, their dimensions were not equated psychophysically for auditory step size. As a result, the weighting functions could have been reliably reflecting the informativeness of the psychoacoustic cues as opposed to the physical acoustic cues. In our experiments, the cue-weights were computed across equated psychoacoustic scales and thus reflect perceptual cue weighting.

What accounts for this salience? One possibility is that in the natural acoustic environment, CF is a more informative cue to auditory category identity than is MF. That is, in identifying a sound, the carrier frequency may provide more information or may be more reliably related to the distal source than is the rate at which the sound is modulated. Whereas we have no direct evidence that this is the case, it does have intuitive appeal. If this is the case, then listeners may come into the experiment with experience that is relevant to the task. A default higher weighting for CF could be a result of an innate predisposition (because the informativeness of CF has led to a fixed adaptation) or learned through experience. Similar patterns of differences in cue salience have been witnessed in studies of cross-modal integration. Battaglia, Jacobs, and Aslin (2003) report that visual cues are weighted more than auditory cues for spatial location even when the visual cues are not as reliably related to spatial location (see also Ernst and Banks, 2002, for an example of visual cue salience over haptic cues). Battaglia *et al.*. were able to model this influence by proposing that perceivers had *a priori* expectations of the reliability of visual information for spatial location (defined as a prior probability distribution for visual cue variance). Thus, perceivers appear to be weighting suboptimally within the context of the information in an experiment, but they may perceive optimally within the larger context of overall perceptual experience. The interaction of short-term (or local) informativeness and long-term (or global) informativeness will be an important area for future investigation.

A second factor that may affect cue weighting is distributional variance. Because our definition of informativeness is based on a criterion-based decision, variance and informativeness are not equivalent. In Experiment 3, we took the distributions from Experiment 2 and increased the within-category distribution variance along CF and decreased it along MF. The result was a small change in relative informativeness but a large change in the relative within-distribution variance for the two dimensions. Whereas the change in informativeness from Experiments 1 to 2 resulted in no change in perceptual cue weighting, the manipulation of within-category distribution variances led to a substantial change in weighting, with MF now the dominant perceptual cue. Because of the design of the experiments, it is not possible to establish exactly what caused the change in perceptual weighting from Experiment 2 to 3. The change in within-distribution variances resulted in changes in informativeness, the ratios of within- to between-distribution variance and correlations with the feedback for each cue. Further studies will be necessary to manipulate these factors independently. What is clear from these experiments is that changes in the central tendencies of the distributions (with concomitant changes in informativeness) do not appear to be effective in changing perceptual weighting functions, whereas changes in within-category variance can lead to significant perceptual change. Another open question is the influence of the pres-

ence of feedback on perceptual weighting. In Experiment 3, the CF dimension with greater within-category variance during training with feedback received the lower perceptual weighting. In Experiment 4, we presented the range of variability in MF with no feedback prior to the categorization training. This exposure resulted in a moderate increase in the relative weighting of the varied cue.[4] The present experiments do not allow us to fully dissociate the possible roles of variability along an acoustic cue with versus without explicit feedback for categorization; further studies will be necessary to investigate these manipulations independently. However, the present data are interesting in that they indicate that manipulation of distribution characteristics of acoustic categories plays a role in listeners' perceptual cue weighting.

A third factor that can affect cue weighting is how each dimension is represented by the auditory system. We attempted to account for some of the effects of auditory encoding by scaling our dimensions to equivalent auditory step sizes based on pilot testing of the stimuli. Whereas it is true that CF and MF are not independent in frequency modulation detection across a large range of values, we saw no evidence of interaction in our constrained ranges; we avoided very low MF values and created stimuli within a MF range where these dimensions are relatively independent (Demany and Semal, 1989). We did not model the effects of internal noise on the representation of our cue values. We correlated responses with the step size values of our stimuli as though these values are represented perfectly by the auditory system. The General Recognition Theory (GRT, Ashby and Townsend, 1986), a model of categorization developed mainly from visual categorization experiments, represents stimuli as multivariate probability distributions in perceptual space as opposed to points. Likewise, the COSS (conditioned-on-a-single-stimulus) analysis of perceptual weights developed for multicomponent psychoacoustic tasks (Berg, 1989) internal noise is an explicit parameter. It is likely that models of cue weighting in phonetic and other auditory categorization tasks will require a more detailed estimate of the auditory representation of the stimuli than we have used here.

The fourth potential factor affecting cue weighting is task. All of the experiments presented here were categorization tasks in which subjects were asked to maximize accuracy, so the present data do not speak to manipulating cue weighting through task adjustments. Nevertheless, task manipulations could be easily implemented in the present paradigm and would be particularly interesting to investigate in future research. In speech perception research, it is common to examine speech categories with categorization or discrimination paradigms. If one presumes that both paradigms access the same category representation, then they can provide converging evidence about auditory categories. However, if listeners modify their weighting functions to maximize performance given particular task constraints then discrimination and categorization data may not always coincide (e.g., Mirman *et al.*, 2004).

A. Implications for L2 phonetic acquisition

In the Introduction, we discussed the proposal that one of the obstacles to acquiring L2 phonetic categories could be a mismatch between weighting functions appropriate for L1 and L2. For example, Lotto *et al.* (2004) estimated the phonetic distributions for syllable-initial /r/ and /l/ from native-English productions and determined that the optimal weighting strategy for categorization was a very heavy perceptual weight on F3-onset frequency and a much lower weight on F2-onset frequency. In a follow up, Lotto *et al.* demonstrated that the optimal weighting pattern for the Japanese distinction between /w/ and /r/ (a distinction often considered to interfere with acquisition of the English liquid contrast) was just the opposite, with a higher weighting of F2-onset frequency. Thus, one may predict that Japanese listeners would have difficulty with the English contrast (and vice versa) because of the mismatch in learned weighting functions. In fact, perception and production studies both indicate that native Japanese listeners weight F2-onset higher than F3-onset for the English contrast (Yamada and Tohkura, 1990; Iverson *et al.*, 2003, 2005; Lotto *et al.*, 2004). This inappropriate weighting of F2 resembles the salience of CF in the current studies. Given the similarity, our results detailing effective and ineffective means of shifting weighting functions may have relevance for training L2 learners. In particular, the lack of weighting shift between Experiments 1 and 2 suggests that changing the informativeness of cues by shifting the average cue values (e.g., enhancing the difference in F3-onset values or decreasing the differences in F2-onset for /l/-/r/ stimuli presented to Japanese listeners) may not result in a significant change in categorization. On the other hand, the reversal of the perceptually dominant cue from Experiments 2 to 3 demonstrates that adding variance to an over-utilized cue (e.g., allowing F2 onset to vary independent of whether the sound is /l/ or /r/) may be an effective strategy to change listeners' weighting functions. In addition, the effect of preexposure to cue variance in Experiment 4, emphasizes the notion that the proper manipulation of variance may be essential to appropriate category acquisition.

In some earlier attempts to train adults on non-native contrasts the training sets did not contain much variance (e.g., Strange and Jenkins, 1978). As a result, the trainees were able to learn to discriminate the training stimuli but were not able to transfer this learning to new stimuli and contexts. In more recent attempts at L2 training, it has become clear that added variance in the training set (e.g., more speakers, more phonetic contexts) aids learning and generalization (Bradlow *et al.*, 1997; Jamieson and Morosan, 1989; Lively *et al.*, 1993). It is quite possible that a major benefit of high-variability training is that less informative cues will vary more within a category across multiple exemplars while more informative cues will be relatively more stable. As in Experiment 3, these changes in relative variance may be most effective at changing weighting functions. The importance of variance for category learning may even be seen in L1 acquisition. The speech that adults direct to their children ("motherese" or infant-directed speech) is actually *more* acoustically variable than speech directed to other adults

(Kuhl *et al.*, 1997). This seems counter intuitive unless one appreciates that infants must use the variance in speech input to determine which features are phonetically-relevant and which are not.

VII. SUMMARY

Of course, it is important to keep in mind that the categories of the present experiments were formed on the basis of an hour or so of experience whereas phonetic categories have the benefit of much greater experience, even among the youngest listeners. Nevertheless, we believe there is much to be gained for an understanding of phonetic cue weighting from laboratory studies that investigate cue weighting in general auditory (nonspeech) categorization. The purpose of the experiments presented here was to examine cue weighting in auditory categorization with well-defined training distributions for which experience could be entirely controlled. This allowed us to define the informativeness of each cue for the task and examine how changes in training distributions affect weighting functions. The long-term goals of this project are to establish a framework for testing models of cue weighting in phonetic categorization. It should be noted that there have been many models of perception, in general, and speech perception, in particular that have included cue weighting explicitly or implicitly. For example, Massaro's (1987; 1998) fuzzy logic model of perception (FLMP), Nosofsky's (1992) generalized context model (GCM), and connectionist models (e.g., TRACE, McClelland and Elman, 1986; Damper and Harnad, 2000) inherently weight cues in coming to a category decision. In order to provide a strong test of these models as a basis of speech perception, one needs good estimates of the distributions of phonetic category exemplars across the various dimensions. In addition to auditory cues, any full specification of phonetic categories would include visual cues, which can also be weighted in a categorization decision (Massaro, 1998). These kinds of estimates are rare.

We believe that nonspeech category tasks such as the ones presented here will play an important role in the development of models of auditory cue weighting. Whereas they do not have the ecological advantages of speech perception tasks, nonspeech tasks allow a level of stimulus and distribution control that allows one to investigate the constraints of the categorization process in detail.

ACKNOWLEDGMENTS

This work was supported by a National Science Foundation grant (NSF BCS-0078768) to L.L.H. and A.J.L. and by a James S. McDonnell Foundation award for Bridging Mind, Brain, and Behavior, to L.L.H. The authors thank Christi Adams for her essential role in conducting the experiments.

[1]An additional grid of 36 novel stimuli [pictured as filled circles in Fig. 1(d)] was also presented to Experiment 1 listeners for comparison with Experiment 4. Responses to these stimuli are discussed with the results of Experiment 4.

[2]It is typical in cue-weighting experiments in psychoacoustics and speech perception to normalize the weights (e.g., Berg, 1989; Christensen and Humes, 1996; Doherty and Turner, 1996; Lutfi, 1992). The result provides a comparison of the relative weighting of cues. Because this normalization scheme does not take into account the accuracy of listeners' categorizations, it can lead to problems with interpretation. For example, guessing randomly in Experiment 1 or responding perfectly with a rational weighting strategy would both result in equal normalized weights of 0.5. Given the high level of accuracy attained by listeners in the experiments presented here, we believe that this ambiguity is less of an issue than it may be for more complex tasks.

[3]All statistical tests for cue weights were performed only on the CF weights, as the MF weights are simply 1 minus the CF weight.

[4]These results may relate to an issue that we have avoided for the most part. There has been a tendency in the literature to equate cue weighting with selective attention. Some models of categorization explicitly equate weighting with attention (e.g., Nosofsky, 1986). The sensitivity of the auditory system to variability as demonstrated in psychoacoustic tasks (e.g., Lutfi, 2003; Neff and Odgaard, 2004) can be interpreted as variability capturing selective attention. The results of Experiment 4 also fit this interpretation. The change in weighting with variance manipulations between Experiments 2 and 3 demonstrate a possible dissociation between weighting and attention. Francis and Nusbaum (2002) have also called into question the viability of simple attentional models in phonetic category learning.

Ashby, F. G., and Townsend, J. T. (**1986**). "Varieties of perceptual independence," Psychol. Rev. **93**, 154–179.

Battaglia, P. W., Jacobs, R. A., and Aslin, R. N. (**2003**). "Bayesian integration of visual and auditory signals for spatial localization," J. Opt. Soc. Am. A **20**, 1391–1397.

Benguerel, A.-P. and Bhatia, T. K. (**1980**). "Hindi stop consonants: An acoustic and fiberscopic study," Phonetica **37**, 134–148.

Berg, B. G. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tohkura, Y. (**1997**). "Training the Japanese listener to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Christensen, L. A., and Humes, L. E. (**1996**). "Identification of multidimensional complex sounds having parallel dimension structure," J. Acoust. Soc. Am. **99**, 2307–2315.

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

Crowther, C. S., and Mann, V. (**1994**). "Use of vocalic cues to consonant voicing and native language background: the influence of experimental design," Percept. Psychophys. **55**, 513–525.

Damper, R. I., and Harnad, S. R. (**2000**). "Neural network modeling of categorical perception," Percept. Psychophys. **62**, 843–867.

Demany, L., and Semal, C. (**1989**). "Detection thresholds for sinusoidal frequency modulation," J. Acoust. Soc. Am. **85**, 1295–1301.

Doherty, K. A., and Turner, C. W. (**1996**). "Use of a correlational method to estimate a listener's weighting function for speech," J. Acoust. Soc. Am. **100**, 3769–3773.

Ernst, M. O., and Banks, M. S. (**2002**). "Humans integrate visual and haptic information in a statistically optimal fashion," Nature (London), **415**, 429–433.

Flege, J. E., and Wang, C. (**1989**). "Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final /t/-/d/ contrast," J. Phonetics **17**, 299–315.

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (**2000**). "Effects of training on attention to acoustic cues," Percept. Psychophys. **62**, 1668–l680.

Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Grieser, D., and Kuhl, P. K. (**1989**). "Categorization of speech by infants: Support for speech-sound prototypes," Dev. Psychol. **25**, 577–588.

Hazan, V., and Barrett, S. (**2000**). "The development of phonemic categorization in children aged 6-12," J. Phonetics **28**, 377–396.

Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**, 3013–3022.

Holt, L. L., Lotto, A. J., and Diehl, R. L. (**2004**). "Auditory discontinuities interact with categorization: Implications for speech perception," J. Acoust. Soc. Am. **116**, 1763–1773.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English

/r/-/l/ to Japanese adults. J. Acoust. Soc. Am. **118**, 3267–3278.
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–B57.
Jamieson, D., and Morosan, D. (**1989**). "Training new, normative speech contrasts: A comparison of the prototype and perceptual fading techniques," Can. J. Psychol. **43**, 88–96.
Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, MA).
Jusczyk, P. W., Pisoni, D. B., Walley, A., and Murray, J. (**1980**). "Discrimination of relative onset time of two-component tones by infants," J. Acoust. Soc. Am. **67**, 262–270.
Keating, P. A. (**1984**). "Phonetic and phonological representation of stop consonant voicing," Lang. **60**, 286–319.
Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. B. (**1998**). "Role of experience in language-specific functional mappings for vowel sounds as inferred from human, nonhuman, and computational models," J. Acoust. Soc. Am. **104**, 3568–3582.
Kuhl, P. K. (**1991**). "Human adults and human infants show a 'perceptual magnet effect' for the 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.
Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by six-months of age," Science **255**, 606–608.
Kuhl, P. K., and Miller, J. D. (**1975**). "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," Science **190**, 69–72.
Liberman, A. (**1996**). *Speech: A Special Code* (MIT Press, Cambridge, MA).
Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). "Perception of the speech code," Psychol. Rev. **74**, 431–461.
Lisker, L. (**1986**). ""Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang Speech **29**, 3–11.
Lisker, L., and Abramson, A. S. (**1964**). "A cross-linguistic study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.
Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.
Lotto, A. J., Sato, M., and Diehl, R. L. (**2004**). "Mapping the task for the second language learner: The case of Japanese acquisition of /r/ and /l/," *From Sound to Sense:* 50+ *Years of Discoveries in Speech Communication*, Research Laboratory of Electronics at MIT, Cambridge, MA.
Lutfi, R. A. (**1992**). "Informational processing of complex sound. III: Interference," J. Acoust. Soc. Am. **91**, 3391–3401.
Lutfi, R. A. (**1993**). "A model of auditory pattern analysis based on component-relative-entropy," J. Acoust. Soc. Am. **94**, 748–758.
Lutfi, R. A. (**2003**). "Perturbation analysis of acoustic cues for sound source identification," J. Acoust. Soc. Am. **113**, 2326.
Lutfi, R. A., and Doherty, K. A. (**1994**). "Effect of component-relative-entropy on the discrimination of simultaneous tone complexes," J. Acoust. Soc. Am. **96**, 3443–3450.
Massaro, D. W. (**1987**). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry* (Lawrence Erlbaum Associates, Hillsdale, NJ).
Massaro, D. W. (**1998**). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT Press, Cambridge, MA).
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.
McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE Model of Speech Perception," Cogn. Psychol. **18**, 1–86.
Mirman, D., Holt, L. L., and McClelland, J. M. (**2004**). "Categorization and discrimination of non-speech sounds: Differences between steady-state and rapidly-changing acoustic cues," J. Acoust. Soc. Am. **116**, 1198–1207.
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**, 331–340.
Morrongiello, B. A., Robson, R. C., Best, C. T., and Clifton, R. K. (**1984**). "Trading relations in the perception of speech by five-year-old children," J. Exp. Child Psychol. **37**, 231–250.
Neff, D. L., and Odgaard, E. C. (**2004**). "Sample discrimination of frequency differences with distracters," J. Acoust. Soc. Am. **116**, 3051–3061.
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., 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., Miller, M. E., Crowther, C. S., and Manhart, M. J. (**2000**). "The effect of segmental order on fricative labeling by children and adults," Percept. Psychophys. **62**, 266–284.
Nosofsky, R. M. (1992). "Similarity scaling and cognitive process models," Annu. Rev. Psychology 43, 25–53.
Nosofsky, R. M. (**1986**). "Attention, similarity, and the identification-categorization relationship," J. Exp. Psychol. Gen. **115**, 39–57.
Parnell, M. M., and Amerman, J. D. (**1978**). "Maturational influences on the perception of coarticulatory effects," J. Speech Hear. Res. **21**, 682–701.
Pisoni, D. B. (**1977**). "Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops," J. Acoust. Soc. Am. **61**, 1352–1361.
Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., and Papanicolaou, A. C. (1998). "MEG correlates of categorical perception of a voice onset time continuum in humans," Cognit. Brain Res. **7**, 215–219.
Simos, P. G., and Molfese, D. L. (**1997**). "Electrophysiological responses from a temporal order continuum in the newborn infant," Neuropsychologia **35**, 89–98.
Sinex, D., and McDonald, L. (**1989**). "Synchronized discharge rate representation of voice-onset time in the chinchilla auditory nerve," J. Acoust. Soc. Am. **85**, 1995–2000.
Sinex, D., McDonald, L., and Mott, J. (**1991**). "Neural correlates of non-monotonic temporal acuity for voice onset time," J. Acoust. Soc. Am. **90**, 2441–2449.
Steinschneider, M., Volkov, I., Noh, M., Garell, P., and Howard, M. (**1999**). "Temporal encoding of the voice onset time phonetic parameter by field potentials recorded directly from human auditory cortex," J. Neurophysiol. **82**, 2346–2357.
Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics **17**, 3–45.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).
Strange, W., and Jenkins, J. J. (**1978**). "Role of linguistic experience in the perception of speech," in *Perception and Experience*, edited by Pick, H. L. and Walk, R. D. (Plenum, New York), pp. 125–169.
Walley, A. C., and Carrell, T. D. (**1983**). "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 1011–1022.
Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Circuit Tree **7**, 49–63.
Yamada, R. A., and Tohkura, Y. (**1990**). "Perception and production of syllable-initial English /r/ and /l/ by native speakers of Japanese," *Proceedings of International Conference on Spoken Language Processing*, Kobe, Japan, 757–760.

Early learners' discrimination of second-language vowels

Anders Højen[a] and James E. Flege[b]
Speech and Hearing Sciences Division, University of Alabama, Birmingham, Alabama 35294-2041

(Received 21 March 2005; revised 7 February 2006; accepted 14 February 2006)

It is uncertain from previous research to what extent the perceptual system retains plasticity after attunement to the native language (L1) sound system. This study evaluated second-language (L2) vowel discrimination by individuals who began learning the L2 as children ("early learners"). Experiment 1 identified procedures that lowered discrimination scores for foreign vowel contrasts in an AXB test (with three physically different stimuli per trial, where "X" was drawn from the same vowel category as "A" or "B"). Experiment 2 examined the AXB discrimination of English vowels by native Spanish early learners and monolingual speakers of Spanish and English (20 per group) at interstimulus intervals (ISIs) of 1000 and 0 ms. The Spanish monolinguals obtained near-chance scores for three difficult vowel contrasts, presumably because they did not perceive the vowels as distinct phonemes and because the experimental design hindered low-level encoding strategies. Like the English monolinguals, the early learners obtained high scores, indicating they had shown considerable perceptual learning. However, statistically significant differences between early learners and English monolinguals for two of three difficult contrasts at the 0-ms ISI suggested that their underlying perceptual systems were not identical. Implications for claims regarding perceptual plasticity following L1 attunement are discussed. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2184289]

PACS number(s): 43.71.Ft, 43.71.Es, 43.71.Hw [ARB] Pages: 3072–3084

I. INTRODUCTION

It is well established that the age at which a second language (L2) is learned will exert a long-term effect on segmental phonetic perception. "Early learners" first exposed to their L2 as children resemble native speakers to a greater extent than do "late learners" whose first exposure to the L2 occurred in late adolescence or early adulthood (Flege *et al.*, 1999; Flege and MacKay, 2004; Yamada, 1995). It is uncertain, however, whether early learners will in time come to perceive L2 vowels and consonants just like native speakers of the target L2. Some studies have provided evidence of substantial differences between early learners and L2 native speakers whereas others have shown evidence of little or no difference. These divergent findings raise questions regarding the extent to which the speech perception system remains plastic following attunement to the native language (L1) sound system. The aim of this study, therefore, was to further evaluate early learners' perceptual abilities. It did so by examining the discrimination of English vowels by native Spanish speakers who had learned English as an L2 in childhood.

A. Previous research

Infants begin to perceive speech in a language-specific way during the first year of life (e.g., Kuhl *et al.*, 1992; Werker and Tees, 1984). As they develop perceptual processes specialized for perceiving phonemic contrasts used to contrast meaning (words) in the ambient language, infants' sensitivity to contrasts not serving such a function declines. The influence of this attunement to the L1 sound system is evident in the difficulties that adult L2 learners may have in distinguishing certain nonnative phonemic contrasts (e.g., in native Japanese speakers' difficulty in distinguishing English /ɹ/ and /l/).

It has been proposed that as people grow older, L2 learning is hindered by a loss of neuroplasticity that accompanies normal neural maturation (e.g., Long, 1990; Scovel, 2000). More recently, a loss of plasticity for speech perception has been discussed in the context of the effects of L1 perceptual attunement on subsequent perceptual learning rather than, or in addition to, a maturationally defined decrease of neuroplasticity that affects basic aspects of sensory coding, storage, and retrieval associated with speech learning. More specifically, it has been argued that L2 speech acquisition is impeded by the acquisition of the L1 phonological system (e.g., Flege, 1995; Iverson *et al.*, 2003; Kuhl, 2000; McClelland, 2001; Pallier *et al.*, 1997, 2003; Sebastián-Gallés and Soto-Faraco, 1999).

One influential account of the effect of L1 perceptual attunement on L2 perception is the native language magnet (NLM) theory (e.g., Kuhl *et al.*, 1992; Kuhl and Iverson, 1995). According to the NLM, during the first year of life L1 phonetic prototypes begin to act as perceptual "magnets" that "warp" auditory input. This would be expected, for example, to reduce the discriminability of L1 vowel tokens near the category center (the prototype) whereas equally distant tokens from the category boundary region, or from two different categories, would be expected to remain highly discriminable.

[a]Present affiliations: University of Southern Denmark and Højen Research. Address for correspondence: Praestegaardsvej 2 DK-8900 Randers, Denmark; electronic mail: anders@andershojen.dk
[b]Present affiliation: Institute of Cognitive Sciences and Technologies, National Research Council, Rome, Italy.

 0001-4966/2006/119(5)/3072/13/$22.50

The NLM predicts that as an L1-specific "perceptual map" is developed, the identification of L1 speech sounds will become more robust whereas learning to perceive instances of new, L2 categories may become difficult. This might hold true, for example, if tokens of two distinct L2 vowel categories straddle the center of a vowel category in the L1 of a learner. Tokens of both L2 vowels would tend to be "attracted" by the same L1 prototype, effectively reducing perceptual distance between the L2 vowel tokens. For a native speaker of the L2, on the other hand, the perceived distances between the two tokens would be augmented by the perceptual magnet effect given that the tokens would be differentially categorized, and thus attracted by two different prototypes (see also Flege *et al.*, 1994).

Iverson *et al.* (2003) suggested that an L1-conditioned warping of the auditory properties of incoming speech occurs at an early phonetic or late auditory level of processing and may be difficult to reverse later in life. This view was supported by the results of five studies examining the perception of Catalan vowels by young native Spanish adults who had begun learning Catalan in Barcelona by the age of 6 years. Using different testing techniques, all assessed the perception of Catalan /e/ and /ɛ/, which straddle the category center of Spanish /e/. In each study, the Spanish early learners of Catalan differed from native speakers of Catalan in perceiving Catalan /e/ and /ɛ/ (Bosch *et al.*, 2000; Pallier *et al.*, 1997, 2001; Sebastián-Gallés and Soto-Faraco, 1999). Most recently, Sebastián-Gallés *et al.* (2005) presented Catalan words and nonwords (created by changing /e/ to /ɛ/ and /ɛ/ to /e/) for lexical decision. The early learners obtained significantly lower scores than Catalan native speakers did, and also showed a wider range of scores.

The research on early native Spanish learners of Catalan suggested a number of important conclusions. One was that the speech perception system "does not seem able to easily develop new phonemic categories" after perceptual attunement to the L1 phonological system (Bosch *et al.*, 2000, p. 193). Another was that early learners exhibit a "lack of behavioral plasticity" (Pallier *et al.*, 1997, p. B9) with respect to learning to perceive contrastive phonemes of an L2. Still another was that "there are severe limitations to the malleability of the initially acquired L1 phonemic categories even under conditions of early and extensive exposure" (Sebastián-Gallés and Soto-Faraco, 1999, p. 120).

The results of two studies examining the perception of English vowels by native Italian immigrants to Canada suggested a different conclusion regarding the perceptual ability of early learners. Both of these studies used a triadic categorial discrimination test (CDT) in which the task was to identify the serial position of an odd item out (in change trials) or to indicate that all three tokens were instances of a single category (in no-change trials). Flege *et al.* (1999) tested early native Italian learners of English who reported using Italian relatively often ("high-L1-use" group) or seldom ("low-L1-use" group). The low-L1-use group did not differ from native English (NE) monolinguals in discriminating English vowels whereas the high-L1-use group was found to differ marginally from the NE group. A similar result was obtained by Flege and MacKay (2004), who observed a significant vowel discrimination difference between NE speakers and an early high-L1-use group but not an early low-L1-use group. The authors interpreted this finding to mean that early exposure to the L2 does not guarantee the formation of new phonetic categories for L2 vowels, but that category formation will be more likely for individuals who use their L1 infrequently.

The research with Spanish learners of Catalan suggested that early learners are unlikely to perceive L2 vowels in a nativelike way whereas the research with Italian learners of English suggested that early learners may closely resemble L2 native speakers, at least those who use the L1 infrequently. The basis for the apparent divergence in findings obtained for early L2 learners in Canada and Spain (Barcelona) is uncertain. It was not reported how often the early learners in Barcelona used their L1 (Spanish). Perhaps they differed from native speakers because they used their L1 as frequently as the "high-L1-use" participants in the Canadian research (*viz.* an approximately equal L1-L2 use). Other possible explanations for the differing results obtained in Canada and Spain include differences in the perceived relation between the L1 and L2 vowels examined, use of diverse testing techniques that might have induced different levels of processing, the age of the participants, and the amount of foreign-accented L2 input the participants might have received.

Obtaining a better understanding of early learners' perception of L2 vowels is important because it is relates to the more general issue of perceptual plasticity in human speech. Native-nonnative differences might be attributed to a loss of plasticity arising either from a reduced capacity to use one or more processes needed to develop language-specific, long-term memory representations for vowels and consonants, or from the persistent influence of prior learning on subsequent learning [e.g., the early-level "warping" of input due to the effect of L1 "magnets;" see Iverson *et al.* (2003)]. Alternatively, native-nonnative differences might be attributed to differences in attention and/or input, or to the inability to sustain an optimal representation of more than one linguistic system (Grosjean, 1989).

Even assuming that the speech perception system remains fully plastic after the L1 sound system has been established, it is probably unreasonable to suppose that early learners' perception of L2 vowels could ever be identical to that of L2 native speakers. This is because perceptual maps are presumably based on the statistical properties of individuals' linguistic input (e.g., Kuhl, 2000; Kluender *et al.*, 1998), which can never have been the same for individuals with different L1s. That being the case, identical perception by groups of early learners and L2 native speakers should probably not be used as a litmus test for whether the speech perception system does or does not remain plastic following L1 perceptual attunement.

B. The present study

The aim of most previous L2 vowel perception research evaluating perceptual plasticity has been to determine if groups of early learners did or did not differ significantly from a control group of L2 native speakers. A more informa-

tive way to assess perceptual learning by early learners may be to compare their performance to that of two control groups, *viz.* L2 native speakers (as in past research) and monolingual native speakers of the early learners' L1. The present study compared the English vowel discrimination of early native Spanish learners of English to that of English monolinguals and Spanish monolinguals. Comparing the early learners' performance to that of English monolinguals provided a way to determine if their perceptual learning was complete. Comparing their performance to that of Spanish monolinguals provided a way to determine how much the early learners had learned, even in the absence of complete learning.

Including Spanish monolinguals ruled out the use of testing techniques that presupposed, even indirectly, a knowledge of English, e.g., forced choice identification or goodness rating. Although the CDT used in the research on Italian learners of English has certain advantages, it does not control response bias as effectively as another discrimination test widely used in cross-language and L2 research, the categorial AXB test (e.g., Best *et al.*, 1996, 2001; Harnsberger, 2001; Polka, 1995; Polka and Bohn, 1996).

A problem for the categorial AXB test, on the other hand, is that it has often yielded near-ceiling discrimination scores (90%–100% correct) for pairs of foreign vowels likely to have been perceived as instances of a single L1 category (Best *et al.*, 1996; Polka, 1995; Polka and Bohn, 1996). This suggests that the AXB paradigm may not constitute a "phonetically sensitive" test of L2 vowel discrimination in the sense that pairs of L2 vowels showing a 2-to-1 cross-language mapping pattern might be successfully discriminated on the basis of low-level (perhaps auditory) memory codes rather than different language-specific phonetic codes. Such a response would present a problem for the present study, given the plan to use the Spanish monolingual control group as a baseline to assess how much perceptual learning the early learners had shown. The Spanish monolingual reference group could not serve this function if it obtained near-ceiling scores due to the use of nonphonetic encoding strategies. Thus, for purposes of this study, an ideal ("phonetically sensitive") AXB test would be one that yielded near-ceiling scores for English monolinguals but near-chance scores for Spanish monolinguals for English vowels heard as instances of a single Spanish vowel category.

The decision was made to use the categorial AXB paradigm here, but to carry out preliminary research (experiment 1) to find a variant that would reduce scores to near-chance levels for contrasting foreign vowels heard as instances of a single L1 category. This ambitious goal was not fully met in experiment 1, but this experiment did nonetheless reveal a way to make the AXB test more phonetically sensitive. The procedures developed in experiment 1 were therefore used in experiment 2 to assess the discrimination of English vowels by native Spanish early learners, English monolinguals, and Spanish monolinguals.

II. EXPERIMENT 1

The aim of this experiment was to identify a phonetically sensitive variant of the categorial AXB discrimination test, which would yield near-ceiling scores for listeners who perceive contrasting vowels as distinct phonemes, but near-chance scores for listeners who perceive contrasting foreign vowels as instances of a single L1 category. Such a pattern of scores for Spanish and English monolinguals in experiment 2 would make it possible to assess the extent of perceptual learning by native Spanish early learners. It was not possible to recruit a sufficient number of Spanish monolinguals for this preliminary experiment as well as experiment 2. Therefore, the decision was made to test English adults on vowels drawn from a language they had never heard before, *viz.* Danish.

Previous cross-language research using the AXB discrimination test has often yielded near-ceiling scores for vowels even when tokens drawn from contrasting foreign vowel categories were classified as instances of a single L1 category (Best *et al.*, 1996; Polka, 1995; Polka and Bohn, 1996). This evidence of successful within-category vowel discrimination indicates that listeners are able to base their judgments in an AXB test on more than just language-specific phonetic codes.

Hessen and Schouten (1992) discussed the nature of processes that might be applied in the discrimination of speech sounds that are perceived as either within-category or between-category contrasts. Combining elements from dual processes theory (Fujisaki and Kawashima, 1970) and trace-context theory applied to speech (Macmillan *et al.*, 1988), Hessen and Schouten's (1992) phoneme perception theory (PPT) posits three processes for the encoding of stimuli drawn from a speech sound continuum: trace coding, context (or position) coding, and labeling. In *labeling*, the trace is mapped onto the permanent context (the language-specific sound system) and is given a phonetic label (Hessen and Schouten term it a "phoneme label"). In *context coding*, the trace is mapped onto the experimental context. The listener positions each stimulus in relation to the full range of stimuli presented and gives it a temporary sublabel, possibly relating it to "anchors" among the stimuli in the form of extreme or best category members. The sensory *trace* is stored with all acoustic information intact (except for sensory noise) but decays rapidly after a few hundred milliseconds. Seen from the perspective of the PPT, foreign vowel tokens given different phonetic labels (between-category contrasts) would be discriminated based on the labels, whereas tokens given the same label (within-category contrasts) must be discriminated using context codes or the sensory trace.

Above-chance categorial AXB discrimination of foreign vowels given the same phonetic label might be based on context codes. (Discrimination is less likely to be based on the sensory traces when all stimuli in a trial are physically different.) For example, native English speakers might label (covertly or overtly) the German vowels in an [y]-[y]-[u] trial as "two foreign-sounding u-vowels" followed by "one overly clear English u-vowel" (or some other *ad hoc* label). The use of such a strategy would be unfortunate in experi-

ment 2, where we wished to determine if Spanish early learners manage to discriminate English vowels like native English speakers by generating contrasting phonetic labels (which, of course, implies the establishment of new perceptual English vowel categories).

Accordingly, procedures that might hinder trace or context-coding strategies were evaluated. According to the PPT (Hessen and Schouten, 1992), increased "context variance"—obtained by increasing the range of stimuli presented—limits listeners' ability to hold context codes in memory. This motivated our decision to evaluate the influence of block structure. In previous applications of the AXB test, the trials testing various contrasts of interest have been presented in separate blocks. This experiment compared the discrimination of four Danish vowel contrasts when presented in separate blocks (small stimulus range) and when presented in one large randomized block (large stimulus range).

A second modification of the standard AXB format evaluated in this experiment was suggested by the results obtained in a study that focused on discrimination of stress placement (Dupoux *et al.*, 2001). The vowel stimuli were produced with either a relatively high or low fundamental frequency ($F0$). The $F0$ variations that occurred within trials were clearly audible, but were independent of category identity. We reasoned that noncriterial $F0$ variation would make context codes less useful for discrimination by diverting attention away from the dimensions that signaled vowel category identity. In addition, even if categorically relevant acoustic information became part of the context code for a stimulus, uncorrelated or misleading $F0$ information might place demands on memory that could negatively affect performance. Finally, $F0$ variation would likely also hamper the use of sensory traces.

The questions raised by this methodological experiment were as follows. First, would English monolinguals obtain lower scores in a nonblocked than blocked condition? Second, would the introduction of $F0$ variation serve to reduce scores? Third, would the two manipulations just mentioned affect performance on vowels perceived as within-category contrasts but not between-category contrasts (for which discrimination could be based on phonetic labels)?

A. Methods

1. Participants

Most of the 45 English monolinguals ("E-monolinguals") recruited for this experiment came from the southeastern portion of the U.S. (28 were from Alabama). None of the E-monolinguals reported being able to speak a language other than English, and none had previous exposure to Danish. Fifteen participants each (12 females) were randomly assigned to one of three conditions: "blocked" (the traditional approach in AXB tests), "mixed" (in which trials testing four contrasts were randomly presented in one large block), or "mixed/$F0$" (one large block, $F0$ variation introduced within trials).

Participants assigned to the three conditions differed little in mean age (blocked=27 years; mixed=28 years, mixed/$F0$=27 years). Three native speakers of Danish (mean age=34 years, 2 females) served as a reference group and took the test in the mixed/$F0$ condition. One Dane had been living in the U.S. for two months when tested, the other two for less than 2 weeks. All 48 participants were paid, and all passed a pure tone hearing screening at octave frequencies from 250 to 4000 Hz (25 dB).

2. Stimuli

Nonwords were used as stimuli because, in L1 development at least, correct discrimination rates may differ between real words and nonwords (e.g., Stager and Werker, 1997). Had real Danish words been used in the present experiment, the Danes might have enjoyed a lexical advantage not shared by the E-monolinguals.

The first author, a native Danish speaker then living in the United States, selected the four vowel contrasts to be examined based on published descriptions of Danish and English vowels (Grønnum, 1998; Steinlen, 2005) and his own intuitions about which Danish vowels might be difficult for E-monolinguals to discriminate. The inference that E-monolinguals will tend to hear Danish [uːˀ] and [ɔ̝ːˀ]) as two distinct English vowels (/u/ and /o/, respectively) was confirmed by a preliminary cross-language classification test in which ten Native English speakers classified the tokens of the eight Danish target vowels used to form the four contrasts. This Danish contrast ([uːˀ]-[ɔ̝ːˀ]) was therefore designated the "easy" contrast. Three other Danish contrasts were designated the "difficult contrasts" (D1, D2, D3) because, as expected, the two vowels in each contrast tended to be classified as a single English vowel: [e̞] and [e] 85% of the time, usually as English /ɪ/; [iːˀ] and [e̞ːˀ] 99% of the time, usually as English /i/; [y] and [ø] 88% of the time, usually as English /u/.

Different flanking consonants were used for the four Danish contrasts: /gl_s/ for the easy [uːˀ]-[ɔ̝ːˀ] contrast, /fl_s/ for the [e̞]-[e] contrast (D1), /kl_s/ for the [iːˀ]-[e̞ːˀ] contrast (D2), and /tj_s/ for the [y]-[ø] contrast (D3). These consonant sequences are legal in English. One might object that the initial /tj/ context for D3 is "foreign" for English speakers; however, its typical phonetic realization ([tᶜ]) closely resembles English [tʃ] (Grønnum, 1998).

The stimuli were produced by a native Danish male and recorded (Shure SM81 microphone, Tascam DA-302 DAT) in a sound booth. All stimuli used in the blocked and mixed conditions, and half of those used in the mixed/$F0$ condition, were produced in a declarative carrier sentence (*Det hed virkelig*_____*engang*, approximately "It really was _____ once"). The remaining half of the stimuli used in the mixed/$F0$ condition were produced in an interrogative sentence (*Hed det virkelig*_____*engang*?, "Was it really _____once?"), which yielded a relatively high $F0$ in the variable target words.

After the digitization (22.05 kHz, 16 bit), the CCVC stimuli were edited out and normalized for rms amplitude. Eight tokens for each of the eight Danish target vowels were selected. The intent was to select tokens that minimized acoustic phonetic variation within the eight tokens of each vowel category. The exception, of course, was $F0$ for tokens

selected for the mixed/$F0$ condition. For this condition, the four tokens per vowel category produced in the declarative frame had a mean $F0$ of 108 Hz at the vowel midpoint (range=102–117), and the four tokens per vowel produced in the interrogative frame had a mean $F0$ of 139 Hz (range =134–146 Hz).

3. Procedure

All 48 participants were tested on 256 AXB trials. Each of the four contrasts was tested by 64 trials with all possible AB combinations of the eight tokens of each vowel presented with an equal number of each trial type (AAB, ABB, BAA, and BBA). There was a complete and balanced rotation of the stimuli in the three positions in the triad; the three stimuli in each trial were always physically different. Participants in the blocked condition heard trials testing the four contrasts in separate, counterbalanced blocks (Latin square). For participants in the mixed condition, trials testing the four contrasts were presented in a single randomized block (with appropriate breaks). The same held true for the mixed/$F0$ condition which, in addition, introduced variation in $F0$.[1]

The stimuli were presented via headphones at a self-selected comfortable level to the participants, who were tested individually in a sound booth. The participants were told to push a button labeled "1-2" if they heard a match between the first and the second stimuli (i.e., A and X), and the "2-3" button if the second and third stimuli (X and B) matched. The interstimulus interval (ISI) for all trials was 1000 ms. A 1200-ms interval occurred between each response and the following trial.

To ensure that participants understood the task, the test was preceded by a practice block with 20 easy trials consisting of stimuli not used in the experiment. Feedback was provided during this block, but not during the experiment. Also unlike the experiment, trials could be replayed. All participants were required to respond correctly to at least 18/20 practice trials before proceeding to the experiment. All did so within four attempts.

The same 256 trials were presented in the blocked and the mixed conditions. In the mixed/$F0$ condition, half of tokens had relatively high (H) $F0$ and half had a relatively low (L) $F0$. Each trial in this condition manifested one of the following types of $F0$ pattern. In "no $F0$ change" trials, all three tokens had either H or L $F0$. In "neutral" trials, $F0$ could not influence responses because the $F0$ of the X stimulus differed from both A and B (HLH or LHL). In "congruent" $F0$ trials, $F0$ similarities "suggested" the correct response because the $F0$ level of the X stimulus was the same as that of the categorically same vowel. Conversely, in "incongruent" trials, the $F0$ suggested the wrong response, i.e., participants would always go wrong if they based their responses solely on $F0$ similarities among the three stimuli in a trial (e.g., HHL $F0$ in an [y]-[ø]-[ø] trial).

B. Results

As expected, the three Danes obtained high scores for the "easy" contrast ($M=100\%$) and also the three difficult contrasts (D1 $M=99\%$, D2 $M=96\%$, D3 $M=94\%$). The

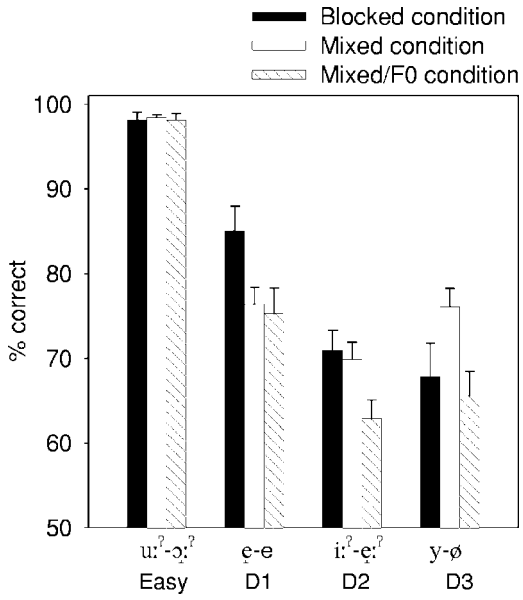


FIG. 1. Mean percent correct AXB discrimination of one "easy" and three "difficult" (D1, D2, D3) Danish vowel contrasts by three groups of English monolinguals assigned to one condition each. Error bars bracket 1 standard error.

E-monolinguals obtained high scores for the easy contrast ($M=98\%$) but, as shown in Fig. 1, substantially lower scores for the three difficult contrasts. Their scores for the difficult contrasts ranged from 63% correct for [iː?]-[e̞ː?], in the mixed/$F0$ condition, to 85% correct for [e̞]-[e], in the blocked condition. Single sample t tests showed that the E-monolinguals' scores for all three difficult contrasts significantly exceeded the chance level of 50% correct in all three testing conditions [$t(14)=4.4$ to 13.5, $p<0.001$].

The E-monolinguals' scores were submitted to an ANOVA in which condition (blocked, mixed, mixed/$F0$) served as a between-subjects factor and vowel contrast (four levels) was a repeated measure. The ANOVA yielded a significant interaction [$F(6,126)=4.2$, $p<0.001$], which was explored through tests of simple effects. The simple effect of condition was nonsignificant for the easy contrast [$F(2,42)=0.1$, $p=0.95$] but significant or marginally significant for the three difficult contrasts [D1 $F(2,42)=3.9$, $p=0.027$; D2 $F(2,42)=3.9$, $p=0.029$; D3 $F(2,42)=3.2$, $p=0.051$]. Tukey tests revealed that lower scores were obtained for D1 and D2 in the mixed/$F0$ condition than in the blocked condition ($p<0.03$). No other differences between conditions reached significance.[2] The lack of a significant difference between the blocked and mixed conditions indicated that simply combining the trials testing the four contrasts in a single randomized block was not sufficient to lower discrimination scores. This suggested that the difference between the blocked and mixed/$F0$ conditions was largely due to the manipulation of $F0$. The effect of $F0$ differences in the mixed/$F0$ condition was therefore explored further.

The E-monolinguals' scores for the easy contrast were at ceiling (98% to 99%) regardless of the $F0$ type of trial. However, for all three difficult contrasts, scores were higher in trials manifesting a "congruent" $F0$ type (D1 89%, D2 80%, D3 81%) than an "incongruent" $F0$ type (D1 63%, D2 50%,

D2 55%). Put another way, higher scores were obtained when reliance on *F*0 similarities would yield a correct response than when reliance on *F*0 would yield an incorrect response. Intermediate scores were obtained for trials with no *F*0 change (D1 80%, D2 67%, D3 68%) and trials with a "neutral" *F*0 type, which suggested neither the correct nor the incorrect response (D1 80%, D2 66%, D3 69%).

C. Discussion

This experiment aimed to develop AXB testing procedures that would yield near-ceiling scores for vowels heard as distinct phonemes (and thus given different labels) but near-chance discrimination scores for foreign vowels heard as instances of a single L1 category (and thus given the same label). This goal was only partially met.

As intended, the E-monolinguals assigned to all three conditions (testing formats) obtained near-ceiling scores for the easy Danish contrast (as did the native Danish reference group for the easy contrast and all three difficult contrasts). It is likely that the E-monolinguals obtained near-ceiling scores for the easy contrast by generating two different phonetic label codes when hearing the Danish vowels making up this contrast. Also as intended, the E-monolinguals obtained much lower scores for three difficult Danish contrasts.

Contrary to our intent, however, the E-monolinguals' scores for all three difficult contrasts were significantly above chance in all conditions. Importantly, though, significantly lower scores were obtained for two contrasts when the effect of both manipulations examined here (i.e., mixed block structure and *F*0 variation) were combined (mixed/*F*0 condition) than when standard AXB testing procedures were used (blocked condition). Assuming that the E-monolinguals based discrimination of the three difficult contrasts on context codes (Hessen and Schouten, 1992), the results suggested that the modifications hindered but did not completely prevent the use of such codes.

The effect of increasing the range of stimuli by presenting four contrasts randomly in a single block (mixed condition) rather than blocked by contrast (blocked condition) was nonsignificant. This suggested that the participants in this experiment did not generate context codes by positioning the stimuli in relation to other stimuli in the experimental context, but rather in relation to phonetic categories stored in long-term memory, a possibility also envisaged by Hessen and Schouten (1992). Context variance has previously been shown to be range dependent in experiments using synthetic stimuli (e.g., Macmillan *et al.*, 1988). The absence of a range effect in the present study, using natural rather than synthetic stimuli, raises the possibility that neither between-category nor within-category discrimination is range dependent for natural stimuli.

The significant differences between the blocked and mixed/*F*0 condition, but the lack of differences between the blocked and mixed conditions, suggested that differences between blocked and mixed/*F*0 were due largely to the *F*0 manipulation. Additional consideration of scores in the four *F*0 types in the mixed/*F*0 condition suggested that *F*0 variation might have been more effective in lowering scores had only the incongruent *F*0 type been used. The scores for the incongruent *F*0 type averaged just slightly above chance (56% correct) for the three difficult contrasts. The decision was made, therefore, to apply the procedures of the mixed/*F*0 condition in experiment 2 but only to use the incongruent and the neutral *F*0 types.

III. EXPERIMENT 2

This experiment compared the discrimination of English vowels by native Spanish adults who had begun learning English in the U.S. as children ("early learners") to that of monolingual native speakers of English ("E-monolinguals") and Spanish ("S-monolinguals"). One of the English vowel contrasts examined here was expected to be easy for the S-monolinguals to discriminate because the two vowels in it were likely to be heard as distinct Spanish vowels. However, the other three contrasts (designated D1, D2, D3) were expected to be difficult because the contrasting vowels were likely to be heard as instances of a single Spanish vowel category. Pilot tests with late learners (native Spanish speakers who had begun learning English as adults) on a range of English vowel contrasts suggested that the difficult contrasts examined here were the three most difficult English vowel contrasts for native Spanish speakers, at least for vowels spoken by the native English speaker who provided the stimuli.

In experiment 1, E-monolinguals with no prior exposure to Danish obtained significantly above-chance discrimination scores for Danish vowels that were heard as instances of a single English vowel category. This suggested that even the most difficult condition (i.e., the mixed/*F*0 condition, in which trials testing four contrasts were mixed together and *F*0 variation was added to the stimuli) did not prevent completely the use of low-level or context codes. Two additional steps were therefore taken in this experiment to further reduce percent correct scores. This was considered important to do given the nature of the question addressed by this experiment: Would the early learners' discrimination of English vowels more nearly resemble that of S-monolinguals or E-monolinguals?

One step involved the *F*0 manipulation. In experiment 1, the *F*0 manipulation lowered E-monolinguals' scores for difficult contrasts mainly when the *F*0 difference among the three stimuli in a trial was of the "incongruent" *F*0 type. Therefore, only the incongruent *F*0 type was included here, along with the "neutral" *F*0 type, to maintain some variability.

The second step was motivated by the evidence (Cowan and Morse, 1986) that vowel discrimination is poorer at an ISI of 0 ms than 250, 500,1000, or 2000 ms (both for within- and between-category trials).[3] We therefore included a 0-ms ISI as well as the 1000-ms ISI used in experiment 1. As described earlier, the PPT of Hessen and Schouten (1992) posits that label codes are generated rapidly, whereas context codes—essentially refinements of the phonetic code—are generated more slowly. If the ISI between items to be discriminated is too short, context codes might be stored in an incomplete and inaccurate form. This should render them less reliable for correct discrimination, thereby lowering

TABLE I. Mean characteristics of the three groups of participants in experiment 2. Ranges are in parentheses.

	Early learners	E-monolinguals	S-monolinguals
Gender	13 f, 7 m	11 f, 9 m	8 f, 12 m
Chronological ages	31 (20–53)	31 (21–45)	29 (19–46)
Years of education	16 (13–22)	17 (13–21)	10 (3–20)
Age of exposure to E[a]	6 (2–10)	…	…
Years of exposure to E[a]	25 (9–45)	…	…
Self-rated S ability[b]	6.0 (3–7)	…	…
Self-rated E ability[b]	6.6 (4–7)	…	…

[a]English.
[b]Self-rated Spanish (S) or English (E) ability, 1=poor, 7=good.

scores in cases where a label code is insufficient. Such interruption of the generation of context memory codes can probably be understood as informational backward recognition masking (see Watson, 2005). For ISI=0 ms, the second and third stimuli in the AXB triads will serve as masks for preceding stimuli. Subsequent masks have been shown to interrupt auditory processing of a stimulus. For example, Massaro (1970) found that identification of a test tone was impeded by an immediately following masking tone but improved with increasing ISI.

A. Methods

1. Participants

The 60 participants were recruited in Birmingham, AL. All were required to be between the ages of 19 and 55 years, to report normal hearing, and to pass a pure-tone hearing screening. Twenty participants each were assigned to one of three groups: "S-monolinguals" (Spanish monolinguals), "E-monolinguals" (English monolinguals), and "early-learners" (native Spanish adults who had begun learning English as an L2 in childhood). As shown in Table I, the three groups did not differ significantly in chronological age [$F(2,57)=0.51$, $p=0.6$], but did differ in years of formal education [$F(2,57)=30.5$, $p<0.001$]. A Tukey test indicated that the S-monolinguals had received less education than the E-monolinguals and early learners ($p<0.001$), who did not differ from one another.

E-monolinguals were required to be native speakers of American English, to have NE-speaking parents, to have never lived outside the U.S., and to be unable to speak a language other than English proficiently. The S-monolinguals were required to have been born and raised in a predominantly Spanish-speaking country (Mexico—16, Guatemala—1, Venezuela—1, Argentina—1, Colombia—1). Despite the fact that the S-monolinguals were living in Birmingham when tested (mean residence in the U.S.=3.8 years, s.d.=4.6), they had either very limited or no proficiency in English according to self-report. To verify this, the experimenter vocally asked candidates for inclusion in the S-monolingual group eight simple questions in English (e.g., *What day was it yesterday*? or *What kind of food do you like to eat*?). The S-monolinguals were able to provide a response—often just a single word—to an average of just 2.7 of the 8 questions (range=0–6).

The early learners were required to have learned Spanish as an L1 from native Spanish-speaking parents and to have learned English as an L2 in childhood. Sixteen early learners were born in a predominantly Spanish-speaking country and had immigrated to the U.S. in childhood. One was born in the U.S. but moved soon thereafter to Mexico before returning to the U.S. later in childhood. The remaining three were born and raised in a predominantly Spanish-speaking community in the U.S.; they reported being unable to understand English upon entering preschool or elementary school.

Participant characteristics are shown in Table I. The age at which the early learners had begun learning English was indexed by self-reported age of first exposure ("AOE") to English. Their English-language experience was indexed by self-reported years of exposure ("YOE") to English. All participants were paid.

2. Stimuli

A total of 64 English vowel stimuli were obtained in the following way. A list of nonwords was created by inserting eight English vowels ([i u ɪ eᴵ ɑ ʌ ʊ oʊ]) into a CC_C frame. A native English male produced the nonwords in declarative and interrogative sentence frames (*I say_____now again*; *Can you say_____now again*?). As in experiment 1, different consonants flanked the vowel pairs of interest: /gl_s/ for vowels forming the [i]-[u] contrast, /kl_s/ for the [ɪ]-[eᴵ] contrast, /tʃ_s/ for the [ɑ]-[ʌ], and /fl_s/ for [ʊ]-[oʊ] contrast. All of the consonant sequences are legal in both Spanish and English.

The sentences were recorded and the CCVC nonword stimuli edited as in experiment 1. For each vowel category, four tokens were taken from the declarative sentence frame and four from the interrogative sentence frame. $F0$ values were measured using an autocorrelation technique by placing a 40-ms window at the vowel midpoint. As shown in Table II, the $F0$ values were higher for stimuli taken from the interrogative than the declarative frame, and so the two sets of stimuli were designated the "high-$F0$" (or H) and "low-$F0$" (L) stimuli.

The frequencies of the first two vowel formants ($F1$, $F2$) of the English stimuli at the vowel midpoint have been plotted in Fig. 2 along with mean $F1$ and $F2$ values for the five vowels of Spanish [as spoken in a /b_t/ context by adult males, from Bradlow (1995)]. The perceived phonetic distance of vowels drawn from two languages cannot be accurately determined by means of distances in an $F1$-$F2$ space. However, the distances evident in Fig. 2 suggest that S-monolinguals might tend to classify the English [ɪ] and [eᴵ] tokens as instances of Spanish /e/, the [ɑ] and [ʌ] tokens as Spanish /a/, and the [ʊ] and [oʊ] tokens as Spanish /o/. The [i] and [u] tokens might be classified as instances of Spanish

TABLE II. Mean vowel duration and fundamental frequency ($F0$) values at vowel midpoint for stimuli produced in interrogative and declarative sentence frames. Ranges are in parentheses.

	$F0$ (Hz)		Duration (ms)	
	Interrogative	Declarative	Interrogative	Declarative
[i]	170 (164–178)	116 (110–120)	149 (143–158)	113 (109–117)
[u]	176 (170–182)	110 (107–111)	128 (122–133)	107 (97–121)
[ɪ]	168 (160–177)	117 (113–123)	110 (101–116)	94 (93–95)
[e^{ɪ}]	169 (163–174)	114 (113–114)	157 (153–161)	132 (121–140)
[ɑ]	163 (153–170)	121 (119–122)	173 (160–188)	151 (132–163)
[ʌ]	167 (160–170)	121 (113–132)	118 (110–124)	114 (110–121)
[ʊ]	172 (166–176)	136 (118–148)	133 (129–134)	113 (91–126)
[o^{ʊ}]	163 (158–172)	129 (124–139)	178 (170–l85)	162 (148–171)

/i/ and /u/, respectively, at least if the fronting of the English [u] did not cause Spanish speakers to perceive tokens of this category as being instances of Spanish /i/.[4]

Inferences based on formant values were formally evaluated following the discrimination experiment described below. The 64 English stimuli from the discrimination test were aurally presented a single time in quasi-random order to the S-monolinguals. Their task was to classify each token as being most like one of the five vowels of Spanish, "i", "e", "a", "o", or "u". (Data for one S-monolingual were lost due to equipment malfunction.) In agreement with the acoustic data, the English [ɪ] and [e^{ɪ}] tokens were usually classified as Spanish /e/ (128 and 124 judgments, respectively, out of a maximum 152); the [ɑ] and [ʌ] tokens were usually classified as Spanish /a/ (in 152 and 118 judgments); and the [ʊ] and [o^{ʊ}] tokens were usually classified as Spanish /o/ (116 and 134 judgments). The [i] and [u] tokens were classified as different vowels, Spanish /i/ (151 judgments) and /u/ (134 judgments).

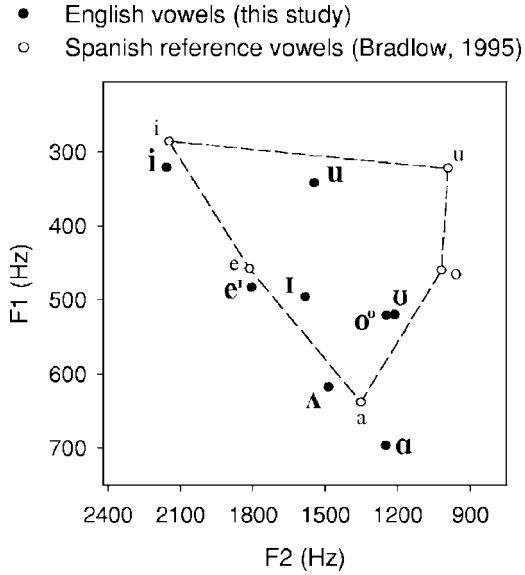


FIG. 2. Vowel chart showing first formant ($F1$) and second formant ($F2$) mean values at vowel midpoint of each of the eight English vowels used in experiment 2 (in bold). The mean values for Spanish are from Bradlow (1995).

3. Procedure

All 60 participants were tested individually in a sound booth where stimuli were presented via headphones at a self-selected level. The 64 trials testing each of the four contrasts were presented in one block of 256 trials. Half of the trials had the neutral $F0$ type (which suggested neither the correct nor the incorrect response), and the remaining trials had the incongruent $F0$ type, which suggested an incorrect response (e.g., LLH $F0$ in an [ɑ]-[ʌ]-[ʌ] trial). The block containing all 256 trials was presented two times in counterbalanced order with an ISI of either 0 or 1000 ms. In both ISI conditions, a 1200-ms interval occurred between each response and the following trial. A separate randomization was used for each participant, but was maintained across the two ISI conditions.

The 256 trials presented in both ISI conditions were preceded by 32 extra trials. Instructions, practice, and admission criterion were similar to those described in experiment 1. All test materials and instructions were translated and given in Spanish to the S-monolinguals. During the interval between blocks, participants responded to a language background questionnaire and produced sentences used for another study.

B. Results and discussion

1. Preliminary analysis

Lower scores were again expected for trials with the incongruent than the neutral $F0$ type. This expectation was tested by a (3) group × (4) vowel contrast × (2) ISI × (2)$F0$ type ANOVA examining the 16 mean scores obtained for each participant. Scores were indeed significantly lower

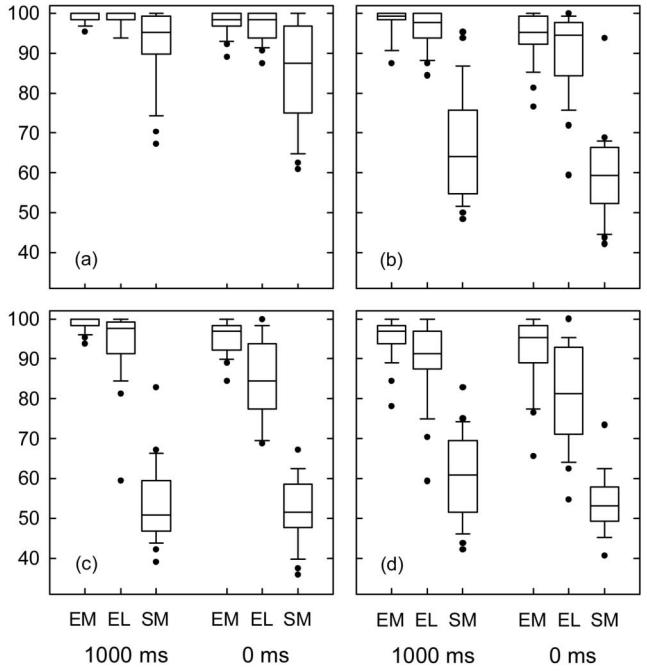


FIG. 3. Boxplots showing percent correct AXB discrimination obtained by English monolinguals (EM), early learners (EL), and Spanish monolinguals (SM). The four panels each display the results for one of the four vowel contrasts, i-u (a), ɑ-ʌ (b), ɪ-eɪ (c), and ʊ-oʊ (d), obtained by the three groups at a 1000 and a 0-ms ISI. Filled circles indicate individual outliers.

overall for trials with an incongruent than neutral *F*0 type [$F(1,57)=34.2$, $p<0.001$]. However, the lack of a group $\times$vowel contrast$\times F0$ type interaction [$F(6,171)=0.9$, $p=0.53$] suggested that the relative effect of *F*0 type on the four contrasts was similar across the three groups. In all subsequent analyses, therefore, scores were averaged over the two *F*0 types.

2. Between-group differences

As shown in Fig. 3, the mean percent correct scores were higher overall for the easy contrast ([i]-[u], $M=95\%$ correct) than for the three difficult contrasts ([ɪ]-[eɪ] 80%, [ɑ]-[ʌ] 84%, [ʊ]-[oʊ] 79%). The early learners' scores were slightly lower ($M=91\%$) than the E-monolinguals' ($M=96\%$), but substantially higher than the S-monolinguals' ($M=65\%$). Lower scores were obtained, on average, for ISI=0 ms than ISI=1000 ms ($M=82$ vs. 87%). Of greatest interest, the S-monolinguals' scores for the three difficult contrasts at ISI=0 ms were all near chance (<60% correct). A series of single sample *t* tests indicated that S-monolinguals' scores for one difficult contrast, [ɪ]-[eɪ], did not significantly exceed the chance level of 50% at ISI =0 ms [$t(19)=1.2$, $p=0.24$] or ISI=1000 ms [$t(19)=1.6$, $p=0.12$].

The eight scores obtained for each participant (4 contrasts $\times$2 ISIs) were examined in a mixed-design (3) group $\times$(2) ISI$\times$(4) vowel contrast ANOVA. All three main effects reached significance [group $F(2,57)=175.3$, $p<0.001$; ISI $F(1,57)=53.6$, $p<0.001$; vowel contrast $F(3,171)=95.8$, $p<0.001$] as did the three-way interaction [$F(3,171)=2.4$, $p=0.033$], which was explored through simple effects tests. As seen in Fig. 3, an important source of the interaction was a differing pattern of between-group differences in the two ISI conditions. The group effect was significant for all four vowel contrasts at both ISI =1000 ms [$F(2,57)=9.7$ to 180.3, $p<0.001$] and ISI =0 ms [$F(2,57)=15.4$ to 158.6, $p<0.001$]. Tukey tests ($\alpha=0.05$) revealed that at ISI=1000 ms, the S-monolinguals' scores were significantly lower for all four contrasts than those obtained by the E-monolinguals and early learners, who did not differ significantly. For ISI=0 ms, the S-monolinguals' scores were lower for all four contrasts than the E-monolinguals'. Importantly, the early learners' scores were lower than the E-monolinguals' for two difficult contrasts, [ɪ]-[eɪ] and [ʊ]-[oʊ]. Although the early learners' scores for difficult contrasts were closer to the E-monolinguals' than to the S-monolinguals' (see Fig. 3), these differences suggested that the early learners' perception of English vowels was not "functionally equivalent" (see Flege and MacKay, 2004) to that of the E-monolinguals.

The lower scores obtained for the easy contrast by the S-monolinguals than by the early learners and E-monolinguals might seem surprising given that all but one S-monolingual classified the English /i/ and /u/ stimuli as instances of two different Spanish vowels (*viz.* Spanish /i/ and /u/). The S-monolinguals' lower scores might be attributed to acoustic phonetic differences between the /i/s and /u/s found in Spanish and English (see Fig. 2).[5] Note, however, that a much smaller difference existed between the E-monolinguals and S-monolinguals for the easy contrast ($\Delta=7\%$) than for the three difficult contrasts ($\Delta=31\%-45\%$).

The simple effect of contrast was significant for all six group$\times$ISI combinations [$F(3,57)=4.3$ to 65.5, $p<0.01$], but the pattern of between-contrast differences varied as a function of group and ISI. The E-monolinguals obtained lower scores for [ʊ]-[oʊ] than [i]-[u] and [ɪ]-[eɪ] at ISI =1000, but [ʊ]-[oʊ] and [ɪ]-[eɪ] did not differ at ISI=0. The S-monolinguals obtained higher scores for [i]-[u] than for all three difficult contrasts at both ISI=0 and ISI=1000, and higher scores for [ɑ]-[ʌ] than [ɪ]-[eɪ] at ISI=1000 ms. For early learners, higher scores were obtained for [i]-[u] than all three difficult contrasts at ISI=0, for [i]-[u] than [ɪ]-[eɪ] and [ʊ]−[oʊ] at ISI=1000 ms, and for [ɑ]-[ʌ] than [ʊ]−[oʊ] at both ISIs.

Tests of the simple effect of ISI revealed that significantly lower scores were obtained at ISI=0 than ISI =1000 ms for 10 of the 12 contrast$\times$group combinations [$t(19)=2.1$ to 4.8, $p<0.05$]. The two exceptions were the highest and the lowest scores obtained, *viz.* the score obtained for the easy contrast by the E-monolinguals, and for the [ɪ]-[eɪ] contrast by the S-monolinguals. The first exception might well have been a ceiling effect (with mean values in the two ISI conditions of 98% and 99% correct), the second a floor effect ($M=52\%$ and 54%).

The effect of ISI for the E-monolinguals—for whom all trials were presumably between-category trials—was unexpected. This finding suggests that ISI influences both the between- and within-category discrimination of natural vowels [as has previously been found for synthetic vowels, see Cowan and Morse (1986)]. Perhaps between-category dis-

TABLE III. Mean characteristics of the four early learners who obtained the highest discrimination scores, and the four obtaining the lowest scores. Ranges are in parentheses.

Variable	Context	4 Highest	4 Lowest
Age[a]	…	36 (22–50)	31 (20–39)
Age of exposure to E	…	3 (2–5)	8 (5–10)
Years of exposure to E	…	33 (21–45)	23 (10–34)
Years of S education[b]	…	0 (0)	3 (0–4)
Years of E education[b]	…	18 (13–20)	13 (10–17)
% E use—last 1[c]	overall	75 (50–90)	68 (50–90)
% E use—last 5[c]	overall	65 (50–80)	70 (50–90)
% E use—last 5[c]	at home	75 (30–100)	58 (0–90)
% E use—last 5[c]	with friends	90 (70–100)	60 (50–70)
% E use—first 5[c]	in general	55 (50–70)	25 (10–40)
% E use—first 5[c]	at home	48 (40–50)	8 (0–30)
% E use—first 5[c]	with friends	88 (80–100)	43 (0–90)
% S-accented E[c]	last 5	15 (0–30)	30 (10–60)
% S-accented E[c]	first 5	20 (0–70)	33 (10–70)
E ability[d]	…	7 (7–7)	5 (4–6)

[a]Chronological age, in years.
[b]Years of education in Spanish (S) or English (E) medium schools.
[c]Self-estimated percentage in the last 1, last 5, or first 5 years in the U.S.
[d]Self-rated ability (1=poor, 7=good).

crimination was based to some extent on the ISI-sensitive context codes in addition to phonetic label codes.

3. Individual differences

The large difference in scores obtained by the E-monolinguals and S-monolinguals suggests that the AXB test used here was phonetically sensitive, perhaps by virtue of hindering use of context codes (Hessen and Schouten, 1992). In the ISI=0 ms condition, the scores obtained by individual E-monolinguals and S-monolinguals were virtually nonoverlapping for the three difficult contrasts. (The exception was a single S-monolingual outlier for [ɑ]-[ʌ][6] and [ʊ]-[oʊ].) Likewise, the scores obtained for individual early learners and S-monolinguals showed virtually no overlap at ISI=0 ms (the exceptions being one outlier in each group for [ɑ]-[ʌ] and [ʊ]-[oʊ]), which provided direct evidence of perceptual learning by nearly all early learners.

Given the theoretical importance of this finding, it is worth considering in greater detail the scores obtained by individual early learners at ISI=0 ms. Averaged over all three difficult contrasts, the scores obtained for the 20 E-monolinguals ranged from 81% to 99% correct, as compared to a range of 64% to 98% for the early learners. The four lowest scores obtained for early learners fell below the E-monolinguals' range (but were nonetheless higher than those obtained for all but one S-monolingual), whereas the four highest-scoring early learners all exceeded the mean score obtained for the E-monolingual group.

Examination of the participants' questionnaires revealed, not surprisingly, that all four of the lowest-scoring early learners reported less proficiency in English than all four of the highest-scoring early learners. In fact, of all 20 early learners, only the four learners who scored below the native English range did not assign themselves the maximum value of "7" when self-rating their ability to speak English.

What enabled some early learners to discriminate English vowels so much better than others? As summarized in Table III, three of the lowest-scoring early learners were exposed to English later in life (i.e., had a later AOE) than all four highest-scoring early learners. Three of the highest-scoring learners reported using English more with friends in the past 5 years than all four of the lowest-scoring learners. And all four highest-scoring learners reported more use of English during their first 5 years of exposure to English than all four of the lowest-scoring learners. Finally, three of the four highest-scoring learners reported having attended school in the U.S. longer than all four of the lowest-scoring early learners. Although these data are only suggestive given the small number of participants (four per subgroup), the pattern seen here suggests that differences in the age of first exposure and amount of L2 use (especially during the first 5 years of L2 exposure) were responsible for important individual differences among the early learners.

IV. GENERAL DISCUSSION

This study evaluated the discrimination of English vowels by native Spanish adults who were first exposed to English in childhood (early learners). The study used a variant of the categorial AXB discrimination test that appeared to hinder vowel discrimination based on low-level or context memory codes (as opposed to language-specific phonetic codes, see experiment 1). The findings relate to the more general question of the extent to which the speech perception system remains plastic following attunement to the L1 sound system in early childhood.

The two English vowels in each of three "difficult" contrasts were both heard by Spanish monolinguals (S-monolinguals) as instances of a single Spanish vowel category. These contrasts were expected to be difficult to discriminate for the S-monolinguals and also for the early

learners unless they had developed new long-term memory representations (phonetic categories) for one or both vowels in the contrasts. The two vowels making up a fourth contrast were heard by S-monolinguals as distinct Spanish vowels. This contrast was expected to be easy to discriminate for Spanish speakers with or without perceptual learning of English vowels.

The AXB test was administered at two ISIs, 1000 vs. 0 ms, and was highly phonetically sensitive in the ISI =0 ms condition in that listeners with English-specific phonetic categories (the E-monolinguals) obtained scores near ceiling for the difficult English contrasts whereas listeners without such categories (the S-monolinguals) obtained scores that were at or near chance. Given this characteristic of the discrimination test, it was well suited to reveal perceptual learning by the early learners. The early learners obtained significantly higher scores than the S-monolinguals at both ISIs and did not differ from the E-monolinguals for any contrast at ISI=1000 ms. However, the early learners obtained lower scores than the E-monolinguals for two difficult contrasts at ISI=0 ms, suggesting that their long-term memory representations for English vowels were not identical to the E-monolinguals'.

Significant differences between early learners and L2 native speakers have often been interpreted as evidence for a loss of perceptual plasticity following attunement to the L1 sound system (e.g., Bosch *et al.*, 2000; Pallier *et al.*, 1997; Sebastián-Gallés and Soto-Faraco, 1999). However, as mentioned in the Introduction, there are *a priori* reasons to expect the perceptual systems of early learners to differ from those of L2 native speakers, and thus to expect differences in performance in sensitive tests such as the one used here. It is for this reason that including the S-monolingual group was important for considering whether the early learners' perceptual system remained plastic. Except for one outlier in each group, the scores obtained for individual early learners and S-monolinguals for the three difficult contrasts were nonoverlapping at ISI=0 ms. In our opinion, this finding provides evidence of extensive perceptual learning in early learners and is inconsistent with the view that the speech perception system loses plasticity (e.g., Pallier *et al.*, 1997; Sebastián-Gallés and Soto-Faraco, 1999), and that new phonetic categories become unlikely (Bosch *et al.*, 2000) following attunement to the L1 sound system.

Research with early learners has tended to focus on between-group differences with a view to determining whether the speech perception system does, or does not, remain plastic. However, as usually holds true in L2 perception research (e.g., Sebastián-Gallés *et al.*, 2005), we observed important individual differences among the early learners tested here. This suggests that an important question to ask in future research is: "What factors influence early learners' L2 speech perception?" A detailed analysis of the scores obtained by individual early learners identified several factors that will need to be examined in greater detail in future research. Just four (of 20) early learners obtained average scores for the three difficult contrasts that fell below the E-monolingual range at ISI=0 ms. These four early learners differed from the four highest-scoring early learners in several potentially important ways: The lowest scoring early learners reported using English less often, especially during their first years of English use, than the highest scoring early learners. This raises the question of how much L2 input early learners must receive, and when, if they are to show a nativelike perception of L2 phonetic segments.

A related question is whether the *kind* L2 input early learners receive affects their performance. For example, in a study examining Spanish speakers who had begun learning their L2 (Catalan) in Barcelona by the age of 4 years, Sebastián-Gallés *et al.* (2005) obtained scores ranging from chance level to values well above the native speaker mean for a vowel contrast not found in the L1 (Spanish). The authors noted that the numerous native Spanish speakers in Barcelona tend to realize both Catalan /e/ and /ɛ/ as an [e]-quality vowel, raising the possibility that the intersubject variability might be explained by how often the early learners had heard their L2 spoken by fellow native speakers of Spanish who neutralize the Catalan contrast.

Language use patterns, of course, are related to language dominance. The Spanish-Catalan bilinguals examined by Sebastián-Gallés and Soto-Faraco (1999) reported higher proficiency in their L1 Spanish than their L2 Catalan. This might explain why they differed from Catalan native speakers in perceiving Catalan speech sounds. It is noteworthy that L2 perception by clearly L2-dominant early learners examined by Flege and MacKay (2004) did not differ from that of L2 native speakers. If the L1 phonetic subsystem in bilinguals becomes weak through lack of use, the possibility exists that effects of cross-language phonetic interference will be minimized, and performance in the L2 will eventually become nativelike.

Indirect support for this hypothesis comes from recent research on Korean adults adopted by French families between the ages of 2 and 9 years, which suggested that early attunement to the L1 sound system may disappear following years of disuse (Pallier *et al.*, 2003; Ventureyra *et al.*, 2004). This interesting research raises the possibility that the L2 performance of early bilinguals with little or no continued L1 use for many years might become indistinguishable from that of native speakers. If so, it would suggest that perceptual learning is limited by language use or dominance patterns rather than a loss of plasticity arising from maturational constraints or irreversible effects of prior learning on subsequent learning. This hypothesis would be strongly supported if clearly L2-dominant early bilinguals with little L1 use could be shown to perform like native speakers in their L2, but show below-native performance in their L1.

The highest-scoring and lowest-scoring early learners also differed in terms of age of first exposure to English in the U.S. (M=3 vs. 8 years). Prior research of "age" effects has focused on differences between early and late L2 learners. However, the differences between the highest- and lowest-scoring early learners in experiment 2 suggested that age effects might exist among participants traditionally classified as "early" learners. Given the small sample size, however, there is a need for additional research, in which factors confounded with age are controlled. If L2 discrimination is indeed better for those first exposed to the L2 at the age of 3

than 8 years, this might be due to a differing effect of L1-on-L2 phonetic interference arising from differences in the state of development of L1 phonetic categories (e.g., Flege, 2003).

In summary, support for the existence of extensive perceptual plasticity in early learners was provided by the finding that most early learners showed nativelike discrimination of three difficult English vowel contrasts which were discriminated at near-chance levels by Spanish monolinguals. A few early learners obtained scores below the native English range, indicating that early exposure to an L2 is insufficient in itself to guarantee nativelike discrimination of L2 vowels. Additional research will be needed to clarify which factors influence early learners' perception of L2 phonetic segments, and to what extent bilinguals can show nativelike performance in both of their languages.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. DC000257. Additional support was provided to the first author by the Danish Research Council for the Humanities and by the Oticon Foundation. We wish to thank A. Bradlow and three anonymous reviewers for very helpful comments on our original submission, and O.-S. Bohn and G. Morrison for comments on a previous draft.

[1]In the blocked condition, eight extra trials were presented before each 64-trial block as a "warm-up." In mixed and mixed/*F*0 conditions, the 256-trial block was preceded by 32 extra trials. All four vowel contrasts were presented in the extra trials, which were not analyzed. All participants received a different randomization of the stimuli.

[2]The simple effect of contrast was significant for all three groups (F values ranging from 43.6 to 61.5, $p<0.001$). Tukey tests showed that higher scores were obtained for the easy contrast than for all three difficult contrasts in each testing condition ($p<0.05$). For participants assigned to the blocked and mixed/*F*0 conditions, higher scores were obtained for difficult 1 than for difficult 2 and 3; in the mixed condition, higher scores were obtained for difficult 1 and 3 than for difficult 2.

[3]The effect for within-category trials might be explained by a lack of time to build up context codes in the 0-ms condition. The effect for between-category trials is surprising assuming that discrimination was based on label codes. However, because the vowel stimuli were drawn from a synthetic continuum with little acoustic difference between the steps, the participants might have relied on context codes in addition to label codes in between-category discrimination. Also, listeners may depend more on context codes for synthetic than natural stimuli (see Hessen and Schouten, 1992).

[4]The mean $F2$ of the [u] tokens in experiment 2 was 1545 Hz, which is somewhat higher (indicating fronting) than previously published $F2$ values for American English [u], e.g., 1238 Hz in a /bVt/ context (Bradlow, 1995) or 870 Hz in an /hVd/ context (Peterson and Barney, 1952). Possibly, the preceding [l] had a fronting effect on [u].

[5]Another possibility is that the S-monolinguals' scores for the easy contrast were reduced somewhat by anxiety because they were unsure about the correct response in the majority of trials (the "difficult" trials). Whatever the explanation, it is worth noting that in previous cross-language research, similar levels of AXB scores have been obtained for instances of a clear "two-category" consonant contrast (Best *et al.*, 2001, p. 787).

[6]Inspection of the classification results revealed that this S-monolingual participant—unlike the other S-monolinguals—consistently classified English [ɑ] and [ʌ] as different Spanish vowels (*viz.* [a] and [e]), which likely explains his high discrimination scores for this English contrast.

Best, C. T., Faber, A., and Levitt, A. G. (**1996**). "Perceptual assimilation of non-native vowel contrasts to the American English vowel system," J. Acoust. Soc. Am. **99**, 2602.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," J. Acoust. Soc. Am. **109**, 775–794.

Bosch, L., Costa, A., and Sebastián-Gallés, N. (**2000**). "First and second language vowel perception in early bilinguals," Eur. J. Cogn. Psychol. **12**, 189–221.

Bradlow, A. R. (**1995**). "A comparative acoustic study of English and Spanish vowels," J. Acoust. Soc. Am. **97**, 1916–1924.

Cowan, N., and Morse, P. A. (**1986**). "The use of auditory and phonetic memory in vowel discrimination," J. Acoust. Soc. Am. **79**, 500–507.

Dupoux, E., Peperkamp, S., and Sebastián-Gallés, N. (**2001**). "A robust method to study stress "deafness"," J. Acoust. Soc. Am. **110**, 1606–1618.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York Press, Timonium, MD), pp. 233–277.

Flege, J. E. (**2003**). "Assessing constraints on second-language segmental production and perception," in *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. O. Schiller (Mouton de Gruyter, Berlin), pp. 319–355.

Flege, J. E., and MacKay, I. R. A. (**2004**). "Perceiving vowels in a second language," Stud. Second Lang. Acquis. **26**, 1–34.

Flege, J. E., MacKay, I. R. A., and Meador, D. (**1999**). "Native Italian speakers' perception and production of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

Flege, J. E., Munro, M. J., and Fox, R. A. (**1994**). "Auditory and categorical effects on cross-language vowel perception," J. Acoust. Soc. Am. **95**, 3623–3641.

Fujisaki, H., and Kawashima, T. (**1970**). "Some experiments on speech perception and a model for the perceptual mechanism," Annu. Rep. Eng. Rese. Insti. (Tokyo) **29**, 207–214.

Grønnum, N. (**1998**). *Fonetik of Fonologi—Almen og Dansk (Phonetics and Phonology—General and Danish)* (Akademisk Forlag, Copenhagen).

Grøsjean, F. (**1989**). "Neurolinguists, Beware! The bilingual is not two monolinguals in one person," Brain Lang **36**, 3–15.

Harnsberger, J. D. (**2001**). "On the relationship between identification and discrimination of non-native nasal consonants," J. Acoust. Soc. Am. **110**, 489–503.

Hessen, A. v., and Schouten, M. E. (**1992**). "Modeling phoneme perception: II. A model of stop consonant discrimination," J. Acoust. Soc. Am. **92**, 1856–1868.

Iverson, P., Kuhl, P. K., Yamada, R. A., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Kluender, K. R., Lotto, A. J., Holt, L. L., and Bloedel, S. L. (**1998**). "Role of experience for language-specific functional mappings of vowel sounds," J. Acoust. Soc. Am. **104**, 3568–3582.

Kuhl, P. K. (**2000**). "A new view of language acquisition," Proc. Natl. Acad. Sci. U.S.A. **97**, 11850–11857.

Kuhl, P. K., and Iverson, P. (**1995**). "Linguistic experience and the 'Perceptual Magnet Effect'," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 121–154.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., and Lindblom, B. (**1992**). "Linguistic experience alters phonetic perception in infants by 6 months of age," Science **255**, 606–608.

Long, M. H. (**1990**). "Maturational constraints on language development," Stud. Second Lang. Acquis. **12**, 251–285.

Macmillan, N. A., Goldberg, R. F., and Braida, L. D. (**1988**). "Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua," J. Acoust. Soc. Am. **84**, 1262–1280.

Massaro, D. W. (**1970**). "Perceptual auditory images," J. Exp. Psychol. **85**, 411–417.

McClelland, J. L. (**2001**). "Failures to learn and their remediation: A Hebbian account," in *Mechanisms of Cognitive Development: Behavioral and Neural Approaches*, edited by J. L. McClelland and R. S. Siegler (Erlbaum, Mahwah, NJ), pp. 97–211.

Pallier, C., Bosch, L., and Sebastián-Gallés, N. (**1997**). "A limit on behavioral plasticity in speech perception," Cognition **64**, B9–B17.

Pallier, C., Colomé, A., and Sebastián-Gallés, N. (**2001**). "The influence of native-language phonology on lexical access: Exemplar-based *vs*. abstract lexical entries," Psychol. Sci. **12**, 445–449.

Pallier, C., Dahaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., and Mehler, J. (**2003**). "Brain imaging of language plasticity in adopted

adults: Can a second language replace the first?" Cereb. Cortex **13**, 155–161.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 585–594.
Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.
Polka, L., and Bohn, O.-S. (**1996**). "A cross-language comparison of vowel perception in English-learning and German-learning infants," J. Acoust. Soc. Am. **100**, 577–592.
Scovel, T. (**2000**). "A critical review of the critical period research," Annu. Rev. Appl. Linguist. **20**, 213–223.
Sebastián-Gallés, N., and Soto-Faraco, S. (**1999**). "Online processing of native and non-native phonemic contrasts in early bilinguals," Cognition **72**, 111–123.
Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (**2005**). "The influence of initial exposure on lexical representation: comparing early and simultaneous bilinguals," J. Mem. Lang. **52**, 240–255.
Stager, C. L., and Werker, J. F. (**1997**). "Infants listen for more phonetic detail in speech perception than in word-learning tasks," Nature (London) **388**, 381–382.
Steinlen, A. (**2005**). *The Influence of Consonants on Native and Non-native Vowel Production*, Language Development (Gunter Narr Verlag, Tübingen).
Ventureyra, V. A., Pallier, C., and Yoo, H.-Y. (**2004**). "The loss of first language phonetic perception in adopted Koreans," J. Neurolinguist. **17**, 79–91.
Watson, C. S. (**2005**). "Some comments on informational masking," Acta. Acust. Acust. **91**, 502–512.
Werker, J. F., and Tees, R. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.
Yamada, R. A. (**1995**). "Age and acquisition of second language speech sounds. Perception of American English /r/ and /l/ by native speakers of Japanese," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, edited by W. Strange (York, Timonium, MD), pp. 305–320.

Phonological versus phonetic cues in native and non-native listening: Korean and Dutch listeners' perception of Dutch and English consonants

Taehong Cho[a)]
Hanyang University, Division of English Language and Literature, Seoul (133-791), Korea and Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

James M. McQueen
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 28 July 2005; revised 14 February 2006; accepted 1 March 2006)

We investigated how listeners of two unrelated languages, Korean and Dutch, process phonologically viable and nonviable consonants spoken in Dutch and American English. To Korean listeners, *released* final stops are nonviable because word-final stops in Korean are never released in words spoken in isolation, but to Dutch listeners, *unreleased* word-final stops are nonviable because word-final stops in Dutch are generally released in words spoken in isolation. Two phoneme monitoring experiments showed a phonological effect on both Dutch and English stimuli: Korean listeners detected the unreleased stops more rapidly whereas Dutch listeners detected the released stops more rapidly and/or more accurately. The Koreans, however, detected released stops more accurately than unreleased stops, but only in the non-native language they were familiar with (English). The results suggest that, in non-native speech perception, phonological legitimacy in the native language can be more important than the richness of phonetic information, though familiarity with phonetic detail in the non-native language can also improve listening performance. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188917]

PACS number(s): 43.71Hw, 43.71Es [ARB] Pages: 3085–3096

I. INTRODUCTION

In the first months of life, infants are typically able to discriminate acoustic-phonetic differences irrespective of whether the phonetic differences are phonologically contrastive in the ambient language or not (see, e.g., Jusczyk, 1997, for a review). By the end of their first year, however, their sensitivity to speech sound contrasts is generally refined according to the phonological properties of the ambient, native language (e.g., Werker *et al.*, 1981; Werker and Lalonde, 1988; see Cutler and Broersma, 2005, for a review). Although the timing of these refinements varies (Polka *et al.*, 2001), the general developmental pattern is that sensitivity to phonetic differences which are phonologically noncontrastive in the native language becomes attenuated relative to sensitivity to contrastive phonetic differences, and that the degree of such an attenuation is typically correlated with age (e.g., Flege, 1995). Eventually, perception of speech sounds by adults is highly biased or modulated by their experience with the phonological systems of their native language (there are effects, e.g., of phoneme repertoire, phonotactics, allophonic distribution, rhythmic structure and morphophonemic alternations; see e.g., Best, 1995; Best and Strange, 1992; Broersma, 2005; Costa *et al.*, 1998; Cutler *et al.*, 1986; Cutler and Otake, 1994; Flege, 1995; Flege and Hillenbrand, 1986; Hallé *et al.*, 1999; Kuhl and Iverson, 1995; Otake *et al.*, 1993; Weber, 2001; Weber and Cutler, 2006). Adult native Japanese listeners, for example, generally perceive both contrastive English sounds /r/ and /l/ as the flap /ɾ/, which is the sound in their native language that is phonetically closest to the two sounds (e.g., Best *et al.*, 1988; Best and Strange, 1992; see also Best, 1995). This is a case in which two contrastive categories in the non-native language are "assimilated" to the same phonological (or phonetic) category in the listener's native language.

Another example of native-language influences comes from differential use of phonetic cues by native and non-native listeners in the perception of contrastive sounds. Flege and Hillenbrand (1986), for instance, showed that in identifying fricatives as /s/ or /z/ (as in *peace* and *peas*) in American English, native English listeners used two well-known phonetic cues to the syllable-final voicing contrast, that is, fricative duration (which is longer for /s/ than for /z/) and preceding vowel duration (which is shorter before /s/ than before /z/; e.g., Denes, 1955, Raphael, 1972; see Watson, 1983 for a review). Non-native listeners of English, specifically, Swedish and Finnish listeners, who have no phonemically contrastive /s/-/z/ pair in their native language, but do have contrastive long and short vowels in their native language, used only vowel length differences to differentiate /s/ from /z/. According to Flege and Hillenbrand, listeners of Swedish and Finnish might have "reinterpreted" the role of phonologically contrastive vowel duration in their native language as a cue to the voicing contrast in non-native listening. Dutch listeners, however, appear not to use vowel duration in identification of voiced versus voiceless English syllable-final fricatives (Broersma, 2005). This may be because, in Dutch, /s/ and /z/ are phonemically contrastive, but not in

[a)]Electronic mail: tcho@hanyang.ac.kr

syllable final position. Dutch listeners may therefore have learned to use cues other than vowel duration in perceiving the fricative voicing distinction.

What has emerged from these studies on native versus non-native listening is that speech perception is tuned according to language experience, resulting in a "phonological reorganization of speech sound percepts" (Cutler and Broersma, 2005), and that one of the driving forces behind this phonological reorganization is the various language-specific constraints that arise within the phonological system of the listener's native language. It is thus not surprising, as Cutler and Broersma (2005) noted, that improving performance of adult listeners in the perception of non-native sounds has been found to be very difficult although training can improve performance to some extent (e.g., Logan *et al.*, 1991; Lively *et al.*, 1993, 1994; Bradlow *et al.*, 1997).

In the present study, we further explore how listening is tuned by language experience. We investigate how listeners of two unrelated languages, Korean and Dutch, process sounds which are phonologically viable or nonviable in their native language. In Korean, stops in syllable- or word-initial position are lenis, fortis, or aspirated, and their phonetic cues are generally present in the release portion and at the beginning of the following vowel (Cho *et al.*, 2002; Kim *et al.*, 2002). In coda position, however, stops are unreleased, at least when the syllable or word is produced in isolation. Coda stops in Korean can be realized either with or without audible releases. When the coda stop is followed by a vowel-initial morpheme (due to morpheme concatenation) or by a vowel-initial word (due to prosodic grouping), it is resyllabified as the onset of the following syllable, and is released with an audible release noise. When followed by another stop, however, coda stops can be released with only an inaudible brief transient noise. Kim and Jongman (1996) reported that when underlying stem-final /t/ was followed by the velar stop /k/ in Korean, it was released with a brief transient 83% of the time. These events were difficult to perceive, but appeared in acoustic analyses. Henderson and Repp (1982) observed similar patterns for English stop sequences, and argued that what are traditionally known as "unreleased" stops are often produced with an inaudible articulatory release gesture. Korean coda stops followed by another stop thus tend, in this sense, to be "unreleased." Korean coda stops in isolated syllables (or utterance finally), however, are unreleased both auditorily and articulatorily. This means that the three-way manner contrast found in onset position in Korean is completely neutralized, without even subtle acoustic differences distinguishing among the three types of stop (Kim and Jongman, 1996). This behavior is often considered to be the outcome of an obligatory phonological rule which states that stops are unreleased when followed by a word boundary or by another obstruent within the word (e.g., Kim-Renaud, 1976; Martin, 1992).

In Dutch, however, voiceless coda stops in word-final position are generally released when words are produced in isolation (cf., Warner *et al.*, 2004; Ernestus and Baayen, in press). Like Korean, Dutch has released stops in syllable-initial (prevocalic) positions. The question we address here, therefore, is whether and how the position-specific phonological difference between the two languages influences speech perception. Released coda stops in isolated syllables are phonologically nonviable in Korean; unreleased stops in this environment are phonologically nonviable in Dutch. It is important to emphasize that we use the term "phonological" to refer to language-specific sound patterns which are driven by obligatory phonological rules. The term "phonological knowledge" is thus used to refer to the native listener's knowledge not only about the native language's phoneme system but also about sound patterns determined by phonological rules. If the way native listeners perceive non-native speech sounds is guided by their native phonological knowledge, listeners should find it harder to process non-native sounds that deviate from native language phonological constraints (phonologically nonviable sounds) than to recognize those that are in conformity with those native constraints (phonologically viable sounds). This will be referred to as the *phonological-superiority hypothesis*.

There is a competing hypothesis, however. Released stops are produced with two distinct sets of acoustic-phonetic cues (those associated with the VC formant transition and those associated with the release burst) whereas unreleased stops are cued solely by the VC formant transition information. It has widely been agreed that the release burst is one of the most important cues to the identity of a stop, such that the presence of an audible release burst (in addition to formant transitions) increases the perceptual recoverability of the stop consonant (e.g., Stevens and Blumstein, 1978; Mattingly, 1981; Silverman, 1995; Smits *et al.*, 1996; Steriade, 1997; Kochetov, 2001; Zsiga, 2003; Wright, 2004; Flemming, 2005). That is, the information available to the listener is richer with than without the release burst.

In a perceptual study on Russian stops in syllable- or word-final coda positions, for example, Kochetov (2001) has shown that the identification of plain and palatalized coronal coda stops in Russian was significantly faster in the presence than in the absence of an audible release burst. With respect to stops in syllable-onset positions, the data reported by Smits *et al.* (1996) showed that in classifications of Dutch stops (/p/ vs /t/ vs /k/), Dutch listeners relied more on the release bursts than they did on the Consonant-Vowel (CV) formant transitions, when the stimuli contained a mismatch between the release burst and the CV formant transition. Similarly, Stevens and Blumstein (1978) showed that the presence of noise bursts at stop onset enhanced overall consonant classification in English. The perceptual importance of the release burst as a cue to stop identity has also been used as phonetic grounds for cross-linguistic phonological patterns of place and laryngeal contrasts for stop consonants (e.g., Steriade, 1997; Silverman, 1995; Wright, 2004; Flemming, 2005). Across languages, phonological place and laryngeal contrasts occur far more often in positions where the release bursts are available to cue the contrasts.

There is therefore considerable phonetic and phonological evidence for the importance of the release burst as a cue to stop identity. Thus, from the perspective of the relative richness of acoustic-phonetic cues, one might expect that released stops (with bursts) are processed more efficiently than

unreleased stops (with no bursts) irrespective of the phonological system of the listener's native language. This will be called the *phonetic-superiority hypothesis*.

In this study, we tested the *phonological-superiority* hypothesis against the *phonetic-superiority* hypothesis, by conducting two phoneme monitoring experiments (Connine and Titone, 1996; Otake *et al.*, 1996). Experiment 1 measured Korean listening performance; Experiment 2 examined Dutch listeners' performance on the same materials. Both groups of listeners heard four sets of speech materials: released and unreleased stops in Dutch, and released and unreleased stops in English. Both Dutch and English speech materials were included in order to examine whether the nature of the input language influences phoneme recognition. In Experiment 1, the language factor was whether the non-native speech materials were familiar or unfamiliar (to Korean listeners). In Experiment 2, it was whether the speech materials were native or non-native (to Dutch listeners).

II. EXPERIMENT 1

Korean listeners were presented with Dutch spoken stimuli in Experiment 1a and with English spoken stimuli in Experiment 1b. The listeners were asked to detect both nasal and oral consonants, in VC syllables. The oral consonants were either released or unreleased stops. We asked how speed and accuracy in phoneme monitoring varies as a function of the phonological viability of the stimuli in the listener's native language and as a function of the phonetic richness of those stimuli. On the one hand, the phonological-superiority hypothesis predicts that Korean listeners will detect unreleased stops more easily than released stops (because word-final stops in words spoken in isolation are always unreleased in Korean). On the other hand, the phonetic-superiority hypothesis predicts that Korean listeners will detect released final stops more easily than unreleased (because released stops are phonetically richer). The nasal targets, which were identical in the released and unreleased conditions, were included partly as distractors. Their inclusion also allowed us to examine whether phoneme monitoring performance on nasal targets varied depending on whether they co-occurred with released or unreleased stops. Finally, we tested whether phoneme monitoring performance is constrained by language experience (unfamiliar versus familiar speech stimuli, i.e., Dutch versus English).

A. Method

1. Participants

Fifty-two Korean student volunteers at Korea University in Seoul were paid to take part. They were divided into four groups of 13 according to the release condition (Released versus Unreleased) and the language condition (Dutch in Experiment 1a and English in Experiment 1b).

2. Materials

Two sets of materials were constructed in each language, one consisting of VC syllables with the release intact (the released condition), and one with the release spliced out (the unreleased condition). Within each set of VC's, there were 30 experimental items with 3 oral (/p, t, k/) and 3 nasal (/m, n, ŋ/) target consonants in their codas in 5 different vowel contexts (/a,e,i,o,u/; 6 target consonants × 5 vowel contexts =30 experimental items). Lists of four-syllable sequences were then constructed for each target-bearing experimental item with the target placed in either the last or the penultimate syllable, as in [V, V, V, V*C*] or [V, V, V*C*, V]. (Note that in the unreleased condition "VC" refers to a stimulus in which the released portion was spliced out, and therefore phonetic cues to "C" were present only in the preceding vowel.) Single V fillers (non-target-bearing syllables) were used instead of VC fillers in order to control for the frequency of occurrence of the target consonants. All V's in each list were different (e.g., [e, u, ak, i]). In addition to the 30 experimental lists, 15 nonexperimental lists were constructed in which the target consonants occurred in filler syllables with a schwa context. For these items, subjects had to detect the target consonants but the responses were not analyzed. These syllables were placed in the antepenultimate syllable ([V, V*C*, V, V]) in the lists, which made the position of target-bearing syllables in lists unpredictable. In addition to these 45 lists, 45 other lists were constructed as foils which had no instances of the specified target (thus no response was required). Among these 45 foil lists, 30 lists were the same as the 30 experimental lists except that the specified targets did not match the coda consonants in the lists. Each foil list was always presented after its occurrence as an experimental (target-bearing) list. The remaining 15 foil lists contained only vowels in the list ([V, V, V, V]). Finally, 15 practice lists were made.

3. Procedures

The speech materials (V's and VC's) were spoken individually by a male native speaker of Dutch (a speaker from Brabant), and by a male native speaker of American English (a speaker from the Midwest). Both speakers were not linguists, and both were naive as to the purpose of the experiment. The materials were recorded in a sound-treated studio directly onto computer and consequently down-sampled to 22.05 kHz (16 bit precision). All V's and VC's were repeated four times in blocks in a pseudorandomized order, and tokens with deviant prosody (e.g., extreme rising or falling intonation) were excluded. Both speakers released the stops naturally in all instances, though they were not instructed to do so. Randomly selected syllables were then combined to form four-syllable stimulus lists using the Praat speech editor (http://www.praat.org). In the stimulus lists, beginnings of syllables were separated by one second from each other. Two sets of stimuli were constructed which were identical except that, in one set, the release portions of the oral consonants in the VC stimuli were intact, while in the other set the release portions were spliced out (at a zero-crossing at the end of the vowel). The duration of the release portions were measured, as will be discussed in Sec. III B. The intensity of tokens across stimuli was then equalized, so that each stimulus had the same peak intensity but the relative amplitude within a stimulus remained intact.

The task of the subjects was to detect a pre-specified target phoneme irrespective of its position within a spoken

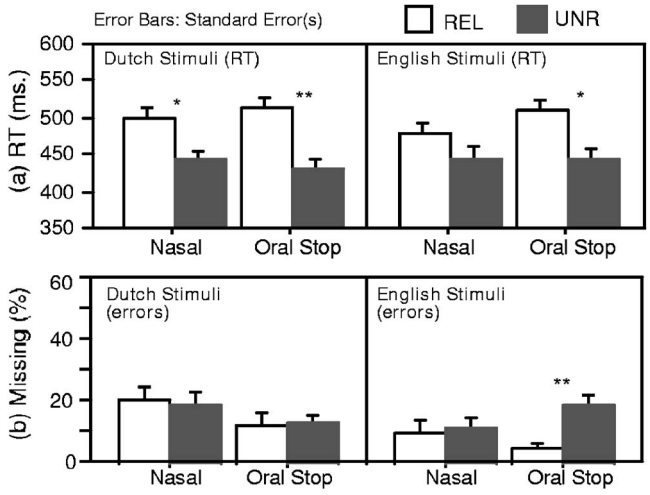


FIG. 1. Korean listeners' mean reaction Times [RTs, (a)] and errors [percent missing responses, (b)] with the spoken stimuli in Dutch (left) vs English (right). REL=released oral stops; UNR=unreleased oral stops. Note the nasal stimuli were identical in the two release conditions. *=$p<0.005$; **=$p<0.001$.

stimulus list. The targets were presented visually in lowercase letters in Roman alphabet for 1 s on a computer screen: **p**, **t**, **k**, **m**, **n**, **ng**, **l**. (/l/ was used in some filler lists.) Note that subjects were sufficiently familiar with the Roman alphabet for targets to be specified in this way. The first syllable of each stimulus list began 300 ms after the target disappeared. A new visual target was presented prior to each list. Subjects were told that they were going to hear some foreign speech (the identity of the language was not revealed) and were instructed to press a button on a response box (with their dominant hand) as fast and as accurately as possible when they detected the targets in the spoken lists.

Participants were tested individually in a quiet room. The computer clock was triggered in synchrony with the onset of the target presentation on the screen and stopped when the response button was pressed. Reaction times (RTs, relative to the offset of the vowel preceding the target phoneme) and errors were recorded. The session lasted approximately 15 min.

4. Analysis

Analyses of Variance (ANOVAs) were performed separately for each language (Dutch/English) with either subjects (F1) or items (F2) as the repeated measure. Both RT and error analyses were carried out (the errors were analyzed in terms of % missing responses). Stop release was a between-subject and within-item factor; manner (Nasal/Oral) was a within-subject and between-item factor. Planned pairwise comparisons between conditions were conducted with separate one-way ANOVAs. Reactions to both oral and nasal targets were analyzed in order to examine the effect of the presence/absence of releases on detection of nasal consonants. Remember that the nasal targets were identical in the released and unreleased conditions.

B. Results

The results are summarized in Fig. 1. Our analyses focus on the manner of articulation of the target consonants (nasals versus stops) and on whether or not the stops were released. Since our hypotheses did not concern either the place of articulation of the targets or the effects of vowel context, we collapsed over those factors. The patterns we report, however, were generally consistent over both place of articulation and vowel context, as shown in the Appendix.

1. Experiment 1a: Korean natives listening to Dutch

Korean listeners, when presented with Dutch spoken stimuli, showed a significant effect of Release ($F1[1,24]=6.969$, $p<0.025$; $F2[1,28]=28.29$, $p<0.001$), such that detection of the targets was faster by 69 ms in the unreleased than in the released condition [mean RTs, 437 vs 506 ms; s.e., 9.5 vs 10.7; see Fig. 1(a)]. Targets were therefore detected faster when they were phonologically viable, even though they were phonetically poorer. Interestingly, however, there was no significant interaction between Release and Manner ($F1[1,24]=2.04$, $p=0.166$; $F2[1,28]<1$), suggesting not only that unreleased stops were detected more rapidly than released stops, but also that nasals which were mixed with unreleased stops were processed more rapidly than those mixed with released stops (mean RTs for stops, 431 vs 513 ms; s.e., 14.4 vs 14.9; for nasals, 443 vs 499 ms; s.e., 11.2 vs 12.3). With respect to errors, neither a main effect of Release ($F1[1,24]<1$; $F2[1,28]<1$) nor an interaction between Release and Manner ($F1[1,24]<1$; $F2[1,28]<1$) was found.

2. Experiment 1b: Korean natives listening to English

In RTs with English stimuli, there was a significant main effect of Release ($F1[1,24]=9.71$, $F2[1,28]=38.72$, both $p<0.001$) and a significant Release × Manner interaction ($F1[1,24]=7.86$, $p<0.025$; $F2[1,28]=4.86$, $p<0.05$). As was the case with Dutch stimuli, Korean listeners were significantly faster by 68 ms in detecting (phonologically viable but phonetically poorer) English unreleased stops than released stops ($F1[1,24]=4.12$, $p<0.05$; $F2[1,14]=21.12$, $p<0.001$; mean RTs, 442 vs 510 ms; s.e., 16.4 vs 14.4), as shown in Fig. 1(a) (right). Detection of nasals was also faster in the unreleased than in the released condition by 36 ms, but this effect was significant only by item ($F1[1,24]=2.38$, $p>0.1$; $F2[1,14]=25.35$, $p<0.001$; mean RTs, 445 vs 480 ms; s.e., 11.2 vs 12.3).

In contrast to the RT data, however, detection of targets was more accurate for the (phonetically richer but phonologically nonviable) released stops than for the unreleased stops [see Fig. 1(b)]. There was a main effect of Release ($F1[1,24]=9.71$, $p=0.005$; $F2[1,28]=6.69$, $p<0.025$) and a significant Release × Manner interaction ($F1[1,24]=7.86$, $p=0.01$; $F2[1,28]=5.02$, $p<0.05$). Planned pairwise comparisons showed that the interaction was due to significantly higher accuracy for the released stops (5% errors; s.e., 1.6) than for the unreleased stops (19% errors; s.e., 2.7; $F1[1,24]=22.29$, $p<0.001$; $F2[1,14]=7.05$, $p<0.025$); there was no difference between the nasals in the released and unreleased conditions ($F1[1,24]<1$; $F2[1,14]<1$).

3. Combined analyses: Experiments 1a and 1b

We also asked whether Korean listeners' phoneme monitoring performance on the stops would be modulated by the language of the stimuli (Dutch versus English). This was a between-subject but within-item factor in two-way (Language × Release) ANOVAs. In RT, there was again a significant effect of Release ($F1[1,48]=11.17$, $p<0.005$; $F2[1,28]=50.53$, $p<0.001$), whereas neither a main effect of Language ($F2[1,48]<1$; $F2[1,28]<1$) nor an interaction between Language and Release ($F1[1,48]<1$; $F2[1,28]<1$) was found. The error analysis also showed a significant main effect of Release by subject ($F[1,48]=9.98$, $p<0.005$), but not by item ($F2[1,28]=2.73$, $p>0.1$) with no main effect of Language ($F1[1,48]<1$; $F2[1,28]<1$). There was also a significant interaction between the two factors by subject ($F1[1,48]=8.45$, $p<0.05$) but not by item ($F2[1,28]=2.73$, $p>0.1$). Planned pairwise comparisons suggested that this interaction was due to the fact that accuracy in detecting released stops was significantly higher with English (5% errors; s.e., 1.4) than with Dutch syllables (13% errors; s.e., 3.1) ($F1[1,24]=5.96$, $p<0.025$; $F2[1,28]=6.47$, $p<0.025$), while there was no language effect on unreleased stops ($F1[1,24]=2.81$, $p>0.1$; $F2[1,28]<1$, $p>0.5$; English versus Dutch mean errors: 19% vs 13%; s.e., 2.7 vs 2.0).

C. Discussion

The latency results showed that Korean listeners processed unreleased stops more rapidly than released stops, even though released stops carry additional phonetic cues to phoneme identity. This was true for both Dutch and English stimuli: As far as the RT data are concerned, the results thus appear to support the view that, in non-native listening, sounds which are phonologically viable in the native language are processed more efficiently than sounds which are phonologically illegitimate in the native language.

Interestingly, however, with respect to errors with English stimuli, accuracy in phoneme detection by Korean listeners improved when the targets were released. In other words, released stops (even though they are nonviable in Korean) were more accurately detected than unreleased stops (in spite of the fact that they are viable in Korean). This suggests that when the acoustic-phonetic cues in the release component of the stops were present, errors on the English spoken stimuli decreased, in spite of the illegality of released codas in Korean. This is at least in part in line with the view that the more cues to segment identity there are in the speech signal, the more efficiently the segment is processed (the phonetic-superiority hypothesis). There appears to be a language factor involved, however, since the phonetic-superiority effect in detection accuracy was observed with spoken stimuli in English—which Koreans are familiar with—but not with Dutch—a language unfamiliar to Koreans. This was further confirmed by the combined analysis which showed that the Koreans' detection accuracy was significantly higher for English than for Dutch consonants only in the released condition. We return to this issue in the General Discussion.

III. EXPERIMENT 2

Experiment 1 (the same materials and procedure) was re-run with Dutch listeners. This experiment had two goals. First, it provided a further test of the phonological-superiority hypothesis. If our interpretation of Experiment 1 is correct, such that Koreans detect unreleased stops more rapidly than released stops because unreleased coda stops are viable in Korean, then Dutch listeners should produce the opposite pattern of responses on the same stimuli. They should detect the released stops more easily than the unreleased stops because released coda stops are viable in Dutch. But they may also do so, of course, because the released stops are phonetically richer. Better performance by Dutch listeners on the released stops could thus be due to either phonological or phonetic influences. The critical analysis, therefore, is not that within the Dutch listeners' data itself, but that comparing the Dutch and Korean data (i.e., Experiment 1 versus Experiment 2). In this comparison, phonetic influences are controlled (the stimuli are the same for both groups of listeners), so if the Dutch and Korean listeners were to perform differently, this would support the phonological-superiority hypothesis. We therefore conducted an analysis in which the results of the two experiments were directly compared. Second, Experiment 2 again tested for effects of language familiarity, but now by contrasting performance on native language stimuli (Dutch) with non-native stimuli (English). This contrast provides a measure of whether non–native phoneme monitoring differs from the way the task is performed on native language materials.

A. Method

Seventy-two student volunteers from the Max Planck Institute subject pool were paid for their participation. They were all native speakers of Dutch, with no known hearing disorders. They were divided into four groups of 18, with two groups in Experiment 2a (Dutch stimuli) and two groups in Experiment 2b (English stimuli). In each sub-experiment, one group heard lists with released stops and the other heard lists with unreleased stops. The four sets of stimuli from Experiment 1 were re-used in Experiment 2. All experimental procedures were identical to those in the previous experiment.

B. Results

Ten subjects (five in each sub-experiment) were excluded from the analysis because they made errors so frequently that no scores were recorded for each of those subjects for at least one target consonant.

1. Experiment 2a: Dutch natives listening to Dutch

The RT analyses revealed that Dutch listeners showed the opposite pattern of performance on the Dutch stimuli to that of the Koreans. There was a main effect of Release on RT ($F1[1,29]=4.80$, $p<0.05$; $F2[1,28]=58.55$, $p<0.001$), which interacted with Manner ($F1[1,29]=19.32$, $p<0.001$; $F2[1,28]=23.02$, $p<0.001$). As shown in Fig. 2(a) (left), the interaction came from the fact that target stops were detected

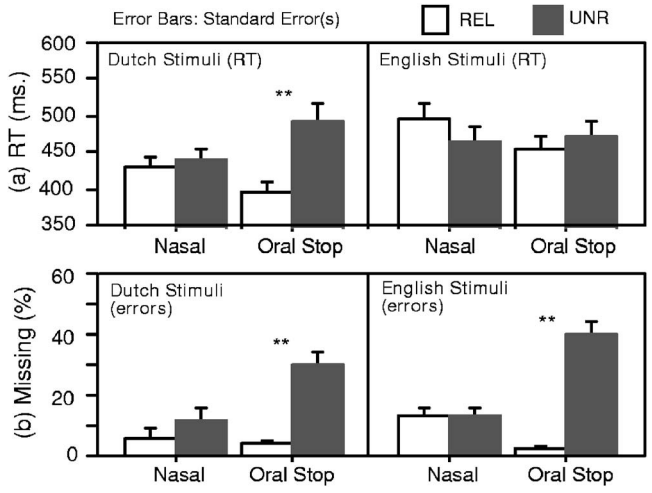


FIG. 2. Dutch listeners' mean Reaction Times [RTs, (a)] and errors [percent missing responses, (b)], with the spoken stimuli in Dutch (left) vs English (right). REL=released oral stops; UNR=unreleased oral stops. Note that the nasal stimuli were identical in the two release conditions. **=$p<0.001$.

more rapidly when released than unreleased by 97 ms ($F1[1,29]=13.09$, $p<0.001$; $F2[1,14]=87.69$, $p<0.001$; mean RTs: 400 vs 497 ms; s.e., 9.4 vs 17.9), while detection of the nasal targets was not influenced by the release condition—that is, irrespective of whether the nasals were presented in the same list as the released stops or mixed with unreleased stops ($F1[1,29]<1$; $F2[1,14]=4.65$, $p>0.07$; mean RTs for nasals: 432 vs 443 ms; s.e., 11.9 vs 11.6). Results with respect to the error rates showed a similar pattern (main effect of Release: $F1[1,29]=52.45$, $p<0.001$; $F2[1,28]=16.71$, $p<0.001$; Release × Manner interaction, $F1[1,29]=13.94$, $p<0.001$; $F2[1,28]=6.18$, $p<0.02$). Detection of stop targets was significantly more accurate when stops were released (4% errors; s.e., 1.3) than when they were unreleased (30% errors; s.e., 3.6; $F1[1,29]=54.89$, $p<0.001$; $F2[1,14]=13.60$, $p<0.005$), while there was only a nonsignificant trend for nasal targets (i.e., $F1[1,29]=3.49$, $p>0.07$; $F2[1,14]=3.11$, $p>0.09$; mean % errors for nasals: 6% vs 12%; s.e., 1.3 vs 2.9), as can be seen in Figure 2(b) (left).

2. Experiment 2b: Dutch natives listening to English

Unlike when they processed stimuli of their native language, Dutch listeners hearing English stimuli were not faster to respond to released than to unreleased stops [Fig. 2(a), right]. In the RT analysis there was no main effect of Release ($F1[1,29]<1$; $F2[1,28]<1$), but there was a significant Release × Manner interaction ($F1[1,29]=5.689$, $p<0.025$; $F2[1,28]=7.62$, $p<0.01$). This interaction was due to the fact that responses were faster for released than for unreleased stops (mean RTs: 459 vs 475 ms; s.e., 12.2 vs 17.8), while this effect reversed for the nasals (mean RTs: 499 vs 471 ms; s.e., 14.2 vs 14.4). But these simple effects were not significant by subjects ($F1[1,29]<1$, for both stops and nasals) or items ($F2[1,14]=3.38$, $p>0.08$ for stops; $F2[1,14]=4.28$, $p>0.05$ for nasals).

In errors, however, there was a significant effect of Release ($F1[1,29]=52.91$, $p<0.001$; $F2[1,28]=23.83$, $p<0.001$) and a significant Release × Manner interaction ($F[1,29]=69.51$, $p<0.001$; $F2[1,28]=20.59$, $p<0.001$). The interaction arose because Dutch listeners' detection of English oral stops was far more accurate for released than unreleased stops ($F1[1,29]=116.16$, $p<0.001$; $F2[1,14]=24.88$, $p<0.0001$; mean % errors, 2% vs 42%; s.e., 1 vs 4.4) whereas no such effect was found for nasal targets (both F's <1; mean % errors, 13% vs 14%; s.e., 2.4 vs 2.5).

3. Combined analyses: Experiments 2a and 2b

We tested if Dutch listeners' phoneme monitoring performance on the stops would be modulated by the language familiarity factor (Dutch versus English). Similar to Experiment 1, this was a between-subject factor but a within-item factor in two-way (Language × Release) ANOVAs. RT analyses showed no main effect of Language ($F1[1,58]<1$; $F2[1,28]=2.09$, $p>01$) and a trend toward a Language × Release interaction which was not significant by subject ($F1[1,58]=3.60$, $p=0.06$) but highly significant by item ($F2[1,28]=20.43$, $p<0.001$). The interaction came from the fact that in processing released (phonological viable) stops, Dutch listeners were faster by 59 ms with native stimuli than with non-native English stimuli ($F1[1,33]=5.75$, $p<0.025$; $F2[1,28]=10.87$, $p<0.003$; Dutch versus English mean RTs, 400 vs 459 ms; s.e., 14.6 vs 20.1), while a nonsignificant opposite pattern was found in processing unreleased (phonological unviable) stops (both F's <1; Dutch versus English mean RTs, 497 vs 476 ms; s.e., 24.3 vs 26.9).

Durational analysis of the released stimuli revealed, however, that the duration of the target consonant (the stop closure and the following release noise) was longer for the English than for the Dutch stimuli (closure, 112 vs 102 ms; release noise, 115 vs 90 ms; closure and release combined, 227 vs 192 ms). We therefore asked whether the fact that the Dutch listeners responded more slowly to the English released stops could be attributable to this durational difference. It is possible that listeners' responses could have been slower when the duration of the target consonant was longer, if their decisions tended to be delayed until they had heard all the available phonetic cues. We tested this possibility by examining, for each subject's data, the correlation between consonant duration and RTs. Variation in RTs was not significantly correlated with consonant duration in most cases. For Experiment 2a, only three out of 18 Dutch listeners (with Dutch stimuli) showed a significant positive correlation (the longer the consonant duration, the slower the RT) at $p<0.05$ (two-tailed), and five listeners showed negative correlation coefficients. For Experiment 2b, none of the Dutch listeners (with English stimuli) showed a significant correlation, and 13 out of 18 listeners showed negative correlation coefficients, which suggests, if anything, that responses were faster when the target consonants were longer. These results indicate that the Dutch listeners' slower responses to the English released stops were not due to the durational difference between the English and Dutch stop releases.

The combined error analysis in Experiment 2 showed a significant main effect of Language ($F1[1,58]=4.11$, $p<0.05$; $F2[1,28]=38.10$, $p<0.001$), such that overall accuracy was significantly higher for Dutch than for English

stimuli (Dutch versus English mean errors, 17% vs 22%; s.e., 1.8 vs 1.8). There was, however, a Language × Release interaction by subject ($F1[1,58]=7.29$, $p<0.01$) but not by item ($F2[1,28]=1.49$, $p>0.2$). Planned pairwise comparisons showed that this trend was due to the fact that when stops were released, there was no accuracy difference between Dutch and English stimuli (both F's <1; Dutch versus English mean errors, 4% vs 2%; s.e., 1.5 vs 1.0); when stops were unreleased, however, accuracy was significantly higher (by 12%) for Dutch stimuli than for English stimuli by subject ($F1[1,25]=5.14$, $p<0.04$; mean errors, 30% vs 42%; s.e., 3.6 vs 3.9), but again not by item ($F2[1,28]=1.09$, $p>0.3$).

4. Combined analyses: Korean and Dutch listeners

Finally, two sets of ANOVAs were run with listener group (Korean versus Dutch) and release condition as independent variables, one with Dutch stimuli (Experiments la and 2a) and one with English stimuli (Experiments 1b and 2b).

In the RT analysis with Dutch stimuli, there was a significant interaction between Listener Group and Release ($F1[1,53]=20.59$, $p<0.001$; $F2[1,28]=98.98$, $p<0.001$). The interaction was due to reverse patterns across the two listener groups. In processing Dutch released stops, Dutch listeners detected the released stops more rapidly than Korean listeners did ($F1[1,29]=19.15$, $p<0.001$; $F2[1,28]=43.92$, $p<0.001$). The reverse was true with unreleased stops: Korean listeners detected the unreleased stops more rapidly than Dutch listeners did ($F1[1,24]=4.78$, $p<0.04$; $F2[1,28]=6.55$, $p<0.02$). In the corresponding error analysis, there was also a significant Listener Group × Release interaction ($F1[1,53]=25.31$, $p<0.001$; $F2[1,28]=7.10$, $p<0.02$). This interaction was again due to reverse patterns of performance: the Dutch were more accurate in detecting released stops ($F1[1,29]=7.66$, $p=0.01$; $F2[1,28]=7.28$, $p<0.02$); the Koreans were more accurate in detecting unreleased stops, significantly so by subject ($F1[1,24]=16.97$, $p<0.001$) and marginally so by item ($F2[1,28]=3.40$, $p<0.08$).

In the analyses with English stimuli, there was a trend towards a Listener Group × Release interaction in the RT analysis by subject ($F1[1,53]=3.10$, $p<0.08$), but the interaction was highly significant by item ($F2[1,28]=22.37$, $p<0.001$). This interaction was again due to opposite patterns in the two listener groups. For the released stops, the difference between listener groups was significant only by item ($F1[1,28]=2.75$, $p>0.1$, $F2[1,28]=5.83$, $p<0.025$); for the unreleased stops this difference was not significant ($F1[1,25]<1$; $F2[1,28]=3.39$, $p<0.08$).

In the corresponding error analysis, however, there was a significant Listener Group × Release interaction ($F1[1,53]=27.23$, $p<0.001$; $F2[1,28]=6.74$, $p<0.02$). This interaction did not reflect opposite patterns in the two listener groups, however: the Koreans were more accurate on the unreleased stops than the Dutch ($F1[1,25]=23.36$, $p<0.001$; $F2[1,28]=4.73$, $p<0.04$), but there was no accuracy difference between the two groups on the released stops ($F1[1,28]=1.88$, $p>0.1$; $F2[1,28]=1.33$, $p>0.2$). This is probably attributable to a floor effect, given that both listener groups were very accurate.

C. Discussion

Perhaps unsurprisingly, Dutch listeners detected released stops in Dutch syllables faster than in English syllables. Similarly, they detected unreleased stops in the Dutch stimuli more accurately than in the English stimuli. Both of these effects reflect global benefits in processing native over non-native speech. The more specific comparison of listening performance on released and unreleased stops, however, provides evidence on the role of phonological knowledge in phoneme monitoring. When presented with native speech materials, Dutch listeners detected released stop targets more rapidly and more accurately than unreleased stop targets. But when they were presented with English spoken stimuli this effect was attenuated: although it was significant in the error analysis, there was no effect in the RT analysis. A violation of the phonological constraints of the listener's native language thus appears to interfere more severely with processing native than non-native speech.

The overall better performance on released stops can be accounted for in terms of the phonological-superiority hypothesis: Speech processing may be better when the sounds are in harmony with the phonological constraints of the listener's native language than when the sounds are phonologically nonviable. It is also possible, however, that the improved performance in detecting released stops is simply due to the richness of cues present in the signal: The more cues to segment identity in the speech signal, the more efficiently the segment is processed (the phonetic-superiority hypothesis). When a stop is released, its identity is cued not only by the formant transition in the vowel but also by the spectral characteristics during the burst, whereas an unreleased stop is perceptually weaker because it is cued only by the formant transition information.

One might argue that the attenuation of the released/unreleased difference in non-native listening (i.e., the Dutch listening to English) offers support for the phonological-superiority hypothesis. If the difference is due to phonological knowledge about the native language, one might expect a weaker effect when the input consists of non-native syllables. But it remains possible that the phonetic cues were stronger in the Dutch than in the American English stimuli. The critical comparisons are therefore the cross-experiment contrasts (i.e., on the Korean versus Dutch listener groups), where phonetic differences, should they exist, are held constant. These comparisons indeed showed that there were significantly different patterns in the two listener groups on both sets of stimuli. This suggests that phonological viability in the native language plays a stronger role in non-native listening than the availability and/or relative richness of acoustic-phonetic cues.

IV. GENERAL DISCUSSION

On the basis of the results of two experiments, we draw three conclusions. First, in line with previous research, per-

ception of non-native and native speech is influenced by the phonological system of the listener's native language. Second, although the richness of acoustic phonetic information also plays a role in consonant identification, this influence in non-native perception can be weaker than that due to native language phonological knowledge. Third, perception of non-native speech is also influenced by the listener's familiarity with the phonetic detail in the non-native language. We discuss each of these points in turn.

In line with the phonological-superiority hypothesis, phoneme recognition is improved in a speech environment which is phonologically viable in the listener's native language: Korean listeners detected unreleased coda stops faster than released coda stops, but Dutch listeners, responding to the same stimuli, detected the released stops more rapidly and/or more accurately. This pattern was confirmed by our cross-experiment comparisons, which showed that, for the unreleased Dutch stops, the Koreans were faster and more accurate than the Dutch, whereas for the released Dutch stops, the Dutch were faster and more accurate than the Koreans. The same pattern, though statistically weaker, was also observed for the English stimuli; we discuss the reasons for the weakening of the effect in the following. Nevertheless, there was still an influence of native phonology in the error rates on the English unreleased stops: Korean listeners missed significantly fewer target phonemes in this condition than the Dutch listeners. Finally, the results indicate that this phonological effect was stronger in native listening (Dutch listeners, Dutch VC's; significant effects in both speed and accuracy) than in non-native listening (Dutch listeners, English VC's; significant effects only in accuracy).

We suggest that listeners' expectations about the acoustic-phonetic form of a stop consonant in coda position, as determined by the phonological sound patterns of the listeners' native language and therefore a lifetime's exposure to those sound patterns, influence ease of phoneme recognition in native and, more strikingly, in non-native listening. Previous research has shown that various aspects of the phonological structure of one's native language influence non-native listening [e.g., phoneme inventory, Costa *et al.* (1998), Flege and Hillenbrand (1986); place assimilation rules, Weber (2001); phonotactic restrictions, Broersma (2005), Weber and Cutler (2006); and rhythmic structure, Cutler *et al.* (1986), Cutler and Otake (1994), Otake *et al.* (1993)]. The present findings add to this body of evidence by showing that native-language experience can determine ease of phoneme identification even in the extreme case where non-native target consonants which are phonologically expected in the listener's native language are in fact phonetically poorer (as when the Korean listeners detected unreleased stops more readily than released stops).

This, then, is our second conclusion: phonological expectations can outweigh phonetic information. It is undeniably true that native and non-native listening depends on the uptake of acoustic-phonetic cues present in the signal. In general, one would therefore assume that speech sounds which are more richly specified (e.g., released coda stops cued by VC formant transition information and by the information in the release burst) would be easier to recognize than sounds which are less well specified (e.g., unreleased coda stops cued only by the VC formant transitions). The present results suggest, however, that at least under some circumstances, poorer acoustic-phonetic information can lead to better phoneme recognition in non-native listening.

The relevant phonological fact that has been the focus of the present study is that Korean stops are obligatorily unreleased in coda position in words spoken in isolation, as compared to Dutch stops, which are generally released in that environment. There is, however, an additional phonological difference concerning coda position in Korean and Dutch which might be relevant. Korean stops are phonologically divided into the categories lenis, fortis and aspirated, and these manners of articulation are cued primarily in the post-release part of the acoustic signal (Cho *et al.*, 2002; Kim *et al.*, 2002). This three-way manner contrast is neutralized completely in coda position (Kim and Jongman, 1996). Furthermore, the three fricatives in Korean (/s,s*,h/) become unreleased alveolar stops in coda position (Kim-Renaud, 1976; Martin, 1992). The situation is very different in Dutch codas. Although there is word-final devoicing in Dutch, this process is incomplete: There are fine-grained phonetic differences between underlyingly voiced and voiceless coda stops (Warner *et al.*, 2004; Ernestus and Baayen, in press). Because of this incomplete neutralization process, five stops can appear in Dutch singeleton codas. Fricatives can also occur in the coda in Dutch. The possible phonetic forms that can occur in coda position are thus more limited in Korean than in Dutch. The functional load in processing consonants in the coda may therefore be lower for Korean than for Dutch listeners. It is possible that this difference may influence perception of coda consonants in the two languages. Our findings do not bear directly on this issue, however, because phonologically viable phonetic forms in the coda position are unreleased in Korean but released in Dutch, and it is this which we manipulated, not the size of the set of coda consonants. A pure test of this set-size hypothesis would require a comparison of two languages with the same type of stops in the coda (either released or unreleased), so that the only relevant factor to consider would be the different number of allowable phonetic forms in the coda stemming from language-specific neutralization processes.

Although poorer acoustic-phonetic information led to faster phoneme recognition in non-native listening by the Koreans, there was some evidence that acoustic-phonetic cues can also be exploited by listeners in non-native speech perception, even if those cues are not the ones used in processing the listener's native language. Koreans were faster on unreleased than on released stops in both Dutch and English. But, for the English stimuli alone, the Koreans were more accurate on the released than on unreleased stops. We suggest that this interaction between the application of phonological constraints and the use of acoustic-phonetic cues in non-native perception depends on language familiarity. Dutch is a truly exotic language to Korean university students, whereas English is, for them, the most familiar non-native language. They are likely to have learned English since they were 12 years old but never to have heard Dutch before. In the course of second-language acquisition, Korean

students might have built up some familiarity with the acoustic-phonetic characteristics of the sounds of English. If so, then even though the violation of the phonological constraints of Korean in English released stops led to slow detection times, familiarity with the form of such stops may have allowed Korean learners of English to exploit the richer set of acoustic-phonetic cues available in the released stops to improve detection accuracy.[1]

On this view, Korean listeners are able to extract efficiently information about stop identity in the formant transitions of the preceding vowel. When there is no release (in both the Dutch and English stimuli), they are therefore able to recognize target phonemes quickly, and quite accurately. When a release follows, however, their phoneme monitoring performance is disrupted, resulting in slower responses to both the Dutch and English released stimuli. Familiarity with English can nevertheless allow Koreans to be very accurate on the English released stops. They have no experience with Dutch and thus the impact of the release-induced disruption is greater, that is, the additional cues in the Dutch released stops are not fully exploited. But the additional information in the English releases, though unable to undo the disruption in processing speed caused by the releases' phonological nonviability, leads to very accurate performance. Thus, if non-native speech contains cues that are not present in the equivalent position in native speech, and if non-native listeners are sufficiently familiar with that language, then those listeners can use those richer cues in phoneme identification.

This use of non-native cues contrasts with the findings of Broersma (2005). She showed that Dutch listeners who were very familiar with English (at least as familiar as the present Koreans were with English) did not process vowel-duration cues to the English coda voicing distinction, unlike a group of native English listeners. The Dutch listeners nevertheless categorized non-native English coda fricatives as voiced or voiceless at least as well as the native English listeners. Broersma suggests that this may have been because the Dutch listeners have learned that vowel duration is not a very informative cue to the voicing distinction in Dutch, and thus do not rely on this cue in non-native listening either. The present results suggest however that it is not necessarily the case that phonetic information that is not valuable in native language perception (or that is even not present in the native language) is ignored when the listener hears a foreign language. One reason for the difference between studies may be due to a difference in cue exposure. The Dutch listeners in the Broersma study have been exposed to differences in vowel duration in their native language, as a cue to vowel quantity (i.e., the phonemic contrast between long and short vowels; cf. Nooteboom, 1972; Nooteboom and Doodeman, 1980) and to other cues to the obstruent voicing contrast. Dutch listeners may thus have learned that vowel duration is a weak cue to obstruent voicing. But the Korean listeners in the present study are never exposed to release bursts in coda position in their native language and therefore cannot learn, during native language exposure, to ignore this information in stop identification or to use it for the recognition of some other sound(s). When Koreans therefore have sufficient exposure to release bursts in non-native listening, they can use the additional information contained in those bursts to improve target recognition accuracy.

The present study has pitted native-language phonological viability against the relative richness of acoustic-phonetic cues in non–native listening. Phonological viability seems to have the upper hand: Phoneme recognition is in general better when the phonetic specification of those phonemes is in keeping with the listener's native phonology, even when that phonetic specification is, in physical terms, poorer. This dominance of phonological knowledge is consistent with theories of non-native speech perception, such as Flege's Speech Learning Model (SLM, e.g., Flege, 1991, 1995) and Best's Perceptual Assimilation Model (PAM, e.g., Best, 1994, 1995, Best *et al.*, 2001). Both models predict that non-native sounds are perceived through some form of phonological filter of the listeners' native language that takes into account the nonnative sounds' phonetic (dis)similarities to native phonetic categories. Previous studies supporting these models, however, have generally been based on cross-linguistic phonological differences concerning the phoneme inventories of the languages in question. The findings of the present study therefore show that the phonological filter created through exposure to a native language is determined not only by the phoneme inventory of that language but also by sound patterns that are driven by phonological (allophonic) rules. As we have just argued, however, the relative richness of phonetic cues can also play a role in non-native listening, at least in a non-native language with which the listeners are familiar. Models of non-native listening must therefore not only take native-language phonological processes into account (both phonemic and rule-driven processes), but also the absolute amount of phonetic information in the nonnative speech stream.

ACKNOWLEDGMENTS

This study was carried out in large part when T. C. was affiliated with the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. We thank Kee-Ho Kim at Korea University for helping us with the acquisition of the Korean data. We also thank Mirjam Broersma, Anne Cutler, Elizabeth Johnson, and Roel Smits for their comments at various stages of this project; and Ann Bradlow and three anonymous reviewers for their comments and suggestions. A preliminary report of this research appears as Cho and McQueen (2004).

APPENDIX

C

TABLE I. Mean reaction times (ms) by place of articulation (/p,t,k/) and vowel context (/a,e,i,o,u/).

Listeners	Stimuli	Vowel context	/p/		/t/		/k/	
			REL	UNR	REL	UNR	REL	UNR
Korean listeners (Experiment 1)	Dutch	/a/	462	394	556	516	480	473
		/e/	479	390	574	498	557	392
		/i/	451	343	546	453	419	347
		/o/	550	395	524	568	478	373
		/u/	472	443	508	432	629	537
	English	/a/	423	380	444	426	478	370
		/e/	501	436	513	407	499	492
		/i/	557	446	533	601	444	430
		/o/	473	391	608	490	544	383
		/u/	528	423	485	389	616	551
Dutch listeners (Experiment 2)	Dutch	/a/	431	490	419	568	363	518
		/e/	371	427	433	495	398	524
		/i/	386	475	448	476	407	509
		/o/	360	421	394	532	368	428
		/u/	367	457	379	489	473	580
	English	/a/	420	411	408	442	358	468
		/e/	417	445	478	486	433	412
		/i/	477	479	510	560	469	501
		/o/	402	409	523	475	473	501
		/u/	480	520	426	541	602	560

NOTE: REL=released stops; UNR=unreleased stops.

TABLE II. Percent errors (missing responses) by place of articulation (/p,t,k/) and vowel context (/a,e,i,o,u/).

Listeners	Stimuli	Vowel context	/p/		/t/		/k/	
			REL	UNR	REL	UNR	REL	UNR
Korean listeners (Experiment 1)	Dutch	/a/	8	15	8	38	31	0
		/e/	8	0	23	0	8	0
		/i/	8	0	8	8	8	8
		/o/	0	8	38	8	8	8
		/u/	8	0	15	8	15	92
	English	/a/	0	8	0	8	0	0
		/e/	15	15	0	15	8	8
		/i/	0	0	8	38	0	0
		/o/	0	8	8	8	8	69
		/u/	0	15	0	8	23	85
Dutch listeners (Experiment 2)	Dutch	/a/	17	15	0	38	0	15
		/e/	0	15	22	69	0	8
		/i/	0	0	6	46	0	46
		/o/	0	0	0	62	11	8
		/u/	6	23	0	15	0	92
	English	/a/	0	7	6	29	6	21
		/e/	0	7	0	43	6	14
		/i/	0	21	0	93	6	71
		/o/	0	14	0	79	6	64
		/u/	0	21	0	43	6	93

NOTE: REL=released stops; UNR=unreleased stops.

[1]This language familiarity effect is further supported by an analysis of detection accuracy on the nasals. (We thank an anonymous reviewer for bringing the detection accuracy data on nasals to our attention.) As can be seen in Fig. 1, Korean listeners were more accurate in detecting (for them more familiar) English nasals than (for them unfamiliar) Dutch nasals, especially in the released condition (10% vs 21% errors; $F1(1,24)=7.08$, $p=0.014$; $F2(1,14)=1.93$, $p>0.1$) but also in the unreleased condition (11% vs 20% errors; $F1(1,24)=3.49$, $p=0.174$; $F2(1,14)=1.31$, $p>0.1$). A similar observation can be made from the Dutch listeners' responses to the nasals (Fig. 2). Dutch listeners were more accurate in detecting Dutch nasals than less familiar (non-native) English nasals, but only in the released condition (6% vs 13% errors; $F1(1,29)=6.38$, $p=0.016$; $F2(1,14)=4.45$, $p=0.053$).

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," J. Acoust. Soc. Am. **109**, 775–794.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, ND), pp. 171–204.

Best, C. T. (**1994**). "The emergence of language-specific phonemic influences in infant speech perception," in *Development of speech perception: The transition from speech sounds to spoken words*, edited by J. Goodman and J. C. Nusbaum (MIT Press, Cambridge, MA), pp. 167–224.

Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phonetics **20**, 305–330.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization of non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Bradlow, A. R., Pisoni, D. B., Arkahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

Cho, T., Jun, S.-A., and Ladefoged, P. (**2002**). "Acoustic and aerodynamic correlates of Korean stops and fricatives," J. Phonetics **30**, 193–228.

Cho, T., and McQueen, J. M. (**2004**). "Phonotactics vs. phonetic cues in native and non-native listening: Dutch and Korean listeners' perception of Dutch and English," in *Proceedings of ICSLP2004 (International Conference on Spoken Language Processing), Jeju Island, Korea* (Sunjin, Seoul), pp. 1301–1304.

Connine, C. M., and Titone, D. (**1996**). "Phoneme monitoring," Lang. Cognit. Processes **11**, 635–645.

Costa, A., Cutler, A., and Sebastián-Gallés, N. (**1998**). "Effects of phoneme repertoire on phoneme decision," Percept. Psychophys. **60**, 1022–1021.

Cutler, A., and Broersma, M. (**2005**). "Phonetic precision in listening," in *A Figure of Speech*, edited by W. Hardcastle and J. Beck (Erlbaum, Mahwah, NJ), pp. 63–91.

Cutler, A., Mehler, J., Norris, D., and Seguí, J. (**1986**). "The syllable's differing role in the segmentation of French and English," J. Mem. Lang. **25**, 385–400.

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Mem. Lang. **33**, 824–844.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Ernestus, M., and H. Baayen. (**in press**). "The functionality of incomplete neutralization in Dutch. The case of past-tense formation," in *Laboratory phonology 8: Varieties of Phonological Competence*, edited by L. M. Goldstein, D. H. Whalen, and C. T. Best (Mouton de Gruyter, Berlin).

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, ND), pp. 133–172.

Flege, J. E. (**1991**). "Orthographic evidence for the perceptual identification of vowels in Spanish and English," Q. J. Exp. Psychol. **43**, 701–731.

Flege, J. E., and Hillenbrand, J. (**1986**). "Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English," J. Acoust. Soc. Am. **79**, 508–517.

Flemming, E. (**2005**). "Speech perception and phonological contrast," in *The Handbook of Speech Perception*, edited by D. Pisoni and R. Remez (Blackwell), pp. 156–181.

Hallé, P. A., Best, C. T., and Levitt, A. (**1999**). "Phonetic vs. phonological influences on French listeners' perception of American English approximants," J. Phonetics **27**, 281–306.

Henderson, J. B., and Repp, B. H. (**1982**). "Is a stop consonant released when followed by another stop consonant?," Phonetica **39**, 71–82.

Jusczyk, P. W. (**1997**). *The Discovery of Spoken Language* (MIT, Cambridge, MA).

Kim, H., and Jongman, A. (**1996**). "Acoustic and perceptual evidence for complete neutralization of manner of articulation in Korean," J. Phonetics **24**, 295–312.

Kim, M.-Y., Beddor, P. S., and Horroks, J. (**2002**). "The contribution of consonantal and vocalic information to the perception of Korean initial stops," J. Phonetics **30**, 77–100.

Kim-Renaud, Y. K. (**1976**). "Korean consonantal phonology," Ph.D. dissertation, University of Hawaii.

Kochetov, A. (**2001**). "Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization," Ph.D. dissertation, University of Toronto.

Kuhl, P. K., and Iverson, P. (**1995**). "Linguistic experience and the perceptual magnet effect," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium, ND), pp. 12l–154.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Lively, S. E., Logan, J. S., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). "Training Japanese listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories," J. Acoust. Soc. Am. **96**, 2076–2087.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–866.

Martin, S. E. (**1992**). *A Reference Grammar of Korean* (Tuttle, Rutland, VT).

Mattingly, I. G. (**1981**). "Phonetic representation and speech synthesis by rule," in *Proceedings of the 13th International Congress of Phonetic Sciences*, edited by T. Myers, J. Laver, and J. Anderson (Royal Institute of Technology and Stockholm University, Stockholm), pp. 574–577.

Nooteboom, S. G. (**1972**). "Production and perception of vowel duration," Vol. 4, Ph.D. dissertation. Rijksuniversiteit Utrecht.

Nooteboom, S. G., and Doodeman, G. J. N. (**1980**). "Production and perception of vowel length in spoken sentences," J. Acoust. Soc. Am. **67**, 267–287.

Otake, T., Hatano, G., Cutler, A., and Mehler, J. (**1993**). "Mora or syllable? Speech segmentation in Japanese," J. Mem. Lang. **32**, 358–378.

Otake, T., Yoneyama, K., Cutler, A., and van der Lugt, A. (**1996**). "The representation of Japanese moraic nasals," J. Acoust. Soc. Am. **100**, 3831–3842.

Polka, L., Colantonio, C., and Sundara, M. (**2001**). "A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern," J. Acoust. Soc. Am. **109**, 2190–2220.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Silverman, D. (1995). "Phasing and Recoverability", Ph.D. dissertation, University of California, Los Angeles.

Smits, R., Ten Bosch, L., and Collier, R. (**1996**). "Evaluation of various sets of acoustic cues for the perception of prevocalic stop consonants. I. Perception experiment," J. Acoust. Soc. Am. **100**, 3852–3864.

Steriade, D. (1997). "Phonetics in phonology. The case of laryngeal neutralization," University of California, Los Angeles (Unpublished) (http://web.mit.edu/linguistics/www/bibliography/steriade.html).

Stevens, K. N., and Blumstein, S. E. (**1978**). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. **64**, 1358–1368.

Warner, N., Jongman, A., Sereno, J., and Kemps, R. (**2004**). "Incomplete neutralization and other sub-phonetic durational differences in production and perception: Evidence from Dutch," J. Phonetics **32**, 251–276.

Watson, I. (1983). "Cues to the voicing contrast: A survey," Cambridge Papers in Phonetics and Experimental Linguistics **2**, 1–34.

Weber, A. (**2001**). "Help or hindrance: How violation of different assimilation rules affects spoken-language processing," Lang Speech **44**, 95–118.

Weber, A., and Cutler, A. (**2006**). "First-language phonotactics in second-language listening," J. Acoust. Soc. Am. **119**, 597–607.

Werker, J. F., and Lalonde, C. E. (**1988**). "Cross-language speech percep-

tion: Initial capabilities and developmental change," Dev. Psychol. **24**, 672–683.
Werker, J. F., Gilbert, J. H. V., Humphrey, K., and Tees, R. C. (**1981**). "Developmental aspects of cross-language speech perception," Child Dev. **52**, 349–355.
Wright, R. (**2004**). "A review of perceptual cues and cue robustness," in *Phonetically Based Phonology*, edited by B. Hayes, R. Kirchner, and D. Steriade (Cambridge University Press, Cambridge), pp. 34–57.
Zsiga, E. (**2003**). "Articulatory timing in a second language. Evidence from Russian and English," Stud. Second Lang. Acquis. **25**, 399–432.

Speech perception in gated noise: The effects of temporal resolution

Su-Hyun Jin[a)]
Division of Communication Disorders, University of Wyoming, Laramie, Wyoming 82070

Peggy B. Nelson
Department of Communication Disorders, University of Minnesota, Minneapolis, Minnesota 55455

(Received 14 June 2005; revised 27 February 2006; accepted 28 February 2006)

Previous investigations have suggested that hearing-impaired (HI) listeners have reduced masking release (MR) compared to normal hearing listeners (NH) when they listen in modulated noise. The current study examined the following questions that have not been clearly answered: First, when HI listeners are amplified so that their performance is equal to that of NH listeners in quiet and in steady noise, do HI listeners still show reduced MR with modulated noise when compared to NH listeners? Second, is the masking release the same for sentences and CV syllables? Third, does forward masking significantly contribute to the variability in performance among HI listeners? To compensate for reduced hearing sensitivity for HI listeners, the spectrum levels of both speech and noise were adjusted based on the individual hearing loss. There was no significant difference between the performance of NH listeners and that of HI listeners in steady noise and in quiet. However, the amount of MR for sentences and for CV syllables was significantly reduced for HI listeners. For sentence recognition, the amount of MR seemed to be more related to hearing sensitivity for low-to-mid frequencies than to forward masking. In contrast, forward masking thresholds appear to be a major contributor to the amount of MR for syllable recognition. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188688]

PACS number(s): 43.71.Ky, 43.71.Es [KWG] Pages: 3097–3108

I. INTRODUCTION

Hearing-impaired (HI) listeners often report difficulty understanding speech in the presence of background noise. Several investigators have suggested that the Articulation Index (AI, French and Steinberg, 1947) can account for the speech recognition performance HI listeners with mild-to-moderate hearing loss (e.g., Dubno and Dirks, 1989; Pavlovic *et al.*, 1986; Zurek and Delhorne, 1987). According to this hypothesis, HI listeners with mild-to-moderate hearing loss should understand speech in quiet as well as normal hearing (NH) listeners if the whole dynamic range of speech were made audible. The effect of steady noise on HI listeners becomes more detrimental than it is to NH listeners because HI listeners already have reduced redundancy of the speech signal in quiet because portions of speech information are below the thresholds (Van Tasell, 1993). Nevertheless, if the AI function for a certain speech material and signal-to-noise ratio (SNR) is known, the speech recognition ability of HI listeners in steady noise can be predicted based on their thresholds.

However, differences in speech recognition performance between NH and HI listeners become more apparent if the noise is fluctuating in level (Bacon *et al.*, 1998). In amplitude modulated noise, NH listeners can benefit from brief moments in which they experience a favorable SNR. This results in a better ability to understand speech in modulated than in steady noise. Depending on the depth and rate of fluctuation the degree of masking release in NH listeners can range from 2 to 30 dB (Gustafson and Arlinger, 1994; Howard-Jones and Rosen, 1993; Miller and Licklider, 1950). Compared to NH, HI listeners show significantly less masking release (Bacon *et al.*, 1998; Eisenberg *et al.*, 1995; Festen and Plomp, 1990; Gustafsson and Arlinger, 1994; Nelson *et al.*, 2003). The amount of masking release remains reduced even when thresholds of HI listeners are close to those of NH listeners except at higher frequencies (Dubno *et al.*, 2002) or when speech signals are presented at above threshold levels (Eisenberg *et al.*, 1995).

In an effort to explain these findings in terms of decreased audibility in HI listeners, Trine (1995) examined the relationship between the amount of masking release and the average pure-tone threshold. He asked listeners to adjust the level of either the noise masker or speech signal to the point where they could understand speech 50% of time. The noise was steady or modulated by an 8 Hz square wave at different modulation depths. Trine found a high negative correlation (correlation coefficient of either −0.81 or −0.88 depending on the overall level of speech and noise) between the degree of hearing loss and the amount of masking release. Therefore, he inferred that if the dips in the modulated noise were above threshold (so that modulation depth was not affected by hearing loss), both NH and HI listeners would show similar amounts of masking release.

However, it has been argued that reduction in masking release in HI listeners could not be accounted for by reduced audibility alone. For example, Eisenberg *et al.*(1995) compared consonant recognition of HI listeners in steady and

[a)]Electronic mail: shjin@uwyo.edu

modulated noise with performance of NH listeners who were tested with and without spectrally shaped background noise that simulated hearing loss. NH listeners performed significantly better in modulated noise than in steady noise, both with and without spectrally shaped noise. In comparison, the investigators examined masking release of HI listeners when the speech and noise were spectrally shaped and amplified to compensate for the loss of hearing sensitivity. Even with speech and modulated noise that were audible, HI listeners still showed less release from masking, relative to NH listeners.

Suprathreshold psychoacoustic abilities other than audibility, such as spectral resolution, temporal resolution, or auditory integration, have been linked to understanding speech in fluctuating noise. For example, Peters *et al.* (1998) suggested that HI listeners have reduced ability to use both spectral and temporal gaps in the background noise because of deficits in frequency selectivity as well as reduced audibility associated with hearing loss. Peters *et al.* measured speech reception thresholds (SRTs) of four different groups of listeners in various noise conditions: young and elderly listeners with normal hearing and young/elderly listeners with moderate to severe hearing loss. The hearing in noise test (HINT) sentences were used to measure SRT of each participant in the presence of six different background noises: a steady noise shaped by the long-term average spectrum of HINT sentences (called HINT noise), a single competing female talker, a HINT noise with temporal fluctuations, a HINT noise with spectral gaps in several frequency regions and a noise with both temporal and spectral gaps. To compensate for reduced audibility for HI listeners, HINT sentences and the background noises were amplified using linear amplification combined with frequency-response shaping (NAL amplification). The results showed that both young and elderly HI listeners had significantly higher SRTs than the normal-hearing groups even when amplified speech and noise were presented to HI listeners. When the background noise was temporally and spectrally modulated, the differences between NH and HI listeners were much greater (7–15 dB) than when the noise was steady (2–3 dB). Reduced audibility as well as suprathreshold deficits were considered as a possible contributor to the poorer SRTs of HI listeners.

It has also been speculated that reduced temporal resolution associated with cochlear hearing loss (Glasberg *et al.*, 1987; Kidd *et al.*, 1983; Nelson and Freyman, 1987) might play a role in speech perception with fluctuating noise (Festen and Plomp, 1990; Dubno *et al.*, 2002). Temporal resolution has been considered as an important characteristic for listeners to understand fast time-varying stimuli like speech. Listeners need to be able to follow acoustic signals in sequence and to put them together into whole meaningful speech. One of the common methods to examine temporal resolution is forward masking which would be a relevant measure in the context of speech recognition in modulated noise (Festen, 1993; Dubno *et al.*, 2002). That is, brief release in dips of fluctuating noise may be affected by recovery from masking from the preceding peaks of noise, which is similar to the forward masking paradigm. A long recovery time may make speech information in dips less audible.

Plack *et al.* (2002) hypothesized that all the major nonlinear processes of forward masking in NH listeners can be accounted for by the nonlinear response characteristics of basilar membrane (BM). One of the nonlinear properties of temporal masking is that increments in masker level do not produce equal increments in masked threshold (Moore and Glasberg, 1983). For example, if the masker level is increased by 10 dB, the amount of forward masking is only increased by 3 dB. This results in nonlinear growth of the forward masking function. Sensorineural hearing loss is presumably related to a loss of nonlinear BM response (Ruggero *et al.*, 1997). If the hypothesis of forward masking is true, HI listeners should show a loss of nonlinear function in forward masking (Plack *et al.*, 2002). In fact, findings from HI listeners in forward masking experiments have supported this assumption. That is, the growth of forward masking increases linearly as masker level increases, which results in linear growth of forward masking in HI listeners (Kidd *et al.*, 1983; Oxenham and Plack, 1997). This means that HI listeners might be more affected by forward masking at a given masker level than would NH listeners.

Bacon *et al.* (1998) speculated that reduced masking release in HI listeners could be partially accounted for by reduced temporal resolution. They measured SRTs in the presence of temporally modulated speech-shaped noises from three groups of listeners: HI listeners, NH listeners, and those with NH who listened to speech in the spectrally shaped noise to simulate hearing loss of HI listeners (NM). The results showed that NM and some HI listeners (5 out of 11 NH listeners) had similar amount of masking release that was less than NH listeners. On the other hand, the rest of the HI group showed significantly less release of masking than the NM group. This suggests that the extra amount of reduced masking release in some HI listeners is not the result of reduced audibility only. Their auditory systems might be more susceptible to forward masking due to cochlear hearing loss so that these listeners may not have as much masking release as other listeners did. Unfortunately, the relationship between masking release and forward masking was not demonstrated because the investigators did not measure forward masking.

Recently, Dubno *et al.* (2002) investigated the effect of aging on masking release and the possible contribution of simultaneous and forward masking thresholds to consonant recognition in modulated noise. Both younger and older listeners in that study had thresholds less than or equal to 20 dB HL at octave frequencies from 0.25 to 4.0 kHz. The average thresholds for older listeners were 10 dB higher than those for younger listeners. To minimize differences in audibility in quiet, Dubno *et al.* presented threshold-matching noise, which resulted in equivalent masked thresholds for all listeners. Results showed reduced masking release and poorer simultaneous and forward masking thresholds in older subjects than those in younger listeners. Older subjects also showed reduced recovery from forward masking compared to younger listeners. Dubno *et al.* concluded that speech rec-

TABLE I. Audiometric hearing thresholds in dB HL as a function of frequency in Hz.

	500	1000	2000	3000	4000	6000	8000
Avg NH	16	12	14	13	14	13	15
HI1	30	45	50	55	55	45	40
HI2	55	70	55	45	50	60	55
HI3	15	15	20	45	65	55	55
HI4	5	15	45	45	55	60	55
HI5	25	25	25	35	45	50	55
HI6	20	40	40	45	50	55	60
HI7	35	55	60	65	70	75	75
HI8	20	25	25	50	40	50	45
HI9	15	15	15	35	50	50	55
Avg HI	24	34	37	47	53	56	55

ognition in modulated noise might be associated with age-related increases in thresholds in forward and simultaneous masking.

However, the measurement of consonant recognition in fluctuating noise introduced some variability in the Dubno *et al.* (2002) study. The timing of each syllable presentation was random and independent of the noise. That is, the location of the consonant portion of the syllable was random with respect to the peaks and valleys of the noise. Critical acoustic cues for consonant recognition mostly lie in the brief onsets of consonants. If 50 ms on- and off-periods of modulating noise (the noise was modulated by a 10 Hz square wave) were randomly placed in the CV syllable, the critical information could be completely covered by the "on" period of noise or not covered at all. In that case, it is not clear whether listeners truly had masking release (listening in dips) or that the critical consonant cues were masked by chance. To test masking release for CV syllables, it may be necessary to present modulated noise at a fixed delay with respect to the onset of consonant rather than at a random delay. In addition, Dubno *et al.* tested forward masking at only one masker delay. Additional detail is needed to fully understand the relationship between forward masking and performance in gated noise.

The relationship between suprathreshold abilities and speech preception in modulated noise is not clear from the studies by Bacon *et al.* (1998) and Dubno *et al.* (2002). In the current investigation, differences in performance of NH and HI listeners for speech perception (sentence and consonant recognition) were examined in various types of noise. A primary goal was to investigate how temporal resolution associated with cochlear hearing loss plays a role in the perception of speech in modulated noise. To lessen the effects of reduced hearing sensitivity on performance of HI listeners both speech and noise for individual listeners were processed through linear amplification with frequency-response shaping based on his/her hearing sensitivity. Amplification was applied based on half-gain rule settings, so that HI listeners would obtain maximal signal audibility. Our goal was that HI listeners would be amplified sufficiently to achieve performance in quiet and in steady noise that was equivalent to NH listeners. Both sentences and consonants were selected as stimuli in an attempt to describe the potential effects of temporal resolution on different types of speech stimuli. For consonant stimuli, both fixed and random timing was employed to further describe potential effects of prior stimulation on the onsets of consonants. In addition, differences in the age range between NH and HI group were made to be small in order to minimize any effect of aging.

Three questions are addressed here. First, when HI listeners are amplified so that their performance is equal to that of NH listeners in quiet and in steady noise, do HI listeners still show reduced masking release when compared to NH listeners? Second, is the masking release the same for sentences and CV syllables? And finally, does forward masking explain a significant proportion of the variability in performance among HI listeners?

II. METHODS

A. Participants

Two groups of native speakers of English, NH and HI, participated in the current study. The age range for the NH group was 20–25 years and for the HI group 20–52 years. The NH group consisted of eight listeners with thresholds no more than 15 dB HL across audiometric frequencies (250–8000 Hz). The HI group consisted of nine individuals with sensorineural hearing loss who have thresholds more than 20 dB HL but no more than 70 dB HL across audiometric frequencies. The individual thresholds for the HI group are shown in Table I.

B. Stimulus preparation and procedures

1. Speech recognition

For speech recognition in noise, sentence recognition and consonant recognition in consonant-vowel (CV) syllables were used. For the sentence recognition test, IEEE sentences spoken by five male and five female talkers were used. Blocks of ten sentences were presented with each block containing one sentence spoken by each talker. Sentences were presented in random order. Each sentence consisted of five key words. Noise stimuli were generated by using the Tucker-Davis wave form generator (TDT WG1) and passed through a 30-band equalizer to match the spectrum of noise with the long-term spectrum of the IEEE sentences.

Noise stimuli were steady or gated with 2 ms cosine-squared ramps, implemented with 50% duty cycle and 100%

modulation depth. The modulation rates were 8 and 16 Hz, which results in the duration of the peaks of the modulating noise at 62.5 ms and 31.25 ms, respectively. Modulation rates of 8 and 16 Hz were selected because Nelson *et al.* (2003) found that the most masking release in NH listeners occurred at these rates. SNR was computed based on the long-term rms of the noise and the speech. SNRs of −5 and −10 dB were used for both NH and HI listeners. The noise always started 10–100 ms prior to the onset of sentences and ended briefly after each sentence. Two blocks of sentences were used per condition. A total of 12 blocks (2 SNRs × 3 modulation rates × 2 repetitions) in each noise condition and 2 blocks in the quiet condition were presented to each listeners in a random order.

Speech stimuli for consonant recognition consisted of 69 CV syllables made by combining 23 English consonants with the vowels /a/, /i/, and /u/ that were spoken by a female talker. Each syllable was presented in random order per block. These syllables were presented in steady-state noise, random-phase modulating noise, and controlled-phase modulating noise at −5 and −10 SNR. Random-phase modulating noise was similar to what Dubno *et al.* (2002) employed. That is, a period of peaks and valleys in modulating noise were implemented in speech randomly so that they were independent of the onset of CV syllables. In the controlled-phase noise condition, on the other hand, peaks of the modulating noise were implemented in a way that the onset of the consonant (which contains critical cues for consonant identification) occurred at the midpoint of a dip in the masking noise. Both random and controlled modulated noises had modulation rates of 8 and 16 Hz. Speech stimuli (both IEEE sentences and CV syllables) were presented about 40 dB above the pure tone average in order to minimize differences in audibility between groups. Listeners were asked to repeat back what they heard after the presentation of each sentence of CV syllable. The experimenter circled the responses on an answer sheet.

For NH listeners, the overall level of speech was fixed at 70 dB SPL. For each HI listener, however, the overall level of the speech and noise (for both sentence and syllable conditions) was adjusted based on his/her hearing sensitivity. For example, from Table I, HI1 has a mild-to-moderate hearing loss at mid to high frequencies for the tested ear. To amplify the level of speech and noise at the frequency range where HI1 has the most hearing loss but to maintain the level at other frequencies, speech and noise were passed through a Rane GE 60 graphic equalizer shaped to approximate the half-gain rule for each individual HI listeners. For NH listeners, natural speech and broadband noise shaped by the long-term spectrum of either IEEE sentences or CV syllables were presented without any amplification. Figure 1 shows the levels of IEEE sentence and speech-shaped noise at −10 dB SNR compared to the hearing sensitivity for NH group on average and for HI1.

2. Forward masking thresholds

The temporal resolution of each participant was examined by using a forward masking paradigm. The levels of the probe tone that were required for a listener to detect the probe following the noise masker were measured for 20 ms probes of 2000 and 4000 Hz. These frequencies were selected because the critical information for speech recognition is located primarily in the higher frequency areas in comparison to low frequencies (French and Steinberg, 1947). In support of this, Dubno *et al.* (2002) reported that the correlation between speech recognition scores and forward masking thresholds at 2000 and 4000 Hz was higher than the correlation between speech recognition and forward masking at 500 and 1000 Hz. Participants' absolute thresholds in quiet were measured for the two tones before measuring masked thresholds in the forward masking paradigm.

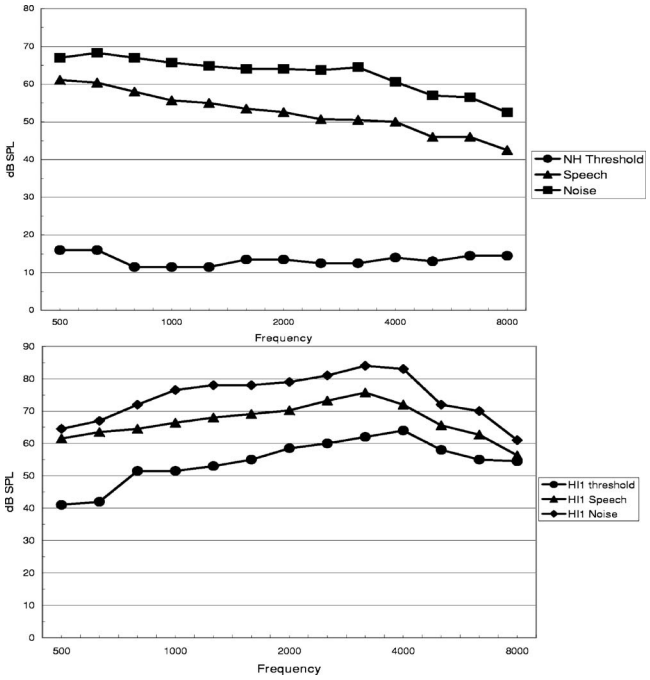


FIG. 1. Long-term spectra for IEEE sentences and speech-shaped noise along with hearing thresholds on average in dB SPL for the NH group in top panel and for hearing-impaired subject 1 (HI1) in the lower panel.

The masking noise was speech-shaped broad band noise that was used in the sentence recognition test. For HI listeners, the 300 ms noise was presented at the two levels that were used to produce −5 and −10 dB SNR in the sentence recognition test. Since the speech signals were presented at 40 dB SL, the absolute level of the noise (in dB SPL) for the HI listeners varied depending on his/her thresholds. Therefore, for HI listeners, the noise levels were 45 or 50 dB SL. The noise levels for all NH listeners were fixed at 70 and 85 dB SPL for two reasons. First, speech-shaped noise at 70 dB SPL represents a typical conversation level. Second, previous investigations (Nelson and Freyman, 1987; Dubno and Dirks, 1989) suggested that some HI listeners show greater amounts of masking only because the higher noise levels in dB SPL (compared to those for NH listeners) were used to compensate for loss of hearing sensitivity. Therefore, in the current investigation, 85 dB SPL was used for the NH group in order to compare temporal resolution at an approximately matched absolute noise level.

The duration between the offset of the masker and the onset of the signal (signal delay) were 8, 16, 32, and 64 ms. There were a total of 18 conditions: 2 frequencies $\times 2$ masker levels $\times 4$ signal delays ($n=16$) plus 2 thresholds in quiet (total $n=18$ conditions). A threshold for each condition was obtained three times. If the variance among the thresholds in a condition were greater than 2 dB, the same condition was repeated until any combination of three thresholds varied within 2 dB. All thresholds were obtained using a three-interval forced-choice adaptive procedure (Levitt, 1971) with a decision rule that estimates the signal intensity required for 79% correct performance. The blocks ended after the 12th reversal. A threshold was calculated from the latter 8 of 12 reversals. The step size was 4 dB for the first two reversals and then decreased to 2 dB for the remainder of the block. A trial consisted of a 200 ms warning interval, three observation intervals separated by 400 ms, and an answer interval terminated by the listener's responses. During each answer interval, the listeners were asked to press either 1, 2, or 3 from a keyboard indicating which observation interval she/he heard the signal. Each interval was marked by a visual cue on the computer monitor. Correct answer feedback was presented on the monitor after each listener's response. Listeners were seated in a sound-treated room and heard a tone and noise through an earphone. The visual cues such as "Ready," "Listen," and "Respond" appeared on the monitor in front of the listeners. Each subject had at least 1 h of practice before data collection began.

C. Analysis

Results of each task were entered into Analysis of Variance (ANOVA) with repeated measures to examine differences between groups (NH versus HI) and main effects within a group (SNR, gate frequencies, signal delay). Thresholds from the temporal masking experiment and hearing sensitivity were correlated with the amount of release from masking in sentence recognition as well as in syllable identification. Those with high correlation coefficients ($r>0.6$) were included in a regression analysis to find which test results best predicted performance in speech recognition in gated noise. For example, if forward masking thresholds and auditory thresholds at 2000 Hz were highly correlated with the masking release in sentence recognition, two regression models were constructed and the fits of the two equations were compared using an F test. Based on these findings, we might be able to further attribute a lack of masking release with deficits in temporal resolution as well as loss of hearing sensitivity.

III. RESULTS

A. IEEE sentence recognition

Figure 2 shows the percent correct key word identification in different noise conditions (no noise, steady, and gated noise) for both NH and HI listeners at the SNR of −5 and −10 dB, respectively. For each noise condition, solid black bars on the far left side represent average percent correct and standard deviation for NH group. Each bar in different patterns shows the performance of each HI listener. The mean key word identification score in quiet was 97.8% (s.d.=3.7) for the HI group and 99.8% (s.d.=0.46) for the NH group. At −5 dB SNR, the range of performance for HI listeners in steady noise was from 34% to 52% ($M=41.5$, s.d.=5.9) and from 37% to 55% ($M=44.4$, s.d.=7.1) for NH listeners. At −10 dB SNR, HI listeners performed from 7% to 19% ($M=10.6$, s.d.=5.8) and for the NH group scores ranged from and 4% to 21% ($M=12.2$, s.d.=4.5). Apparently, the two groups performed very similarly in quiet and steady noise.

To avoid possible ceiling and floor effects, the results were converted using the arcsine transformation [$y=\text{arcsine}(\sqrt{\text{PC}\times 100})$] where PC is the percent correct scores and y is the transformed data. According to Howell (1997), the arcsine transformation is useful to stretch out both ends of data relative to the middle. Transformed data were then analyzed using an ANOVA with a group as between subject factor and the percent correct scores as a within subject factor. Even with transformed data, group differences were not significant in quiet [$F(1,15)=2.007$, $p=0.177$] and in steady noise [$F(1,15)=3.518$, $p=0.08$].

This suggests that the presentation levels of amplified speech for each HI listener yielded equivalent audibility to those for NH group. As a result, both NH and HI listeners showed comparable performance in quiet and in steady noise.

When noise was modulated, however, performance of HI listeners was significantly poorer than that of the NH group as shown in Fig. 2. Whereas NH listeners identified 81%–97% key words at −5 SNR and 64%–96% at −10 SNR for both gate frequencies, HI listeners only scored 36%–74% and 13%–58%, respectively.

Figure 3 shows the amount of improvement (or masking release) as the difference between the percent correct scores for gated noise and in comparison with those for steady noise. The improvement of the NH listeners from gating was about 50% for −5 dB SNR and 70% for −10 dB SNR. This suggests when background noise gets louder, NH listeners benefit more from momentary gaps in the gated noise in terms of absolute change. For the HI group, the amount of masking release (MR) ranged from −8% to 34% at −5 dB SNR and from −3% to 45% at −10 dB SNR. Those who had the most hearing loss (HI2 and HI6) showed the least masking release. Compared to the NH group, HI participants took less advantage of gating and the amount of MR was relatively constant across noise levels.

Since the baseline (percent correct in steady noise) and the best scores (percent correct in quiet) for each listener varied somewhat, it was worthwhile to examine a "normalized masking release" (NMR) relative to each listener's performance difference between baseline and best scores. That is, NMR represents a ratio of the amount of masking release obtained to the amount of possible benefit from gating. For example, HI2 had masking release of 34% at 8 Hz, −5 SNR. His baseline and best scores were 38% (with steady noise) and 99% (in quiet) and the difference between these was 61%. It can be assumed that HI2 could benefit from gating at most by 61%. However, his performance improved by only 34% out of the possible 61%, which results in a ratio of 0.56.

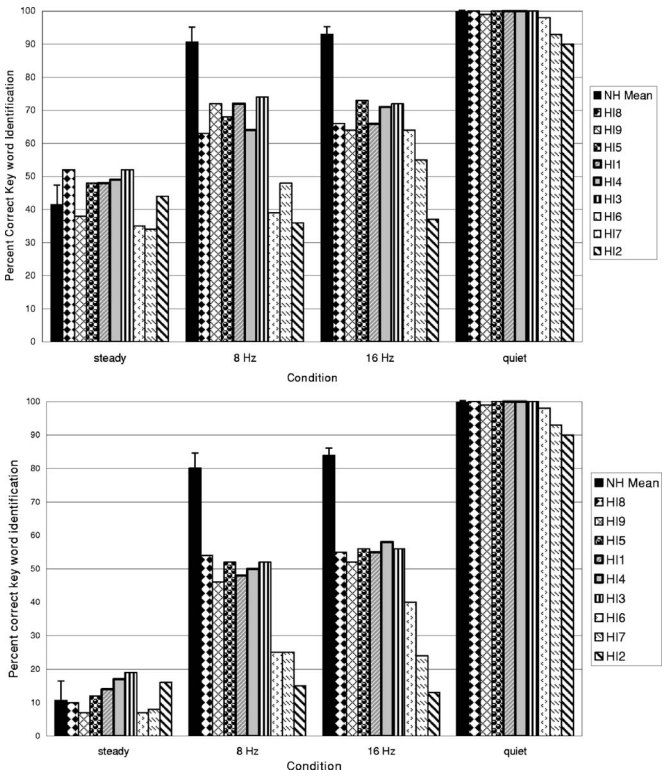


FIG. 2. Percent correct keyword identification of HI listeners and the NH group for sentence recognition at −5 dB SNR (upper panel) and −10 dB SNR (lower panel.) Note: the data for the quiet condition are the same in both upper and lower panels. They are shown for comparison in each panel.

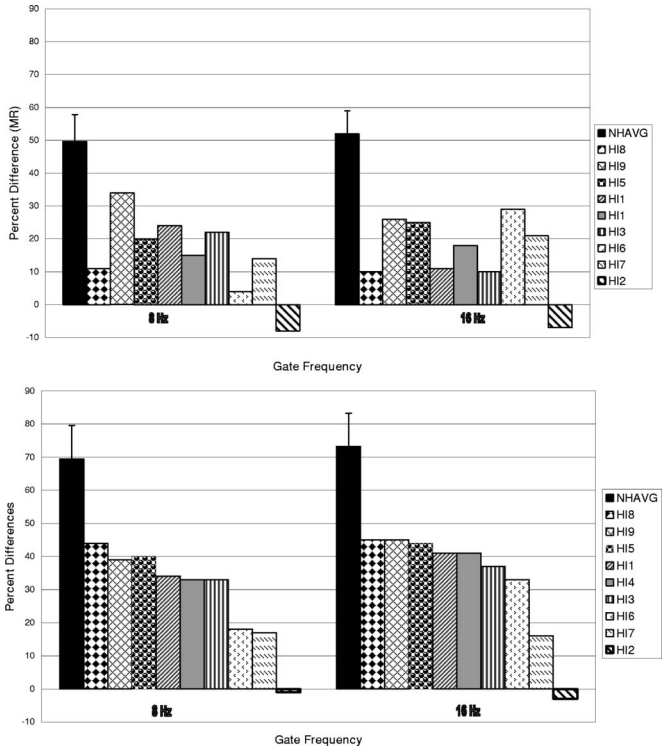


FIG. 3. The amount of masking release (MR) for individual HI listeners and for the NH group for sentence recognition at −5 dB SNR (upper panel) and −10 dB SNR (lower panel).

In the current study, this ratio is called a NMR. The amount of NMR shows that NH listeners could take advantage of gating about 0.8 (80% of the possible gain) whereas HI listeners only could gain 15% to 55%. Even considering those best HI listeners (HI1, HI3, and HI5, for example), their NMR was significantly less than those of NH listeners.

In a two-way ANOVA with repeated measures, group (either NH or HI) was a between-subject factor and the dependent variables were PC, MR, or NMR. ANOVA results indicate significant effects of group on all three dependent variables ($p<0.01$). For example, for NMR, there was significant effect of group [$F(1,15)=60.5$, $p<0.0001$]. In contrast, the level of noise (dB SNR) and gate frequency had no significant effect on NMR ($p>0.05$). There was no significant interaction. Clearly, the HI listeners obtained significantly less masking release than did NH listeners, despite the fact that their performance in quiet and in steady noise were not significantly different from NH listeners.

B. CV syllable recognition

Figure 4 shows the percent correct syllable identification of NH and HI listeners when the onset of the noise burst was either random or fixed relative to the onset of consonants at −5 dB SNR. Similarly, Fig. 5 demonstrates percent correct in fixed and random conditions at −10 dB SNR. In quiet, the mean syllable identification score was 93.25% (s.d.=2.71) for the NH group and 89% (s.d.=4.52) for the HI group. In the presence of steady noise, the means for NH listeners were 43% (s.d.=4.9) at −5 SNR and 12.5% (s.d.=4.5) at −10 SNR whereas those for HI listeners were 43% (s.d.=4.3) and 20.2% (s.d.=2.6), respectively. Consistent with the results from sentence recognition, the ANOVA indicated that the consonant identification scores of the NH and HI groups were similar to each other when syllables were presented in quiet [$F(1,15)=5.530$, $p=0.053$] or with steady noise [$F(1,15)=5.935$, $p=0.058$].

As with sentences, NH listeners performed significantly better than HI listeners did in the presence of gated noise. Figures 6 and 7 show the amount of MR for the different noise conditions. Compared to the random condition, the performance of each listener in both NH and HI groups seems to be better in the fixed condition across gate frequencies. A three-way ANOVA (2 SNR × 2 gate frequencies × 2 conditions) with repeated measures indicates significant effects of condition (random versus controlled phase) on MR [$F(1,15)=34.377, p=0.000$] and on NMR [$F(1,15)=59.87, p=0.000$]. Like the sentence recognition task, NH listeners showed significantly greater MR and NMR than did HI listeners for CV syllable recognition. The controlled-phase scores were significantly better than the random-phase scores. There were no significant differences between gate frequencies.

C. Temporal resolution

As stated previously, for HI listeners, the noise levels were 45 or 50 dB SL except HI2, HI5, and HI7. (HI5 could not finish all sessions because of her limited schedule for the study. HI2 and HI7 could not be tested at higher levels because the equipment was limited to provide 50 dB SL relative to their hearing sensitivity.) A three-way ANOVA with

TABLE II. Slopes of forward masking function.

	Low level		High level	
	2 kHz	4 kHz	2 kHz	4 kHz
Average NH	−3.8	−3.5	−4.69	−4.28
HI1	−1.84	−3.27	−2.8	−3.29
HI2	−2.1	−1.8	CNT	CNT
HI3	−2.98	−0.8	−2.95	−1.6
HI4	−2.68	−2.17	−2.96	−3.07
HI5	DNT	DNT	−1.6	−3.27
HI6	−3.11	−1.8	−4.1	−2.67
HI7	−2.9	−1.2	−2.62	CNT
HI8	−3.38	−2.3	−3.4	−2.4
HI9	−3.1	−2.49	−3.1	−2.25

repeated measures (probe frequency × noise level × time delay) for the NH listeners indicated that there was a significant effect of noise levels on masked thresholds at 2 kHz $[F(1,7)=20.475, p=0.005]$ but not at 4 kHz. The main effects of time delay on masked thresholds are significant for both 2 and 4 kHz tones.

When the forward masked thresholds were plotted as a function of probe delay in ms in log scale, those who had the lower thresholds in quiet showed the better thresholds in the presence of the masker across probe delays. This implies that the masked thresholds were greatly affected by hearing sensitivity for signal tones in quiet. Table II demonstrates the slope of each function of masked thresholds plotted along the time delay in log scale. This slope could be considered as a rate of recovery from nonsimultaneous maskers. NH and HI listeners with better thresholds in quiet at 2 kHz (HI 3, 6, 8, and 9) and at 4 kHz (HI6, 8, and 9) showed relatively faster recovery rate (steeper slope) than other listeners with greater hearing loss. A three-way ANOVA (probe frequency × noise level × time delay) with repeated measures found that both probe frequency $[F(1,15)=4.632, p=0.048]$ and time delay $[F(3,45)=347.464, p=0.000]$ had significant main effects. There was no main effect of the presentation level of the noise $[F(1,15)=0.011, p=0.918]$. Only the two-way interactions between time delay and other factors (group, probe frequency and level) showed significant effects. The differences between groups were also significant $[F(1,15)=106.542, p=0.000]$.

As indicated by Nelson and Freyman (1987), comparing the amount of masking rather than the masked thresholds might provide a way to examine temporal effects independent of hearing sensitivity. In order to examine the temporal resolution of individual participants with minimal effect of reduced hearing sensitivity, thresholds in quiet for each frequency tone were subtracted from the masked thresholds across time delay. Figure 8 demonstrates the amount of masking in dB SPL at 2 and 4 kHz tones for NH and HI listeners. The mean threshold for the NH group at the noise level of 70 dB SPL is plotted along with thresholds of HI listeners obtained at 45 dB SL. Similarly, the mean for the NH group at 85 dB SPL is charted with the results of HI listeners at 50 dB SL. A three-way ANOVA (probe frequency × noise level × time delay) with the amount of forward masking as a dependent variable and with a group as a between-subject factor showed significant group differences $[F(1,15)=16.730, p=0.001]$. The main effects of the level of the noise, and the time delay and only the interaction between group and time delay were significant. In general, NH listeners had greater amount of masking and yet faster recovery (steeper slope) than most of HI listeners with few exceptions.

D. Regression analysis

One of the main purposes of the current investigation was to determine the possible contributions of hearing sensitivity (audiometric pure tone thresholds) and temporal resolution to speech perception in gated noise. Therefore, the results of forward masking and hearing sensitivity were correlated with NMR at −5 SNR for all 17 listeners. When all factors and all listeners were included, the effect of hearing threshold explained a large proportion of the variance. The results showed that the correlation coefficients for the 8 Hz gate frequency were quite similar to those for 16 Hz. The hearing sensitivity of most audiometric frequencies and forward masked thresholds at 2 and 4 kHz were strongly correlated with NMR in sentence recognition ($r>0.8$) but not the amount of forward masking or the slope of forward masking functions. One possible reason for that is because the correlation might be overwhelmingly affected by hearing sensitivity when the results of the NH group were included. Compared to the NH group whose hearing sensitivity and performance in different tasks were very uniform, HI listeners in the current study had various hearing thresholds across the audiometric frequencies and showed different performance levels. Therefore, if only the results of HI listeners were correlated, it might be possible to observe the differential contribution of hearing sensitivity and suprathreshold abilities that affect masking release of HI listeners. Thus, only the results from HI listeners were included in the analysis. Table III shows the relationship between NMR for sentences, hearing sensitivity, and the performance in forward masking (masked thresholds, masking amount, and slope) for HI listeners only. The masking amount was the difference between the threshold of 2 kHz (or 4 kHz) in quiet and the forward masked threshold. Only a few factors (hearing sensitivity at 0.5 and 1 kHz and forward masking threshold at 2 kHz) retained a relatively strong correlation ($r>0.6$) with NMR for sentences. These are highlighted in bold in Table III.

Using these factors, three different regression models were constructed in order to analyze the differential contributions of hearing sensitivity and forward masking threshold to NMR. The NMR for 8 and 16 Hz gate frequencies at −5 dB SNR were analyzed separately because the correlation coefficients were slightly different for the two gate frequencies. The full model containing all significant factors predicted NMR at 8 Hz gating based on performance in forward masking threshold at 2 kHz as well as hearing sensitivity for 0.5 and 1 kHz tones. The fall model for NMR for 16 Hz gating was similar to the one for 8 Hz. Sub-model 1 predicted NMR based on hearing sensitivity only whereas sub-

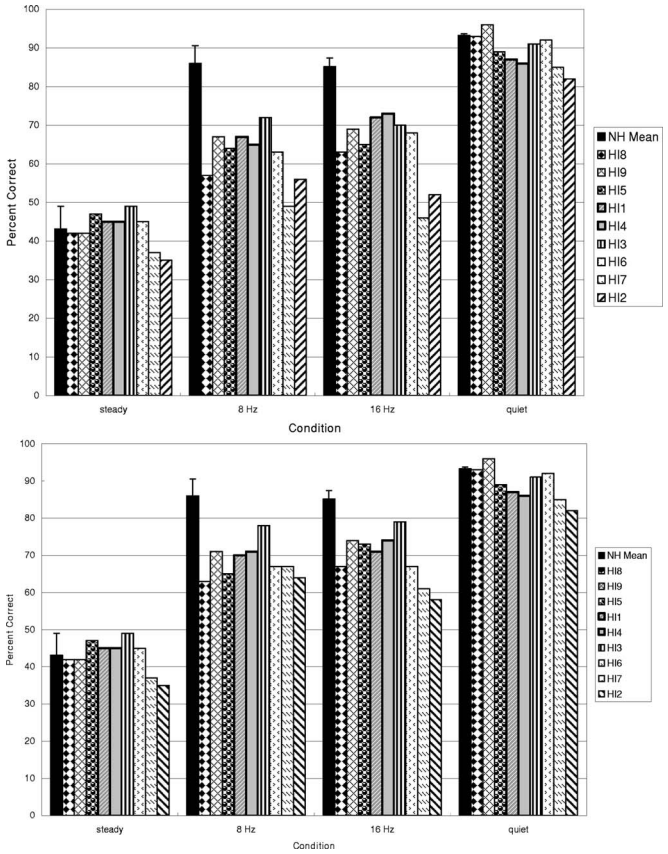


FIG. 4. Percent correct CV syllable identification for individual HI listeners and the NH group for the random phase (upper panel) and the fixed phase (lower panel) conditions at −5 dB SNR. Note: the data for the quiet condition are the same in both upper and lower panels. They are shown for comparison in each panel.

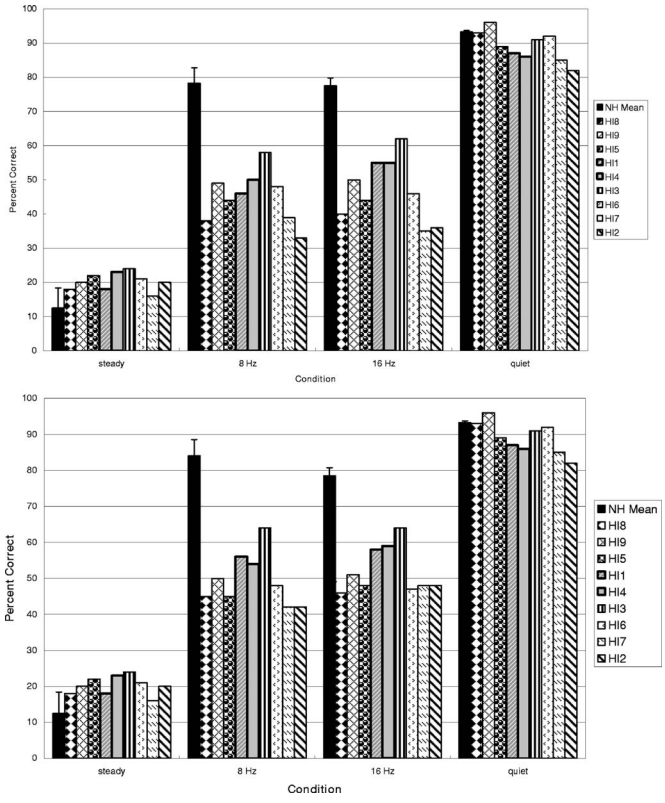


FIG. 5. Percent correct CV syllable identification for individual HI listeners and the NH group for the random phase (upper panel) and the fixed phase (lower panel) conditions at −10 dB SNR. Note: the data for the quiet condition are the same in both upper and lower panels. They are shown for comparison in each panel.

model 2 only included forward masking thresholds (eliminating hearing sensitivity) (see Table IV). Three regression models were constructed to predict the NMR at 16 Hz for −10 SNR. The MI model consisted of forward masking threshold at 2 kHz as well as hearing sensitivity for 0.5 and 1 kHz tones. Sub-model 1 predicted NMR based on hearing sensitivity only whereas sub-model 2 only included forward masking thresholds.

The results show that the dependent variable was well predicted by the factors in the model. The multiple-correlation-squared (R^2) for each model indicates the strength of the relationship between the NMR and the predictors such as hearing sensitivity or forward masking thresholds. For most of the sentence recognition conditions, when the hearing sensitivity at low-to-mid frequencies was taken out of the model (which becomes the sub-model 2), the R^2 decreased considerably. This suggests that hearing sensitivity at low-to-mid frequencies accounts for a substantial amount of the variance in the NMR for sentence recognition. The toward masking threshold did not seem to play a large role in sentence recognition in gated noise because the differences in R^2 between the full model and sub- model 2 (which consists of hearing sensitivity alone with forward masking removed) were relatively small.

Table V shows the correlation between the NMR for syllable identification and the results of other tasks at −5 SNR. Compared to those for sentences, hearing sensitivity across frequencies was not strongly related to the NMR. Only the forward masked thresholds and the masking amount (highlighted in bold in Table V) show moderate correlation ($r > 0.6$). There were no strong correlations between any factors and NMR at −10 SNR. (A full regression analysis for syllable identification is not included here because very few factors showed strong correlation with masking release for CV syllables.)

IV. GENERAL DISCUSSION

Common environmental noises are typically temporally fluctuating. It is important to understand how HI listeners perform when background noise is fluctuating, or gated, rather than steady. In these situations, it may be that performance would not be easily predicted by signal audibility alone. While we cannot completely rule out subtle audibility differences between NH and HI listeners, nevertheless HI listeners in this study were sufficiently amplified so that their performance was comparable to that of normal-hearing listeners in quiet and steady noise. Even under those nearly ideal amplification conditions, the HI listeners showed significant decrements in masking release. This finding was true for both sentences and for CV syllables, suggesting that masking release cannot be accounted for simply on the basis of audibility of speech in noise. When audibility differences between NH and HI listeners were reduced, modest effect of forward masking was seen for CV syllable recognition.

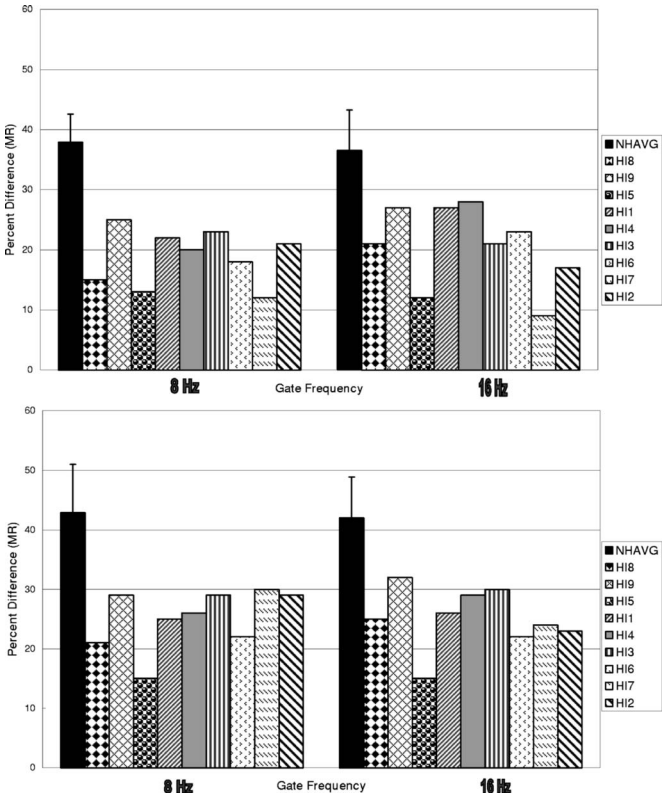


FIG. 6. The amount of MR for individual HI listeners and the NH group for the CV syllables in random phase (upper panel) and fixed phase (lower panel) conditions at −5 dB SNR.

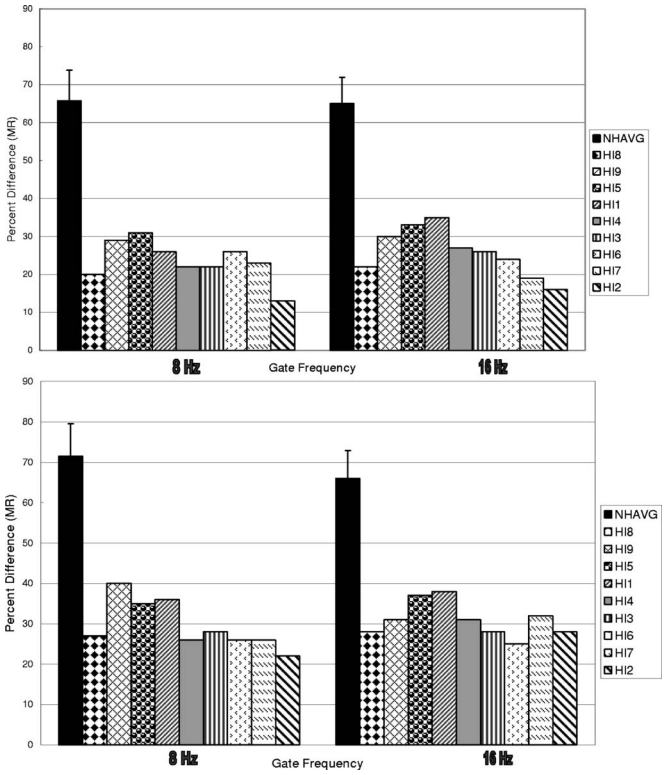


FIG. 7. The amount of MR of individual HI listeners and the NH group for the CV syllables in random phase (upper panel) and fixed phase (lower panel) conditions at −10 dB SNR.

A. Sentence recognition

These findings extend the findings of previous investigations in which amplified speech and noise were used for HI listeners and show that even with appropriate amplification, HI listeners experience significantly reduced MR. For example, Peters *et al.* (1998) employed a linear amplification formula with appropriate frequency response shaping that is commonly used for a hearing aid. The amplification substantially improved. SRTs in quiet and steady noise for the younger and older HI participants. Compared to SRTs for NH listeners, however, the SRTs of the HI listeners were still poorer. Therefore, the amplification used in previous investigations did not seem to provide an adequate amount of speech information to compensate for individual hearing loss. Even when sufficient amplification was used HI listeners were less able to use speech momentary gaps in gated noise. Some investigators (Trine, 1995; Van Tasell, 1993; Peters *et al.*, 1998) have tried to explain deficits of HI participants in gated noise as a result of reduced audibility. That is, even though speech stimuli are audible, those momentary decreases in the intensity of the masker might be below the hearing thresholds so that full "dips" are not available. Therefore, HI listeners would not have as much masking release as NH listeners. In addition, Peters *et al.* (1998) found that even with the amplification, the SRT of HI listeners was worse than the SRT of NH listeners in a background noise with spectral and/or temporal gaps. It was suggested that the linear amplification might not provide sufficient gain to restore audibility of the entire dynamic range of speech.

However, the lack of audibility alone does not seem to explain the reduced masking release of HI listeners in the current investigation because the sentence recognition of HI listeners in quiet (and in steady noise) was very similar to that of NiH listeners. As an alternative explanation, Dubno *et al.* (2002), and Bacon *et al.* (1998), proposed that abnormal forward masking might predict the rack of the masking release in HI listeners. In the current study, however, forward masking does not appear to predict MR for sentence materials. The amount of forward masking varied among the HI listeners, though typically when FM thresholds and quiet thresholds were compared, HI listeners showed less forward masking at short masker durations than did NH listeners. In general, their FM function slopes were shallower than those of NH listeners. This is consistent with previous investigations such as Nelson and Freyman (1987). Correlation coefficients between sentence recognition and amount of FM were low and regression models indicate that the addition of FM to the models adds little to the prediction of performance, apart from the hearing thresholds of the listeners.

To explain the lack of MR in sentence recognition of HI listeners, possible contribution of alternative suprathresholds will be investigated as a follow-up. Peters *et al.* (1998), Mackersie *et al.* (2001) and other investigators speculated that reduced spectral resolution and auditory integration due to the cochlea hearing loss could be responsible for the reduced masking release.

B. CV syllable identification

The results of the current CV syllable identification task were consistent with those of the sentence recognition task.

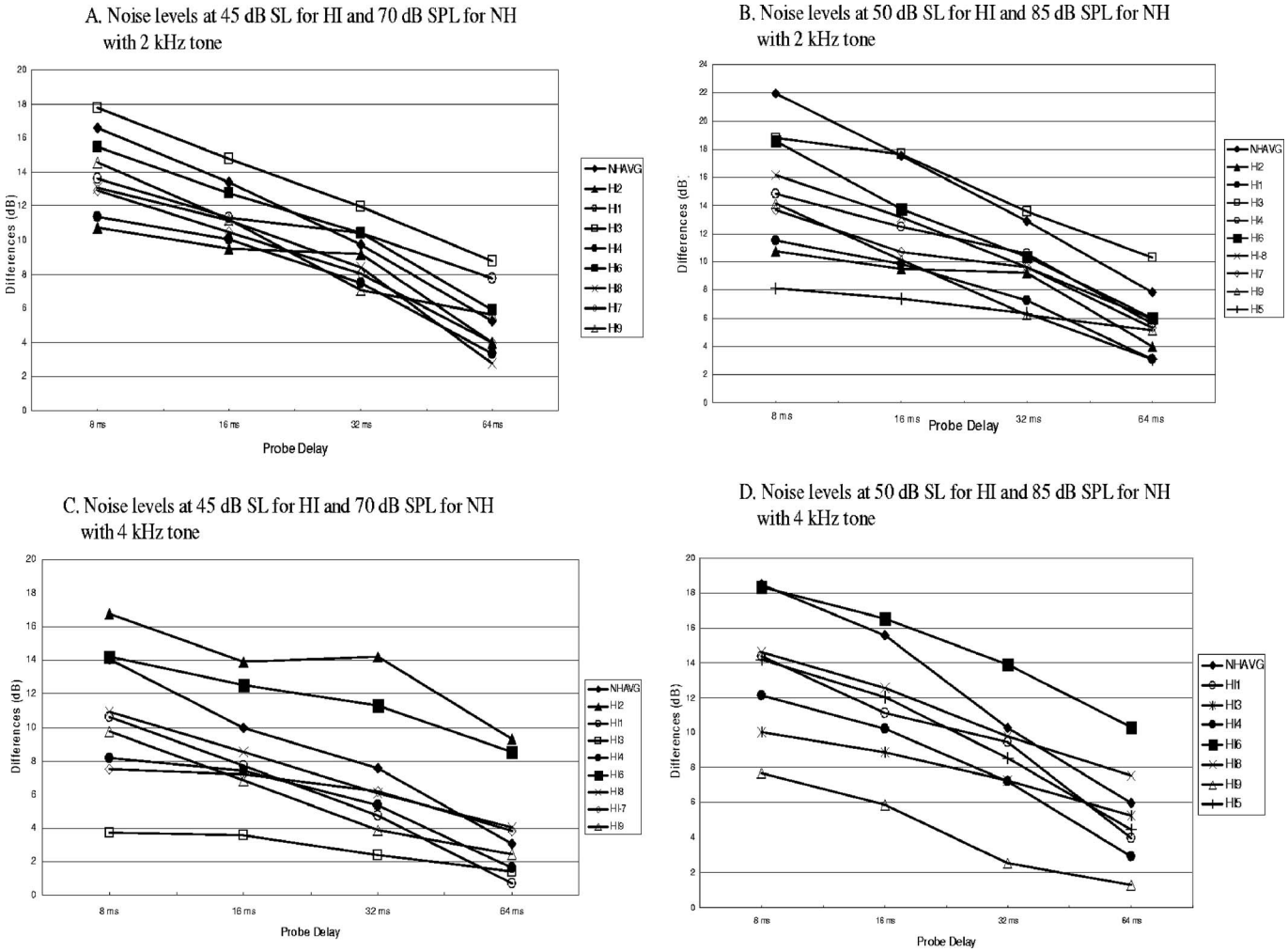


FIG. 8. The amount of forward masking of individual HI listeners at 45 dB SL [(a) and (c)] and 50 dB SL [(b) and (d)] compared to the group average performance of NH listeners at 70 dB SPL [(a) and (c)] and 85 dB SPL [(b) and (d)].

That is, the scores of NH listeners in quiet and in steady noise were not significantly different from those of HI listeners whereas the performance in the gated noise was substantially poorer for HI listeners than NH listeners. The percent correct scores, the amount of masking release, and the normalized masking release were analyzed separately as dependent variables. Consonant recognition of NH and HI listeners was also examined in the presence of no noise, steady noise, and gated noises. Dubno and her colleagues (2002) also investigated benefits of modulated maskers for syllable recog-

TABLE III. Correlation coefficient between normalized masking release sentence recognition and psychoacoustical measures for HI listeners only.

	NMR −5 dB		−10 dB	
Correlation coefficient	8 Hz	16 Hz	8 Hz	16 Hz
Hearing sensitivity (0.25 kHz)	−0.28	−0.55	0.02	−0.6
Hearing sensitivity (0.5 kHz)	**−0.6**	**−0.73**	**−0.33**	**−0.8**
Hearing sensitivity (1 kHz)	**−0.67**	**−0.62**	**−0.29**	**−0.84**
Hearing sensitivity (2 kHz)	−0.55	−0.39	−0.05	−0.57
Hearing sensitivity (4 kHz)	0.12	0.04	0.42	−0.36
Masked threshold (2 kHz)	**−0.61**	**−0.6**	**0.04**	**−0.62**
Masked threshold (4 kHz)	−0.14	−0.14	−0.61	−0.12
Masking amount (2 kHz)	−0.3	0.27	−0.03	0.34
Masking amount (4 kHz)	0.3	−0.27	−0.17	−0.19
Slope (forward masking function, 2 kHz)	0.044	−0.27	0.03	−0.24
Slope (forward masking function, 4 kHz)	−0.27	−0.36	0.18	−0.57

TABLE IV. Summary of regression model analysis (IEEE sentences).

Dependent variable	Full model	Sub-model 1	Sub-model 2
	NMR at 8 Hz in −5 SNR		
R	0.71	0.68	0.43
R^2	0.51	0.46	0.19
F statistic	$F(3,5)=1.73$	$F(2,6)=2.55$	$F(1,7)=1.60$
p value	0.28	0.16	0.25
	NMR at 16 Hz at −5 SNR		
R	0.76	0.74	0.34
R^2	0.57	0.55	0.12
F statistic	$F(3,5)=2.24$	$F(2,6)=3.68$	$F(1,7)=0.92$
p value	0.20	0.09	0.37
	NMR at 16 Hz at −10 SNR		
R	0.85	0.84	0.62
R^2	0.72	0.71	0.39
F statistic	$F(3,5)=4.29$	$F(2,6)=7.65$	$F(1,7)=4.38$
p value	0.075	0.02	0.075

TABLE V. Correlation coefficient between NMR in syllable identification and psychoacoustical measures at −5 SNR for HI listeners only.

	Norm masking release			
Correlation coefficient	8*R* Hz	8*C* Hz	16*R* Hz	16*C* Hz
Hearing sensitivity (0.25 kHz)	0.04	0.5	−0.4	−0.04
Hearing sensitivity (0.5 kHz)	−0.06	0.28	−0.45	−0.3
Hearing sensitivity (1 kHz)	−0.13	0.3	−0.37	−0.31
Hearing sensitivity (2 kHz)	−0.19	0.34	−0.2	−0.16
Hearing sensitivity (4 kHz)	−0.26	−0.43	−0.2	−0.14
Masked threshold (2 kHz)	−0.2	−0.19	−0.6	−0.32
Masked threshold (4 kHz)	0.01	−0.67	−0.39	0.15
Masking amount (2 kHz)	**0.82**	**0.62**	**0.4**	**0.5**
Masking amount (4 kHz)	**−0.23**	**−0.12**	**−0.31**	**−0.62**
Slope (forward masking function, 2 kHz)	−0.09	0.001	−0.47	−0.32
Slope (forward masking function, 4 kHz)	−0.07	0.64	−0.5	0.25

nition (either CV or VC syllables) only when the timing of syllable presentation was random and independent of the noise cycle. Two conditions (either random or control phase) were implemented in the current experimental design. The results show that when the periods of gated noise were controlled relative to the onset of consonants, the performance of both NH and HI listeners was significantly better than the random condition regardless of the noise level. In other words, if the peaks of noise were always placed shortly after the onset of consonants as in controlled conditions the listeners were able to hear out more cues for each consonant. Thus randomly phased noise seemed to be more effective maskers for syllable recognition and are more realistic since fluctuating noise would occur rather by chance in a real life. However, when syllables are used in a speech recognition task, interpretation of th results might need to be made more cautiously.

The recent study by Dubno *et al.* (2002) compared syllable identification of young NH listeners to those of older listeners with near normal hearing up to 4 kHz. Although the older listeners have near normal thresholds, the average thresholds of older group were 10 dB higher at mid-to-high frequency than younger group. The investigators found that forward-masked thresholds of the elderly listeners were elevated and the poorer asked thresholds were significantly correlated with the elderly listeners' reduced masking release. In the current study, the amount of forward masking at 2 kHz was positively correlated with CV recognition in gated noise. (Refer to Table 5.) Interestingly, forward masking was more highly correlated with CV recognition in noise gated at 8 Hz than in noise gated at 16 Hz, which is contrary to the assumption that shorter gap durations in the noise would be more negatively affected by abnormal forward masking.

C. Differences between sentence and CV materials

Over the decades, various investigators have demonstrated the differences in the amount of masking release between NH and HI listeners with a range of speech materials. For example, Dubno *et al.* (1989, 2002) used CV or VC syllables, while Eisenberg *et al.* (1995) used nonsense syllables and SPIN sentences. HINT or IEEE sentences were employed in other studies (Bacon *et al.*, 1998; Mackersie *et al.*, 2001; Nelson *et al.*, 2003). The results from the current investigation suggested similar principles underlying speech recognition in fluctuating noise would be demonstrated regardless of the type of speech stimuli. That is, most listeners are able to benefit from momentary dips in noise, and the amount of masking release varies corresponding to hearing sensitivity of listeners. A simple ANOVA with stimulus type (either IEEE sentence or CV syllable) as a dependent variable showed that when the SNR is fixed, the differences in NMR for CV syllables (in either random or control condition) and IEEE sentences were not significant. Speech material types do not seem to play an important role in measuring speech perception in fluctuating noise. However, it does appear that forward masking has significantly stronger relationship to CV identification than to sentence recognition.

V. CONCLUSION

Listeners with sensorineural hearing loss demonstrate significantly reduced masking release from gated noise, even when stimuli are amplified so that their performance matches that of normal-hearing listeners in quiet and in steady noise. This finding was true for both sentence and CV syllable stimuli. This finding has implications for understanding the common complaints of HI listeners in background noise. Amplification may not be expected to fully restore understanding to normal levels in fluctuating noise. Abnormal forward masking seems to be an incomplete explanation of these findings, especially for recognition of sentences. Further studies will be undertaken that will continue to explore additional factors such as spectral resolution and auditory grouping, which may help to explain the findings further.

Bacon, S. P., Opie, J. M., and Montoya, D. Y. (**1998**). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," J. Speech Lang. Hear. Res. **41**, 549–563.

Dubno, J. R., and Dirks, D. D. (**1989**). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1666–1675.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2002**). "Benefit of modu-

lated maskers for speech recognition by younger and older adults with normal hearing," J. Acoust. Soc. Am. **111**, 2897–2907.
Eisenberg, L. S., Dirks, D. D., and Bell, T. S. (**1995**). "Speech recognition in amplitude-modulated noise of listeners with normal and listeners with impaired hearing," J. Speech Lang. Hear. Res. **38**, 222–233.
Festen, J. M. (**1993**). "Contributions of comodulation masking release and temporal resolution to the speech-reception threshold masked by as interfering voice," J. Acoust. Soc. Am. **94**, 1295–1300.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Glasberg, B., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear hearing impairments and their relationship to the ability to understand speech," Scand. Audiol. **32**, 1–25.
Gustafsson, H. A., and Arlinger, S. D. (**1994**). "Masking of speech by amplitude-modulated noise," J. Acoust. Soc. Am. **95**, 518–529.
Howard-Jones, P. A., and Rosen, S. (**1993**). "The perception of speech in fluctuating noise," Acustica **78**, 258–272.
Howell, D. C. (**1997**). *Simple Analysis of Variance: Transformations. Statistical Methods for Psychology*, 4th ed. (Duxbury, Belmont, CA), pp. 323–329.
Kidd, G., Mason, C. R., and Feth, L. L. (**1983**). "Temporal resolution of forward masking in listeners having sensorineural hearing loss," J. Acoust. Soc. Am. **75**, 937–944.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Mackersie, C. L., Prida, T. L., and Stiles, D. (**2001**). "The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentence by listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **44**, 19–28.
Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Growth of forward masking for sinusoidal and noise maskers as a function of signal delay," J. Acoust. Soc. Am. **73**, 1249–1259.
Nelson, D. A., and Freyman, R. L. (**1987**). "Temporal resolution in sensorineural hearing-impaired listeners," J. Acoust. Soc. Am. **81**, 709–720.
Nelson, P. B., Jin, S-H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.
Oxenham, A. J., and Plack, C. J. (**1997**). "A behavioral measure of basilar membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.
Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). "An Articulation Index based procedure for predicting the speech recognition performance of hearing-impaired individuals," J. Acoust. Soc. Am. **80**, 50–57.
Peters, R. W., Moore, B. C., and Baer, T. (**1998**). "Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people," J. Acoust. Soc. Am. **103**, 577–587.
Plack, C. J., Oxenham, A. J., and Drga, V. (**2002**). "Linear and nonlinear processes in temporal masking," Acta Acust. **88**, 348–358.
Ruggero, M. A., Rich, N. C., Reico, 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.
Trine, T. D. (**1995**). "Speech recognition in modulated noise and temporal resolution: Effects of listening bandwidth," Doctoral dissertation, University of Minnesota, Twin Cities (unpublished).
Van Tasell, D. J. (**1993**). "Hearing loss, speech, and hearing aids," J. Speech Hear. Res. **36**, 228–244.
Zurek, P., and Delhorne, L. (**1987**). "Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment," J. Acoust. Soc. Am. **82**, 1548–1559.

A rule-based emotion-dependent feature extraction method for emotion analysis from speech

Vladimir Hozjan and Zdravko Kačič[a)]
Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova ul. 17, SI-2000 Maribor, Slovenia

(Received 28 July 2003; accepted 21 February 2006)

This paper presents a rule-based method to determine emotion-dependent features, which are defined from high-level features derived from the statistical measurements of prosodic parameters of speech. Emotion-dependent features are selected from high-level features using extraction rules. The ratio of emotional expression similarity between two speakers is defined by calculating the number and values of the emotion-dependent features that are present for the two speakers being compared. Emotional speech from Interface databases is used for evaluation of the proposed method, which was used to analyze emotional speech from five male and four female speakers in order to find any similarities and differences among individual speakers. The speakers are actors that have interpreted six emotions in four different languages. The results show that all the speakers share some universal signs regarding certain emotion-dependent features of emotional expression. Further analysis revealed that almost all speakers in the analysis used unique sets of emotion-dependent features and each speaker used unique values for the defined emotion-dependent features. The comparison among speakers shows that the expressed emotions can be analyzed according to two criteria. The first criterion is a defined set of emotion-dependent features and the second is an emotion-dependent feature value. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188647]

PACS number(s): 43.72.Ar, 43.70.Fq [DOS] Pages: 3109–3120

I. INTRODUCTION

Over the last few decades human emotions and behavior have been research topics within different scientific areas. Researchers from the fields of acoustics and prosody have also become interested in examining human emotions. Most research activities on emotional speech focus on analyzing the prosody of speech signals. The majority of papers about emotions in speech concerns the English language (Murray and Arnott, 1993; Gobl and Ni Chasaide, 2000; Cornelius, 1996). Very often emotional expression in speech is examined from acted speech (Kwon *et al.*, 2003; Ververidis *et al.*, 2004; Pereira, 2000). Emotions in acted speech are usually called simulated emotions. Over the last few years several analyses of emotional expression in speech have been accomplished, such as using a Wizard of Oz (WOZ) approach, from television talk shows, or data from call centers (Batliner *et al.*, 2003; Amir *et al.*, 2003; Ang *et al.*, 2002). These sources normally provide nonacted emotional expression, so-called authentic emotions. However, actors seem to interpret emotional speech with an artistic goal that is far from producing speech similar to nonacted speech (Auberge *et al.*, 2004). In acted emotional speech it is hard to evaluate whether the production of simulated emotions is identical to authentic emotions (Auberge *et al.*, 2004). Batliner *et al.* (2003) compares results from the automatic emotion recognition of simulated emotions as interpreted by actors, with the authentic emotions produced by the WOZ method. Simulated emotions in this research were recognized significantly better than authentic emotions. Douglas-Cowie *et al.* (2000), Batliner *et al.* (2000), and Murray and Arnott (1993) present and discuss, in detail, the usages of simulated and authentic emotions in the research of emotion expression in speech.

It is well-established that speech prosody also contains information about a speaker's emotions. Prosody is divided into several components: intonation, tempo, and loudness. Intonation is the changing voice pitch pattern over an utterance, used to convey information (Arnfield, 1994). Tempo is the speaking rate and is usually measured in syllables or words per second. Loudness is an auditory impression of the amount of energy in sounds (Roach, 1994). In speech, intonation is represented as pitch, tempo as duration of phonemes, and loudness as the energy of a speech signal.

In speech analysis a dilemma always exists over the usages of short-term or long-term features. Emotional speech prosody has mostly been analyzed using long-term features rather than short term. Long-term features are usually determined over a time window at the sentence or paragraph level. Short-term features are determined over windows with a smaller duration. The window length for short-term features is on the order of the duration of words, syllables, or even smaller units. Researchers for this topic state that long-term features identify emotion better than short-term (Li and Zhao, 1998; Schuller *et al.*, 2003). Most researchers focus on the influence of emotions on fundamental frequency ($F0$) and duration (Banse and Scherer, 1996; Fairbanks and Pronovost, 1939; Williams and Stevens, 1972; Fairbanks and Hoaglin, 1941; Murray and Arnott, 1993; Noad *et al.*, 1997). Research into emotions in speech also includes work on the influence of emotion on speech quality (Gobl and Ni Chasaide, 2000; Kienast *et al.*, 1999). Many researchers have investigated simulated and authentic emotions in emotional

[a)]Electronic mail: kacic@uni-mb.si

speech in regard to basic emotions (anger, joy, fear, and sadness) (Ekman, 1972). They found that joy, anger, and fear have a higher mean $F0$ and amplitude, with greater $F0$ and amplitude variations than neutral speech. Sadness has an opposite influence on $F0$ and amplitude than emotions such as anger, joy, and fear. Sadness has low mean $F0$ and amplitude, and low variations of $F0$ and amplitude (Banse and Scherer, 1996; Fairbanks and Pronovost, 1939; Williams and Stevens, 1972; Noad *et al.*, 1997). Speech rate also varies among emotions. The highest speech rate is for neutral and angry speech, and the slowest for sad speech. Some prosody parameters of emotional speech differ from study to study, while others are consistent (Cowie *et al.*, 2001). Sad speech usually has low values of pitch, but some studies have shown pitch to be high in sad speech (Pereira, 2000). There is also disagreement on the durational aspects of anger, happiness, and fear. Some report longer duration, some report faster speech rate, and some report lower speech rate (Cowie *et al.*, 2001; Davitz, 1964; Williams and Stevens, 1972; Coleman and Williams, 1979; Fonagy, 1978).

In the area of human-machine communication, emotions could be used for emotional speech synthesis and emotion recognition. Emotion recognition usually relies on statistical methods, which require large databases with many speakers. For recognition of emotional speech, a study of authentic emotions would be more appropriate than using simulated emotions. It may well be sufficient, for the purpose of synthesis, to study a single speaker so that his or her manner of emotional expression could be modeled (Douglas-Cowie *et al.*, 2000). The aim there would be to modify the output of speech synthesis systems in order to produce speech that sounds naturally emotional. This is essentially the same goal as that of actors (Johnstone, 1996).

Research into cross-cultural emotional expression shows both universality, and differences, in emotional signs between different cultures. Cross-cultural research of emotional expression in speech has tended to be of less interest than emotions by facial expression. This is due to the methodological problems of data collection, and vocal parameters being more difficult to measure than facial expressions (Frick, 1985). Facial signs of emotional expression have been researched by, among others, Ekman (1972), Izard (1971), and Bruner and Tagiuri (1954). Research into emotional expression in speech between different cultures has been carried out by Wallbott and Scherer (1995), Frijda (1986), Ohala (1984), and Amir and co-workers (2004). Research into emotion recognition for the English and Japanese languages has been done by Tickle (1999, 2000), when it was found that people recognize emotions in other languages, but not as accurately as in their own. Researchers who investigated emotional expressions of the face found that there are some universal signs of emotional expression, but there are also differences in different languages (Ekman, 1994).

Thus, there is an increasing need for emotion modeling. Various software agents need emotion models to communicate with humans in the most natural ways. These software agents need models for emotion recognition using the face, body movements, or speech, and for synthesis of emotions in the animation of a software agent's face and body movements, and in synthesized speech (Camurri and Coglio, 1998). Almost all researchers model emotion for speech synthesis using rules. A review of emotional speech synthesis is presented by Schröder (2001), who gathered and defined rules for specific emotions, which are combined from the works of Burkhardt (2000), Cahn (1990), and Murray (1989). Murray and co-workers (1989), who designed the HAMLET system, and Cahn (1990), who designed the AFFECT EDITOR system, also used rules for emotion modeling. These models are used for emotional speech synthesis using formant synthesizers. The disadvantage of a formant synthesizer is the low naturalness of synthesized speech. Diphone and corpus-based synthesizers reach a higher naturalness of synthesized speech, but have lower control over emotion parameters (Schröder, 2001). Modification of the prosody parameters in synthesized speech from diphone or corpus-based synthesizers can be done, but modification of speech quality parameters is more difficult. Researchers like Heuft and coworkers (1996), Rank and Pirker (1998), Murray (1989), and Montero and co-workers (1998) carried out mainly diphone or corpus-based emotional speech synthesis. All these authors used models with rules.

The aim of this current research is to investigate several questions regarding the modeling of prosody for emotional speech synthesis. The first question is: How many speakers should be used for modeling emotional speech prosody? As already mentioned, Douglas-Cowie and co-workers (2000) proposed the use of a single speaker so that his or her manner of emotional expression can be modeled. If a group of speakers was to have similar emotional speech characteristics, the prosody model could have some universal features. This kind of prosody model, therefore, could possibly be used for an emotional speech synthesizer having different voices with less, or no, adaptation to emotion expression for a particular voice. One of the major tasks facing automatic emotion detection is the automatic recovery of relevant features (Cowie *et al.*, 2001). As speech from several speakers should be used for the recovery of relevant features (Kwon *et al.*, 2003; Ververidis *et al.*, 2004), it might be expected that all speakers may have the same set of features affected by emotions. Therefore, the next question is, which and how many features of emotional speech are common to different speakers? To answer this, relevant features should be extracted from each speaker, separately.

This paper presents a novel rule-based method for defining emotion-dependent speech prosody parameters and their values. As prosody was the main concern of this research, the characteristics of pitch, energy, and duration of speech segments were analyzed. A proposed rule-based method is described in detail and the analyzed results are given for nine speakers. The output of the proposed method is compared to the results of principal component analysis (PCA), which is used as a statistical method for extracting relevant features from emotional speech. Emotion-dependent features are compared between different speakers, and a ratio of emotional expression similarity is calculated for each pair of speakers.

II. RULE-BASED EMOTION-DEPENDENT FEATURE EXTRACTION METHOD

In the proposed rule-based emotion-dependent feature extraction method, the pitch contour ($F0$) of the speech signal, durations of phonemes and pauses, energy of the speech signal, derivative ($\Delta F0$) of speech signal pitch contour, and derivative of speech signal energy contour are defined as low-level features. These features represent intonation in the speech signal.

The high-level features used in the proposed method are composed of statistical measurements of prosody parameters such as pitch, duration, and energy. Different researchers have found many prosody high-level features that were correlated to emotions (Banse and Scherer, 1996; Noad and coworkers, 1997; Murray and Arnott, 1993). Based on reports from the literature, we defined a set of high-level features as an initial set of features. A similar set of high-level features was also used by Banse and Scherer (1996), Noad nad coworkers (1997), Murray and Arnott (1993), Lee and Narayanan (2003), Kwon and co-workers (2003), and Lee and coworkers (2004).

The proposed method selects those features, from a large set of high-level features, that determine specific emotions. At the beginning of the process, all high-level features are normalized. Equation (1) represents normalized high-level features to reduce intraspeaker variability.

$$xn_{i,j} = \frac{x_{i,j}}{\bar{x}_i}, \tag{1}$$

where $xn_{i,j}$ is the normalized value of the ith feature in the jth sentence, $x_{i,j}$ is the ith feature in the jth sentence, $\bar{x}_i$ is the mean value of the ith feature. The xn stands for the normalized value. Each high-level feature is normalized with a speaker's high-level feature mean value, according to Eq. (1).

Pell (2001) found that sentence length has an influence on some prosody measurements. Therefore, test recordings should be separated according to sentence length. In our case, the sentences used for evaluation of the method were divided into five groups. The first group contained isolated words, the second short sentences, the third middle-length sentences, the fourth long sentences, and the fifth group was composed of context-dependent sentences.

Five classes of emotional feature values were used, equivalent to Banse and Scherer's (1996) classification of emotional feature values. The histogram of all the high-level feature values ranges from 0 to 2. As an illustrative example a histogram was determined for the first English male speaker (see Fig. 1), from the database used in the experiments. The mean value of all the normalized features of the first English male speaker was 1, and the standard deviation 0.094. The range of normalized high-level feature values was divided into five classes. Table I shows the output value and corresponding range of each class. Values 1.2 and 0.8 represented a deviation from mean value that was equal to 2 $\times$standard deviation. The values 1.05 and 0.95 were deviations from the mean value for half of the standard deviation. Two classes were above the value 1 and two classes were below the value 1 (see Table I). Class 1 represented values that were well below mean, class 2 values below mean, class 3 values around mean, class 4 values above mean, and class 5 values that were well-above mean.

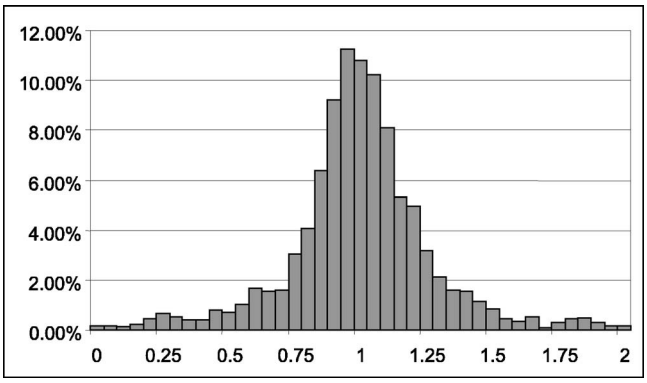


FIG. 1. Histogram of normalized high-level features for English male speaker (M1) for all values of high-level features.

The mean values of each group's normalized high-level features and for each session ($\bar{X}_{i,e,r,s}$) were classified according to Table I, and the result was denoted as $C_{i,e,r,s}$. $C_{i,e,r,s}$ was the value of the ith normalized high-level feature of the eth speaking style, of the rth group, and of the sth session of recordings (more recordings of the same sentences were made for the same emotion over different time instants—e.g., to enable emotion consistency evaluation). For example, $C_{i,e,r,s}$ could be the vowel duration value for anger, for a group of long sentences, session E1, and is denoted by $C_{\text{vowel duration,anger,long sentences,E1}}$.

The following rules were defined for extracting emotion-dependent features:

(1) Rule 1. If $C_{io,eo,ro,E1}$ and $C_{io,eo,ro,E2}$ have different values, then feature xn_{io} is denoted as an emotion-independent feature for emotion e_0. For example, if the vowel duration value for anger, group of long sentences, session E1 is different from the vowel duration value for anger, group of long sentences, session E2, then the vowel duration is denoted as an emotion-independent feature for anger.
(2) Rule 2. If $C_{io,eo}$ for all groups of the same emotion has different values, then the feature xn_{io} is denoted as an emotion-independent feature for emotion e_0. For example, if the vowel duration values for anger, for groups from isolated words to context-dependent sen-

TABLE I. Class number, range of high-level feature values that belong to the class, output value for high-level feature, and a brief description of which high-level feature values belong to which class.

Class num.	Range	Output value	Description
1	$xn_{i,j} < 0.8$	−2	Values are well below mean.
2	$0.8 \leqslant xn_{i,j} < 0.95$	−1	Values are below mean.
3	$0.95 \leqslant xn_{i,j} < 1.05$	0	Values are around mean.
4	$1.05 \leqslant xn_{i,j} < 1.2$	1	Values are above mean.
5	$1.2 \leqslant xn_{i,j}$	2	Values are well above mean.

$xn_{i,j}$-normalized high-level feature of ith feature in the jth sentence

tences, do not have the same values, then the vowel duration is denoted as an emotion-independent feature for anger. These two rules are applied for each feature and for each emotion. Features that are not denoted as emotion-independent features are subject to one additional rule.

(3) Rule 3. If C_{io} has the same values in all speaking styles, then the feature xn_{io} is also denoted as an emotion-independent feature for all emotions. For example, if the vowel duration mean value for all defined emotions is the same, then the vowel duration is defined as an emotion-independent feature for all emotions.

If feature xn_{io} is unclassified as an emotion-independent feature, then feature xn_{io} is denoted as an emotion-dependent feature. Emotion-dependent features should discriminate between speaking style according to emotion and not to sentence type or recording times. The first rule assures that emotion-dependent features are consistent with recording times. The second rule assures that emotion-dependent features are consistent with sentence type, and the last rule assures that emotion-dependent features discriminate between different emotions.

A. Definition of a feature vector

The rule-based method we use relies on high-level input features. In our case, high-level means statistical representations of low-level features. The low-level features were calculated from speech and determined prosody parameters.

1. Calculation of low-level features

Low-level features represent the acoustical characteristics of speech prosody. The pitch contour ($F0$) of the speech signal, duration of phonemes and pauses, energy of speech signal, derivative ($\Delta F0$) of speech signal pitch contour, and derivative of speech signal energy contour were defined as low-level features. These features represent intonation in the speech signal.

Various methods could be used for the calculation of pitch, energy, and duration. Methodology using a normalized cross-correlation function (Qian and Kimaresan, 1996) was used to extract the $F0$ from speech signals. This method calculated one pitch value per 10 ms. The rms (root-mean-square) method calculated energy from speech signals. Rms calculated energy in a square window that was 10 ms long at a frame rate of 5 ms. Phoneme segmentation was carried out using the HTK toolkit (Young *et al.*, 1997). Here, the speech signal was first pre-emphasized with a factor of 0.95. Twelve mel-cepstral coefficients and energy were used as a feature vector, as well as the first and second derivatives of the mel-cepstral coefficients. The cepstral coefficients were calculated using a Hamming window (5 ms long) at a frame rate of 2.5 ms. Normalization of cepstral mean value was used before acoustic modeling, where 3-state left-right hidden Markov models (HMMs) were applied. The emission probabilities were modeled using continuous Gaussian mixture densities. The models for all phonemes, for pauses between words, and the start and end points of sentences were used in the acoustic modeling.

The training database consisted of all the sentences of one speaker (speech recorded in all sessions and for all emotional states). Global mean values and variances were estimated on initialization. HMM parameters were re-estimated with six iterations of the Baum-Welch algorithm. In the next stage, the number of Gaussian mixture densities was increased in the following order: 1, 2, 4, 6, 8. Two iterations of the HMM parameter estimations were performed after each increase in the number of Gaussian mixture densities. Segmentation was done on all the sentences of one speaker, over all the emotional states.

2. Definition of high-level features

All of our high-level features were sorted according to their particular low-level features. The first group of high-level features was derived from $F0$ and $F0$ derivative ($\Delta F0$), and consisted of the mean value, standard deviation, minimal value, maximal value, and range of $F0$. The high-level features derived from $\Delta F0$ were the mean value, standard deviation, minimal value, maximal value, and range of $\Delta F0$. The range represented the difference between minimal and maximal values.

The second group of high-level features encompassed features that were calculated from the energy contour and its derivative: mean value, standard deviation, word and syllable energy, a derivative's mean value, and a derivative's standard deviation. Syllables were defined from phonetic transcriptions.

The third group consisted of features that were extracted from phoneme duration: words, syllables, vowels, fricatives, affricates, sonorants, plosives, and pause duration.

High-level features represent information about the speaker's prosody over an entire sentence.

III. EXPERIMENT

The analysis of emotional speech requires various methods from acoustics, speech science, and statistics. Emotional speech prosody with long-term features was used for the experiments described in this paper. The window used for the long-term features corresponded to the duration of sentence length, which was also the length of the recording. To evaluate the proposed rule-based method, we decided to use emotional speech from the Interface[1] databases (Hozjan *et al.*, 2002) in order to analyze the prosody patterns of emotional speech. The Interface databases contain acted emotional speech from speakers of different nationalities and genders.

All the high-level features, mentioned in Sec. II, represented input data for the emotion-dependent feature extraction method. This method was used on the speech of each speaker independently. It selected 14 emotion-dependent features from the defined 26 high-level features. These 14 features represented a union of these features from all nine speakers. The results for each speaker are presented in detail in Sec. IV.

The complete analysis was done automatically in three stages on an MS Windows-based PC machine. The low-level

and high-level features were calculated in the first and second stages, and the emotion-dependent features were determined in the third. Each stage used different software. The software for the calculation of pitch, energy, and derivatives was developed using the algorithms described in Sec. II A 1. The calculation of high-level and emotion-dependent features was done with Perl scripts and further analysis used MS EXCEL.

A. Principal component analysis (PCA)

Statistical methods, such as principal component analysis (PCA), which analyzes the significance of individual feature vector components, have been used in previous analyses of emotional features (Li and Zhao, 1998). In these analyses the speech of all speakers involved was used for the extraction of emotional features, and no speaker-dependent features were analyzed.

The results of the rule-based method proposed in this paper were compared with the results of statistical methods found in the literature, as well as with the results of PCA analysis.

PCA was used to verify the significance of the defined 26 high-level features, from the standpoint of discriminative significance. PCA examines the variance structure of the data and finds those directions in the feature space that have significant variations (Everitt and Dunn, 1992). PCA was also used in this study to analyze the importance of individual components in the feature vector. Altogether 14 emotion-dependent features were extracted for all speakers in the databases, using the rule-based method described in Sec. II. PCA was used in such a way as to reduce the 26-dimensional feature space to 14.

Linear correlations between individual components of the vector y can be transformed using the following expression:

$$y = A^T \cdot x, \tag{2}$$

where y represents a new feature vector, x an input feature vector, and A the transformation matrix. PCA analysis begins with the following expression:

$$C \cdot A = A \cdot \lambda, \tag{3}$$

where C is the covariance matrix, A the transformation matrix, and λ the diagonal matrix with the diagonal elements being the eigenvalues. If Eq. (3) is solved, then the values for the individual elements of the diagonal matrix λ and all the values of the transformation matrix A become results. The individual values of the diagonal matrix λ correspond to the individual feature variances. In the next step of the PCA analysis, the elements of the diagonal matrix λ and corresponding columns of the transformation matrix A were sorted according to the values of the diagonal matrix λ. The number of rows in the transformation matrix A, introduced in Eq. (3), was reduced to the size of the new vector y. At the end of the PCA analysis, the new vector y was calculated using Eq. (2). The transformation matrix A is often called the eigenvector matrix. A more detailed explanation of the PCA analysis is described in Everitt and Dunn (1992).

TABLE II. Norm of 14 dimensional transformation vectors calculated from the reduced transposed transformation matrix A and corresponding features sorted by significance.

	Features	Norm of transform vectors
1	Mean value of Δenergy	0.999
2	Duration of affricates	0.979
3	Duration of plosives	0.907
4	Duration of sonorants	0.905
5	Duration of fricatives	0.866
6	Standard deviation of energy	0.864
7	Duration of syllables	0.853
8	Mean value of $\Delta F0$	0.816
9	Duration of vowels	0.778
10	Minimal value of $F0$	0.719
11	Mean value of $F0$	0.684
12	Duration of phonemes	0.663
13	Maximal value of $F0$	0.656
14	Syllable energy	0.626
15	Maximal value of $\Delta F0$	0.619
16	Minimal value of $\Delta F0$	0.607
17	Duration of words	0.595
18	$F0$ Range	0.589
19	Standard deviation of $F0$	0.586
20	Duration of pauses	0.579
21	Standard deviation of $\Delta F0$	0.552
22	$\Delta F0$ Range	0.544
23	Word energy	0.513
24	Energy of high frequencies	0.508
25	Standard deviation of Δenergy	0.452
26	Mean value of energy	0.433

Instead of the last step of the PCA analysis, the norm of transformation vectors should be calculated in order to determine the significance of feature variance. The norm of transformation vectors is calculated from the reduced transposed transformation matrix A. In the reduced transformation matrix A, the values in the rows represent the amount of high-level feature contribution to the new feature vector y. The transformation vector z_i from the reduced transposed transformation matrix A is extracted for each high-level feature i. The norm of transformation vector z_i is calculated for each high-level feature i. The norm of transformation vector z_i represents the portion by which the high-level feature i contributes to the new feature vector y. The norm of transformation vector z_i represents the significance of the ith feature variability.

The norms of 14 dimensional transformation vectors z_i for all 26 high-level features are given in Table II. The high-level features were sorted in order of significance. According to PCA, the first feature (Δenergy mean value) is the most important and the 26th feature (energy mean value) is the least important.

Eight of the 14 emotion-dependent features from the rule-based analysis lie in the first half of Table II. Some emotion-independent features, such as the mean value of Δenergy or affricate durations, were placed in higher positions of importance according to PCA. PCA handles features by considering variations in the features' values. In contrast, our rule-based method additionally considers important con-

ditions such as recording times, sentence type, and discrimination between emotions. Recording time is determined by the recording session, and the sentence type ranges from isolated words to context-dependent sentences. Nevertheless, PCA analysis confirms that the majority of emotion-dependent features has significant variations and that these features are suitable for further analysis. Six such features are listed in the second part of Table II and their variances have less discriminative power. The rules defined in Sec. II assure that all emotion-dependent features discriminate between emotions, and therefore all the emotion-dependent features were used in further analysis.

B. Databases

All speakers from the Interface databases are used in our analysis. The Interface databases have nine speakers, who belong to four nationalities: English, Slovenian, French, and Spanish. The recordings are made by actors. Each database contains recordings of actors for one language. The English database uses one female and two male speakers. The Slovenian, French, and Spanish databases contain one male and one female speaker each.

1. Corpus of Interface databases

The corpora for the English, Slovenian, French, and Spanish Interface databases are designed mainly for emotional speech synthesis. Therefore, the corpora of the databases contain all phonemes in each language. The English corpus contains 186, Slovenian 190, French 175, and Spanish 184 sentences, respectively. All corpora are developed within the same specifications. In all languages the number of sentences is higher than that required by the specifications, in order to increase the frequency of rare phonemes in the corpora. The structure of the corpora is the same for the English, Spanish, Slovenian, and French Interface databases. The minimal number of isolated words, short, medium, and long sentences, and sentences that are context-dependent is defined in the specification of the Interface corpora. Short sentences are composed of five to eight words, medium sentences of nine to 13 words, and long sentences of 14 to 18 words.

2. Speech material

The Interface databases were recorded by the various research teams that were partners in the Interface project. The Slovenian and English Interface databases were recorded at the University of Maribor, Slovenia, the Spanish at the Universitat Politecnica de Catalunya, Spain, and the French at Scansoft, France. These corpora were recorded for six emotions and various numbers of neutral speaking styles. The emotions used in the databases are those defined by Cornelius (1996) as the "big six." Different types of neutral styles are recorded, as the Interface databases may also be used for nonemotional speech synthesis. The number of neutral speaking styles was not defined in the specifications, so each partner decided individually the number and types of neutral speaking styles. The English and Slovenian databases include two neutral styles: slow/soft and fast/loud neutral styles. The French database includes basic neutral, slow/soft, and fast/loud speaking styles. The Spanish database includes five different neutral styles: basic, slow, soft, fast, and loud.

This paper tries to find those prosody parameters that are influenced by emotions. Therefore, a minimal number of neutral speaking styles was chosen in the analysis. The basic neutral speaking style was selected from the Spanish and French databases. The English and Slovenian databases do not include the basic neutral speaking style, so both slow/soft and fast/loud neutral styles were used in the analysis described in this paper.

Databases consist of two sessions (E1 and E2). Each session contains recordings of the entire corpus for all six emotions and all defined neutral speaking styles. Session E2 was recorded at least 14 days after session E1, according to specifications. The aim of this was to ensure emotional speech parameter consistency. Each speaker uttered at least 2.000 recorded sentences. In the analysis 23.520 sentences were used altogether.

The Slovenian and Spanish Interface databases were evaluated by a subjective listening test to evaluate emotion expression in speech. In the Spanish Interface database emotions were recognized, on average, with an accuracy of 80% and in the Slovenian Interface database the accuracy of emotion recognition was 58%. In the listening test the intensity of emotions was evaluated and not the naturalness of the emotions. Details of the subjective evaluation test are described in Hozjan *et al.* (2002).

3. Recording conditions

The recordings of the Interface databases were made in studio environments using a high-quality studio microphone. The sampling frequency was 48 kHz with a sample size of 16 bits. Feature calculation used recordings that were downsampled to a sampling rate of 16 kHz.

IV. RESULTS AND DISCUSSION

Table III presents the emotion-dependent features for all speakers involved in the analysis. This table presents the number of these features and their values for each speaker and for each defined emotion. The missing values in Table III are a consequence of the rules described in Sec. II. The emotion-independent features for specific emotions are noted by the missing values. The shaded columns in Table III represent the emotion-independent features for each specific speaker. The differences between the sets of emotion-dependent features are outlined using shaded columns. The values of emotion-dependent features in Table III range from −2 to 2, according to the classes of output values of emotion-dependent features given in Table I. The emotion-dependent features were determined for each speaker individually. Each speaker, except Spanish ones, had a different set of emotion-dependent features. They had, altogether, 14 emotion-dependent features. These features were: mean value of $F0$ (mean $F0$), standard deviation of $F0$ (SD $F0$), minimal value of $F0$ (min $F0$), maximal value of $F0$ (max $F0$), range of $F0$ (range $F0$), standard deviation of $\Delta F0$ (SD $\Delta F0$), minimal value of $\Delta F0$ (min $\Delta F0$), maximal value of $\Delta F0$ (max $\Delta F0$),

TABLE III. Values of emotion-dependent features for all nine speakers in the Interface databases for each emotion. The emotion-dependent features in the table range from −2 to 2, according to the classes of output values described in Table I.

English male 1	F0 features					ΔF0 features				Duration features				
	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	1			1		-1	1				-1	0	-1	-2
Disgust	-1			-1				-1	-1		0	1	0	0
Fear	-1			-1		-1	-1	1	0		-2	0	-1	1
Neutral slow	0			1			1	1	1		0	-1	-1	-1
Joy	1			1		0	1				-1	-1	-2	-1
Neutral fast	-1			0			0	1	1		0			1
Surprise	1			1		1	1		1		-1	-1	-2	-1
Sadness	-1			-1		-1	-1	-1	-1		-2	1	1	1
English male 2	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	1		1	2		2	2			-2	-1	-1	-1	-1
Disgust	-2	-1	-1	-1		-2				1	0	2	1	2
Fear	1	-2	2	1			0			-1	-2	-1	-1	0
Neutral slow	0	1	-1	0		0				-2	-2	-2	-2	-2
Joy	1		0	1		2	2			-1	-1	1	0	0
Neutral fast	-1	-1	-2	0		-2	-1			-1	-1	0	-1	-1
Surprise	1	2	0	1		2	2			-2	-2	-1	-2	-1
Sadness	-2	-2	-1	-1		-2				-1	-1	1		
English female	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	0	0		0			0	-1	0			-1	-1	-2
Disgust	-1	-2		-1								1	0	0
Fear	0	-1		0		-1	0					2	1	2
Neutral slow	0			0		-1	-1		0				-1	-1
Joy	0	1		1		1	1	1	1			-1	-1	-1
Neutral fast	0	-1		-1		-1		0	-1			1	1	1
Surprise	1			1		2	1	1	1			-1	-1	-1
Sadness	0	-2		0		-2		-1				1	1	1
Slovenian male	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	2			1		2	2			-2	-1	-1	-1	-1
Disgust	0			0		-2				1	0	2	1	2
Fear	1		-2	0			0			-1	-2	-1	-1	0
Neutral slow	-1		-2	-1		0				-2	-2	-2	-2	-2
Joy	0			0		2	2			-1	-1	1	0	0
Neutral fast	-1		-1	-1		-2	-1			-1	-1	0	-1	-1
Surprise	1		1	1		2	2			-2	-2	-1	-2	-1
Sadness	-1		1	-1		-2				-1	-1	1		
Slovenian female	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	1			0		2	2			-2	-1	-1	-1	-1
Disgust	-1			-1		-2				1	0	2	1	2
Fear	1			0			0			-1	-2	-1	-1	0
Neutral slow	0			1		0				-2	-2	-2	-2	-2
Joy	1			0		2	2			-1	-1	1	0	0
Neutral fast	2			-1		-2	-1			-1	-1	0	-1	-1
Surprise	0			0		2	2			-2	-2	-1	-2	-1
Sadness	-1			-1		-2				-1	-1	1		
French male	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	1	0					1	1	1	-2	-2	-1	-1	-1
Disgust	1	1			1		1	1	1	0	-1	0	1	1
Fear	1	0			1					-2	-2	-1	-1	-1
Neutral	1	1			1					-1	-1	-1	-1	-1
Joy	1	0			0		2	1	2					
Surprise	0	1			1		1	2	1	-1	-1	-1	-1	-1
Sadness	0	0			1					-1	-1	1	0	2
French female	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	1	0		0	0					-2	-2	-2	-2	-2
Disgust	2	1		0	0					0	0		1	1
Fear	2	0		0	0					-2	-2	-1	-1	-1
Neutral	2	0		0	0					-2	-2	-2	-2	-1
Joy	1	0		1	1									
Surprise	2	1		0	0					-2	-2	-1	-1	-1
Sadness	-2			1	1					-2	-2	-2	-2	-1
Spanish male	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	1	0		-1						-2	-2	-2	-1	-2
Disgust	1	0		1						0	-1	1		1
Fear	0	-1		0						-2	-2	-2	-2	-2
Neutral	1	0		0						-2	-1	-2	-2	-1
Joy	0	0		0						-1	-2	-1	-1	-1
Surprise	1	1		-1						-1	-1	-1	0	0
Sadness	0	0		1						-1	-2	-1	1	1
Spanish female	Mean F0	SD F0	Min F0	Max F0	Range F0	SD ΔF0	Min ΔF0	Max ΔF0	Range ΔF0	phoneme	vowel	fricative	sonorant	plosive
Anger	1	0		-1						-2	-2	-1	-2	-2
Disgust	1	0		1						1	1	1	2	1
Fear	0	-1		0						-1	-1	-1	-1	0
Neutral	1	0		0						-1	-1	-1	-1	-1
Joy	0	0		0						-1	-1	-1	-1	-1
Surprise	1	1		-1						-1	-1	-1	-1	-1
Sadness	0	0		1						1	-1	1	2	2

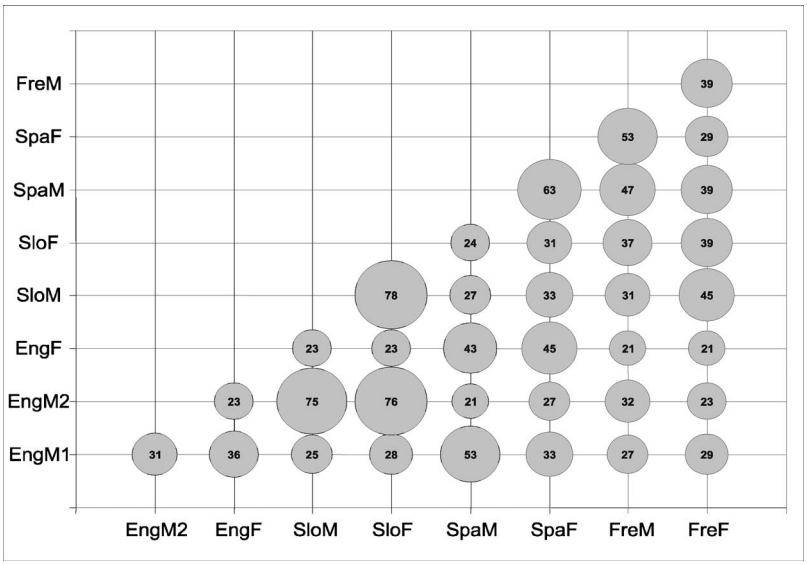


FIG. 2. The area of a bubble is equal to the ratio of similarity between speakers defined in rows, and speakers defined in columns. The values in the centers of the bubbles represent the ratio of similarity expressed as percentages.

range of $\Delta F0$ (range $\Delta F0$), duration of phonemes (phoneme), duration of vowels (vowel), duration of fricatives (fricative), duration of sonorants (sonorant), and duration of plosives (plosive). The names in the brackets are used in Table III.

The results of the analyses in this study are similar to the results of other authors such as Banse and Scherer (1996), and Noad and co-workers (1997). The difference between our study and others is in the interpretation of results. Other authors compare emotions to a neutral speaking style. Emotion-dependent feature values were extracted in this research according to the mean value defined from all emotions, including neutral speaking styles. This comparison was made to allow emotion-dependent feature comparison between different speakers. When the emotion-dependent feature values in our research are compared to neutral speech, the results are similar to the results of other authors. For emotions with high arousal (anger, joy, and surprise), the high values of $F0$ features and the low duration feature values can be compared favorably with the neutral speaking style. Emotions with low arousal (disgust and sadness) have low values of $F0$ features and low duration feature values compared to neutral speech.

The values for emotion-dependent features for the same emotion differed for all speakers. Common emotion-dependent features were defined to enable a comparison of emotion expression with the extracted emotion-dependent features. They were defined for two speakers, and were those features that were present for both. For example, the set of common emotion-dependent features for the Spanish male and the Spanish female was composed of all emotion-dependent features (see Table III) for both, because these two had the same set of emotion-dependent features. Furthermore, the English female and the French female from the database had the following set of common emotion-dependent features (see Table III): mean value of $F0$, standard deviation of $F0$, maximal value of $F0$, duration of fricatives, sonorants, and plosives. Each pair of speakers may have had different sets of common emotion-dependent features. The ratio of emotional expression similarity was defined using common emotion-dependent features. The number of common emotion-dependent feature values that are equal for the same emotion was estimated, and is defined as the number of features with the same value. The ratio of emotional expression similarity is defined as the ratio between the number of common emotion-dependent features with the same value and the number of all common emotion-dependent features of two speakers being compared. In Fig. 2 the ratio of similarity is expressed as a percentage and is depicted as a bubble for each pair of speakers. The area of the bubble is equal to the ratio of similarity.

Two neutral styles were used for analysis of the Slovenian and English speakers, and one neutral style was used for the French and Spanish speakers. A basic neutral speaking style was used for the French and Spanish speakers. When the ratio of similarity was calculated between the English or Slovenian and French or Spanish speakers, then the fast neutral speaking style was excluded from the comparison. Our method for calculating the ratio of similarity requires an equal number of emotions for both speakers. In our experience from the Interface databases, the slow neutral speaking styles of the English and Slovenian speakers have a greater similarity with the Spanish and French neutral speaking styles than the fast neutral speaking style.

In Fig. 2, two groups of speakers had a high ratio of emotional expression similarity. The first group consisted of the Slovenian male, the Slovenian female, and the second English male (in Fig. 2 the second English male speaker is marked as EngM2). The second group of speakers encompassed the Spanish male, the Spanish female, and the French male. The members of both groups had a high ratio of emotional expression similarity values among each other. The first English male (EngM1), the English female, and the French female speakers were not included in these two groups, but they had a high ratio of similarity values with several speakers of the Interface databases. The first English male and English female had a high ratio of emotional expression similarity with the Spanish male and the Spanish

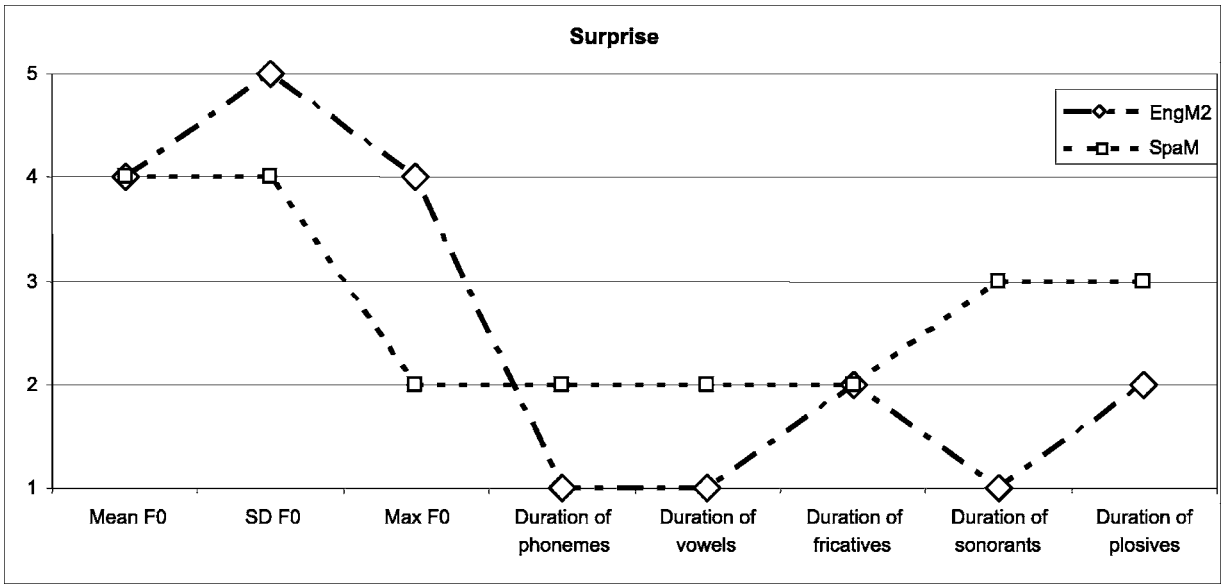


FIG. 3. Common emotion-dependent features for the second English male (EngM2) and Spanish male (SpaF) speakers for surprise.

female. The French female had emotional expression similar to the Slovenian male, Spanish male, Slovenian female, and French male.

Figures 3 and 4 show common emotion-dependent features for speakers with the lowest and the highest ratios of emotional expression similarity, which can be seen from Fig. 3, as most of the common emotion-dependent features have different values. Figure 3 presents the common emotion-dependent features of the second English speaker (EngM2) and the Spanish speakers for emotion surprise. This emotion was selected as it exhibits a similar ratio of common emotion-dependent features and had the same values as the ratio of emotional expression similarity between the two compared speakers for all emotions. These two speakers had the lowest ratio of emotional expression similarity. Figure 4 presents the common emotion-dependent features of the Slovenian male and female speakers for the emotion anger. The Slovenian speakers had the highest ratio of similarity, as most of the common emotion-dependent features had the same values (see Fig. 4).

The experimental results show that our rule-based method defines a set of emotion-dependent features and their values when a speaker expresses emotion. The first two rules of our rule-based method select features for each emotion separately; therefore, the features are specified by emotion alone. Only the last rule considers features among all emotions.

The results of our rule-based method for emotion-dependent feature extraction demonstrate that the described method is capable of extracting such features from speakers of different gender and nationality. The ratio of emotional expression similarity shows that some similarities can be de-

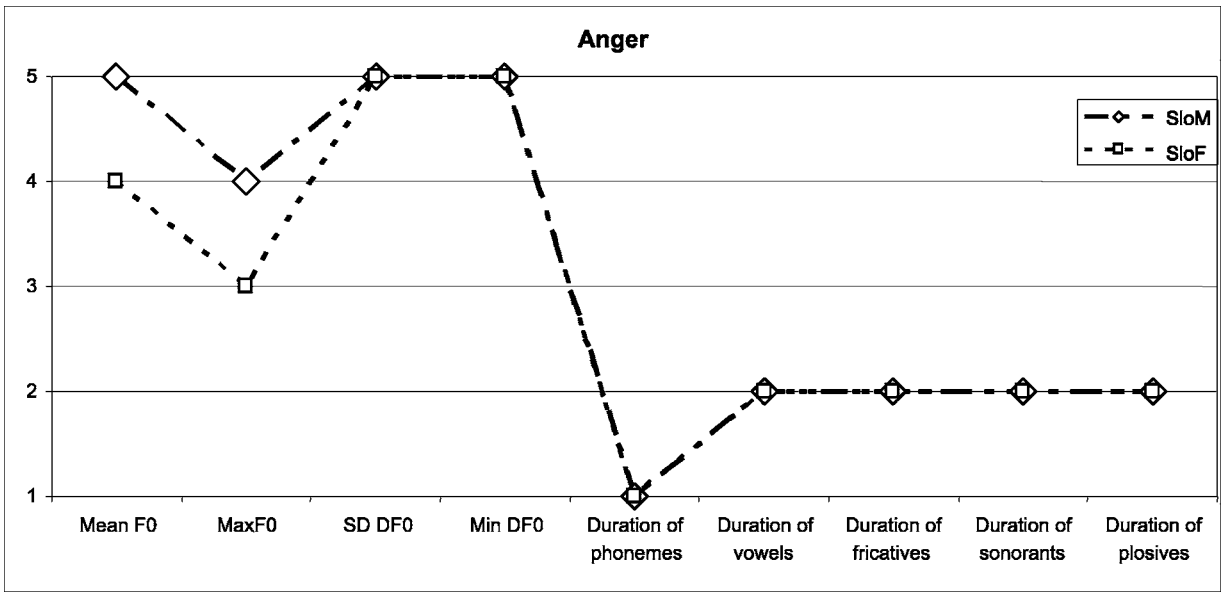


FIG. 4. Common emotion-dependent features for Slovenian male (SloM) and female (SloF) speakers for anger.

tected for all speakers. For example, all speakers had high values for almost all $F0$ features for anger (high values are values of 3 or greater), except for the standard deviation value of $\Delta F0$ for the first English male speaker, the maximal value of $\Delta F0$ for the English female speaker, and the maximal value of F0 for both Spanish speakers.

The results also show that the speakers from the Interface databases use different sets and values of emotion-dependent features to express their emotions in speech. Pereira (2000), and Cowie with co-workers (2001) show in their literature review that some features are remarkably consistent across all studies while others are quite variable. Similarly, it has been found in our analysis that there are some similarities of emotion-dependent feature values among all speakers, and the values of some emotional features are different for the same emotion.

Detailed analyses of the emotion-dependent features indicate that almost all speakers have unique sets of emotion-dependent features, except for the Spanish speakers. Fourteen emotion-dependent features were extracted altogether for all speakers. The significance of the emotion-dependent features was confirmed using PCA analysis. PCA was used as a statistical method for the extraction of features during emotion detection for all speakers involved in the analysis. Eight of the 14 emotion-dependent features were among a significant half of all 26 high-level features. These high-level features placed in the significant half are high-level features with significant variations according to PCA. The difference between PCA and the rule-based method is that PCA, as a statistical method, uses statistical measurement (variance) as a measure of significance. Our rule-based method uses, as a measure of importance, the sentence type, recording time, and emotions. Because both methods use different measures of significance, the PCA and the rule-based methods together confirm that the majority of emotion-dependent features is unaffected by recording times and sentence types, but is affected by emotions and have significant variations. Out of 14 emotion-dependent features, only five were present for all speakers. These five features could represent universal signs of emotional expression in speech for the speech databases used, and could serve as general emotion models for emotional speech synthesis. All speakers also had unique emotion-dependent feature values for all emotions, and each speaker, except for the Spanish speakers, had a unique set of emotion-dependent features. This indicates that each speaker in the databases expresses emotions differently. It can, therefore, be concluded that each speaker can perform each emotion in speech differently from other speakers using different sets of emotion-dependent features and/or values. Therefore, in speech synthesis, emotions should be modeled using the emotion-dependent parameters of only one speaker. If emotional expressions were to be modeled from more than one speaker, then the differences in emotional expression between different speakers should not be modeled, but only those characteristics of emotions that are common to all speakers. The best modeling for emotional speech synthesis, therefore, should be expected if only one speaker is selected for the modeling of emotions, as proposed by Douglas-Cowie and co-workers (2000). The extracted emotion-dependent features could possibly be used in emotional speech synthesis, as a small number of parameters controlling speech prosody are usually used in speech synthesis. The speakers from the Interface databases had from 8 to 11 emotion-dependent features. This number of features represents the amount of prosody parameters that could be controlled, at the same time, in emotional speech synthesizers. From these the $F0$ features and duration features are already used in emotional speech synthesizers, as can be found in the literature. According to the available literature, $\Delta F0$ features have not been used in emotional speech synthesis so far. As $\Delta F0$ features have proved to be emotion dependent, it would be reasonable to use them along with $F0$ and duration features to control prosodic parameters in emotional speech synthesis.

An analysis of authentic emotions should be made in order to verify our rule-based method for emotional speech in general. The results for authentic emotions are not expected to be the same as the results presented in this article, because it was shown in the literature (Batliner, 2003) that authentic emotions are less recognizable than acted emotions. In contrast, the results from our study could suggest that authentic emotion on a speaker level could be as recognizable as acted emotions, but the differences among different speakers who express authentic emotions are higher than the differences among speakers who express acted emotions. Usually this effect decreases the recognition rate.

V. CONCLUSION

This paper presents a rule-based method for the extraction of emotion-dependent features and an analysis of the experimental results. Statistical measurements of prosody parameters were used to define high-level features, from which the emotion-dependent features were selected using the extraction rules. Based on predefined rules, this method selected emotion-dependent features for each speaker individually, and independent of each other. The method thus enables comparison of emotion-dependent feature values between different speakers. Furthermore, the ratio of emotional expression similarity between two speakers was defined by calculating the number and values of the emotion-dependent features that are present for the two speakers being compared.

The proposed rule-based method was used to analyze emotional speech from four different languages, using male and female actors. In the analysis, nine speakers were used in order to find any similarities and differences among individual speakers, and to compare the results with those from other studies. The ratio of emotional expression similarity was calculated in order to compare the expressions of emotion between different speakers.

The results show that all the speakers showed some universal signs regarding certain emotion-dependent features of emotion expression. For example, all speakers had values of emotion-dependent features higher or equal to 0 for the emotions of anger and joy. However, further analysis showed that each emotion was expressed differently, which can be seen through the different sets of emotion-dependent features

and/or with different emotion-dependent features' values, when compared with results from all the speakers.

The comparison among speakers using the proposed emotion-dependent feature extraction method shows that the expressed emotions can be analyzed according to two criteria. The first criterion is a defined set of emotion-dependent features and the second is an emotion-dependent feature value. The results also show that almost all speakers from the Interface databases use a unique set of emotion-dependent features and that all the speakers had unique values of emotion-dependent features for each emotion.

The ratio of emotional expression similarity, however, indicates that speakers could be grouped together according to the values of the emotion-dependent features they used for emotional expression. The ratio of similarity for all cases is higher than 20% for the speakers in the Interface databases.

Our conclusions can be drawn mainly for prosody modeling in emotional speech synthesis, as both the emotional speech synthesizer and the actors have the same goal in the production of emotional speech. The extracted emotion-dependent features could be particularly appropriate for modeling individual speakers' emotions in emotional speech synthesis, as a set of emotion-dependent features for a specific speaker contains a small number of features that could be used for controlling prosody parameters in an emotional speech synthesis system.

[1]"Multimodal Analysis/Synthesis Systems for Human Interaction to Virtual and Augmented Environments" (INTERFACE), IST-1999-10036.

Amir, N., Ziv, S., and Cohen, R., (**2003**). "Characteristics of authentic anger in Hebrew speech," in *Proceedings of Eurospeech '03*, Geneva, pp. 713–716.

Amir, N., Almogi, B. C., and Gal, R. (**2004**). "Perceiving Prominence and Emotion in Speech—a Cross Lingual Study," in *Proceedings of Speech Prosody '04*, Japan, pp. 375–378.

Ang, J., Dhillon, R., Krupski, A., Shilberg, E., and Stockle, A. (**2002**). "Prosody-based automatic detection of annoyance and frustration in human-computer dialog," in *Proceedings of ICLSP '02*, Denver, Colorado, pp. 2037–2040.

Arnfield, S. C. (**1994**). "Prosody and Syntax in Corpus Based Analysis of Spoken English," Ph.D. dissertation, The University of Leeds School of Computer Studies.

Auberge, V., Audibert, N., and Rilliard, A. (**2004**). "E-Wiz: A Trapper Protocol for Hunting the Expressive Speech Corpora in Lab," in *Proceedings of LREC'04*, Lisbon, pp. 179–182.

Banse, R., and Scherer, K. R. (**1996**). "Acoustic profiles in vocal emotion expression," J. Pers Soc. Psychol. **70/3**, 614–636.

Batliner, A., Fisher, K., Huber, R., Spilker, J., and Noeth, E. (**2000**). "Desperately seeking emotions or: Actors, wizards and human beings," in *Proceedings of ISCA-Workshop on Speech and Emotion*, Newcastle, Northern Ireland, pp. 195–200.

Batliner, A., Fisher, K., Huber, R., Spilker, J., and Noeth, E. (**2003**). "How to find trouble in communication," Speech Commun. **40**, 117–143.

Bruner, J., and Tagiuri, R. (**1954**). "The perception of people," in *Handbook of Social Psychology*, edited by G. Lindzey (Addison-Wesley, Cambridge, MA), Vol. **2**.

Burkhardt, F. (**2000**). "Simulation emotionaler Sprechweise mit Sprachsyntheseverfahren," Ph.D. thesis, Berlin.

Cahn, E. J. (**1990**). "Generating expression in synthesized speech," Master thesis, Massachusetts Institute of Technology.

Camurri, A., and Coglio, A. (**1998**). "An architecture for emotional agents," IEEE Multimedia, October - December, 24–33.

Coleman, R. and Williams, R. (**1979**). "Identification of emotional states using perceptual and acoustic analyses," in *Care of the Professional Voice*, edited by V. Lawrence and B. Weinberg (The Voice Foundation New York), Vol. **1**.

Cornelius, R. (**1996**). *The Science of Emotion* (Prentice-Hall, Englewood Cliffs, NJ).

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., and Taylor, J. (**2001**). "Emotion recognition in human–computer interaction," IEEE Signal Process. Mag. **18**(1), 32–80.

Davitz, J. R. (**1964**). *The Communication of Emotional Meaning* (McGraw-Hill, New York).

Douglas-Cowie, E., Cowie, R., and Schröder, M. (**2000**). "A new emotional speech database: Consideration, sources and scope," in *Proceedings of ISCA-Workshop on Speech and Emotion*, Newcastle, Northern Ireland, pp. 39–44.

Ekman, P. (**1972**). "Universal and cultural differences in facial expression of emotion" in *Nebraska Symposium on Motivation*, edited by J. R. Cole (Nebraska University Press, Lincoln, NE), pp. 207–283.

Ekman, P. (**1994**). "Strong evidence for universals in facial expressions: A reply to Russell's critique," Psychol. Bull. **115**, pp. 268–287.

Everitt, B. S., and Dunn, G. (**1992**). *Applied Multivariate Data Analysis* (Oxford University Press, New York.)

Fairbanks, G., and Hoaglin, L. W. (**1941**). "An experimental study of duration characteristics of the voice during the expression of emotion," Speech Monograms, **8**, pp. 85–90.

Fairbanks, G., and Pronovost, W. (**1939**). "An experimental study of the pitch characteristics of the voice during the expression of emotions," Speech Monographs **6**, 87–104.

Fonagy, I. (**1978**). "Emotions, voice and music," Lang Speech **21**, 34–49.

Frick, R. W. (**1985**). "Communicating emotion: The role of prosodic features," Psychol. Bull. **97**(3), 412–429.

Frijda, N. H. (**1986**). *The Emotions* (Cambridge University Press, Cambridge).

Gobl, C., and Ni Chasaide, A. (**2000**). "Testing affective correlates of voice quality through analysis and resynthesis," in *Proceedings of ISCA-Workshop on Speech and Emotion*, Newcastle, Northern Ireland, pp. 178–183.

Heuft, B., Portele, T., and Rauth, M. (**1996**). "Emotion in time domain synthesis," in *Proceedings ICSLP '96*, pp. 1974–1977.

Hozjan, V., Kačič, Z., Moreno, A., Bonafonte, A., and Nogueiras, A. (**2002**). "Interface Databases: Design and Collection of a Multilingual Emotional Speech Database," in *Proceedings of LREC '02*, pp. 2024–2028.

Izard, C. E. (**1971**). *The Face of Emotion* (Appleton-Century-Crofts East Norwalk, CT).

Johnstone, I. T. (**1996**). "Emotional speech elicited using computer games," in *Proceedings of ICSLP '96*, pp. 1985–1988.

Kienast, M., Paeschke, A., and Sendlmeier, W. (**1999**). "Articulatory reduction in emotional speech," in *Proceedings of Eurospeech '99*, pp. 117–120.

Kwon, O-W., Chan, K., Hao, J., and Lee, T-W. (**2003**). "Emotion recognition by speech signals," in *Proceedings Eurospeech '03*, Geneva, pp. 125–128.

Lee, C., and Narayanan, S. (**2003**). "Emotion recognition using a data-driven fuzzy inference system," in *Proceedings of Eurospeech '03*, Geneva, pp. 157–160.

Lee, C-M., Yildirim, S., Bulut, M., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., and Narayanan, S. (**2004**). "Emotion recognition based on phoneme classes," in *Proceedings of ICSLP '04*, Korea, pp. 889–892.

Li, Y., and Zhao, Y. (**1998**). "Recognizing Emotions in Speech Using Short-term and Long-term Features," in *Proceedings of ICSLP '98*.

Montero, J., Gutierrez-Arriola, J., Palazuelos, S., Enriquez, E., Aguilera, S., and Pardo, J. M. (**1998**). "Emotional Speech Synthesis: From Speech Database to TTS," in *Proceedings of ICLSP '98*, Sydney.

Murray, I. R. (**1989**). "Simulating emotional speech," Ph.D. thesis, University of Dundee, UK.

Murray, I. R., and Arnott, J. L. (**1993**). "Toward the simulation of the synthesized speech: A review of literature on human vocal emotion," J. Acoust. Soc. Am. **93**, 1097–1108.

Noad, J. E., Whiteside, S. P., and Green, P. D. (**1997**). "A macroscopic analysis of an emotional speech corpus," in *Proceedings of Eurospeech '97*, pp. 517–520.

Ohala, J. J. (**1984**). "An ethological perspective on common cross-language utilization of *F*0 of voice," Phonetica **41**, 1–16.

Pell, M. D. (**2001**). "Influence of emotion and focus location on prosody in matched statements and questions," J. Acoust. Soc. Am. **109**(4), 1668–1680.

Pereira, C. (**2000**). "Perception and expression of emotion in speech," Ph.D. thesis, Macquarie University, Australia.

Qian, X., and Kimaresan, R. (**1996**). "A Variable Frame Pitch Estimator and Test Results," in *Proceedings of IEEE ICASSP '96*, Vol. **1**, pp. 228–231.

Rank, E., and Pirker, H. (**1998**). "Generating emotional speech with concatenative synthesizer," in *Proceedings of ICSLP '98*, pp. 671–674.
Roach, P. (**1994**). "Conversion between prosodic transcription systems: 'Standard British' and TOBI," Speech Commun. **15**, 91–99.
Schröder, M. (**2001**). "Emotional Speech Synthesis: A Review," in *Proceedings of Eurospeech '01*, pp. 561–564.
Schuller, B., Rigoll, G., and Lang, M. (**2003**). "Hidden Markov Model-Based Speech Emotion Recognition," in *Proceedings of ICASSP '03*, pp. 401–405.
Tickle, A. (**1999**). "Cross-language vocalisation of emotion: Methodological issues," in *Proceedings of ICPhS '99*, pp. 305–308.
Tickle, A. (**2000**). "English and Japanese speakers' emotion vocalization and recognition: A comparison highlighting vowel quality," in *Proceedings of ISCA-Workshop on Speech and Emotion*, Newcastle, Northern Ireland, pp. 104–109.
Ververidis, D., Kontropoulos, C., and Pitas, I. (**2004**). "Automatic emotional speech classification," in *Proceedings of ICASSP'04*, pp. 593–596.
Wallbott, H. G., and Scherer, K. R. (**1995**). "Cultural determinants in experiencing shame and guilt," in *Self-conscious Emotions: The Psychology of Shame, Guilt, Embarrassment, and Pride*, edited by J. P. Tangney and K. W. Fischer (Guilford, New York), pp. 465–487.
Williams, C. E., and Stevens, K. N. (**1972**). "Emotions and speech: Some acoustical correlates," J. Acoust. Soc. Am. **52**, 1238–1250.
Young, S., Ollason, D., Valtchev, V., and Woodland, P. (**1997**). *The HTK Book (for HTK 2.1)* (Entropic Cambridge Research Laboratory, Cambridge).

Interaction of reed and resonator by sound generation in a reed organ pipe

András Miklós,[a] Judit Angster,[b] and Stephan Pitsch
Fraunhofer-Institut Bauphysik, Nobelstrasse 12, D-70569 Stuttgart, Germany

Thomas D. Rossing[c]
Physics Department, Northern Illinois University, DeKalb, Illinois 60115

(Received 31 August 2005; revised 24 February 2006; accepted 26 February 2006)

Interaction of reed and resonator in reed organ pipes without and with the resonator has been investigated. The wave form of the sound generated without the resonator attached is a periodic train (with the frequency of reed vibration) of decaying high frequency oscillation. The frequency of this oscillation is independent of the frequency of the reed vibration; it corresponds roughly to the fundamental acoustic resonance of the shallot. Another rather surprising phenomenon was also observed: without the pipe resonator, the reed vibration frequency, and correspondingly the sound frequency is a continuous function of the reed vibrating length, but when the resonator is added "forbidden" frequency domains occur in the vicinity of the eigenfrequencies of the resonator, and the sound frequency jumps abruptly from below to above each resonance frequency of the resonator, when the reed is shortened. This effect can be explained by the coupling between two oscillating systems, the reed and the acoustic resonator. The presented investigations result in a better understanding of the acoustical and mechanical aspects of sound generation in reed pipes, which may allow organ builders and voicers to easier pipe design and more accurate voicing to obtain the desired sound. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188372]

PACS number(s): 43.75.Np, 43.75.Pq, 43.25.Ts, 43.28.Ra [MFH] Pages: 3121–3129

I. INTRODUCTION

Although nearly all pipe organs include reed (lingual) pipes as well as flue (labial) pipes, much less research has been reported on reed pipes than on labial pipes.[1–3] In most reed pipes the reed vibrates against a fixed shallot and modulates the flow of air passing through the shallot into the resonator. In reed pipes there is a rather complicated interaction between the vibrating reed and the pipe resonator.[4–6]

A typical reed pipe is shown in Fig. 1. Air under pressure flows from the windchest through the bore into the boot and through the thin opening into the shallot. The wind sets the thin, flexible tongue (reed) into vibration; this in turn modulates the flow of air passing through the shallot into the resonator. The reed is pressed against the open side of the shallot by a tuning wire ("Stimmkrücke") that can be adjusted up and down to tune the vibrating reed. The pipe voicer generally adjusts the reed and the resonator to produce the best sound and then tunes the pipe by adjusting the tuning wire.[7,8]

In an earlier paper we reported on measurements of reed vibration in reed organ pipes without the resonators.[9] We reported three rather surprising phenomena: the frequency of the reed plucked by hand is shifted upwards for large-amplitude plucking; the blown frequency is significantly higher than the plucked one; and peaks halfway between the harmonics of the fundamental frequency appear in the spectrum of the reed velocity.

In this paper, we discuss reed vibrations with resonator and sound generation in the reed pipe with and without the resonator.

II. EXPERIMENTAL METHOD

The experimental apparatus is similar to that described in our earlier paper.[9] The pipes were tested on a slider chest, with the wind supplied by an organ blower and pressure regulator. Wind pressure was measured on a water manometer. Because the response of a reed pipe is very much dependent on the speed of the pressure onset, the pipe was switched on by means of a pallet activated by an electromagnet. We found that a sudden pressure onset was needed in our experiments for stable reed oscillations.

A window was installed on the boot so that the reed could be observed while it vibrated. The velocity was recorded by means of a laser vibrometer (Polytec OFV 3000). A $\frac{1}{4}$-inch microphone (B&K 4135) was inserted into the shallot wall to record the sound pressure inside the shallot, and another microphone (B&K 4165) recorded the sound pressure near the open end of the pipe.

The reed velocity and the sound pressure inside the shallot, as well as the sound pressure near the open end of the pipe were recorded as the vibrating reed length was varied by means of the tuning wire. The wind pressure was set to the optimal value (80 mm water gauge) for sounding the pipe, as recommended by an experienced organ voicer. Sound pressure and reed velocity wave form were recorded both with and without the pipe resonator in place. Sound spectra, velocity spectra, and frequency were obtained with the help of a dual-channel FFT analyzer (HP 35670A).

[a]Electronic mail: andreas.miklos@urz.uni-heidelberg.de
[b]Electronic mail: angster@ibp.fhg.de
[c]Electronic mail: rossing@physics.niu.edu

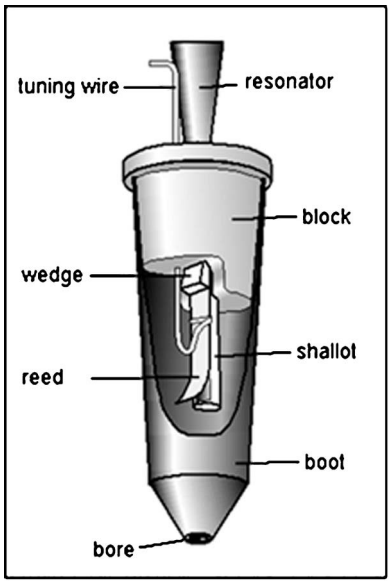


FIG. 1. The foot (boot) of a reed pipe.

The natural resonances (eigenfrequencies) of the pipe resonator were determined by measuring the frequency response of the resonator to an external acoustic excitation. Pink noise was radiated by a small loudspeaker placed close to the larger opening (mouth) of the conical resonator, and the sound pressure at the narrow end (throat) was measured by a small microphone.

A G4 Trompette pipe was used for the study presented in this paper. The conical resonator of the pipe was 687 mm long with diameters 9.2 and 65 mm at the ends (conical angle, 4.7 degree). The shallot was also slightly conical (internal diameters, 6.5 mm at upper end and 8 mm at the lower end), with a length of 56.5 mm and a wall thickness of 1.5 mm. The opening on the side of the shallot has a triangular shape with 4.8 mm maximal width and 40.4 mm height. The reed (tongue) was made from bronze. It has a trapezoidal shape with length of 55.8 mm and widths of 7.2 mm and 5.5 mm at the ends. The thickness of the reed is 0.312 mm. The 5.5 mm width end was clamped by means of a wedge into the block of the pipe. The total length of the reed from the wedge to the free end was 45 mm. The vibrating length could be adjusted from 37.5 mm to 15 mm.

III. RESULTS

A. Sound in the shallot without the resonator

In this case the resonator is not attached to the boot of the pipe. The system, as shown in Fig. 1, consists of the boot, the shallot inserted into the block and the reed, attached to the shallot. This system, when blown, produces quite a strong sound, whose frequency can be tuned continuously by the tuning wire. It has been shown[9] that the oscillation of the reed is generated and maintained by hydrodynamic forces due to the air flow through the gap between the shallot and reed; therefore, no acoustic feedback is necessary for the reed vibration. Sound is produced by the modulated volume flow into the shallot.

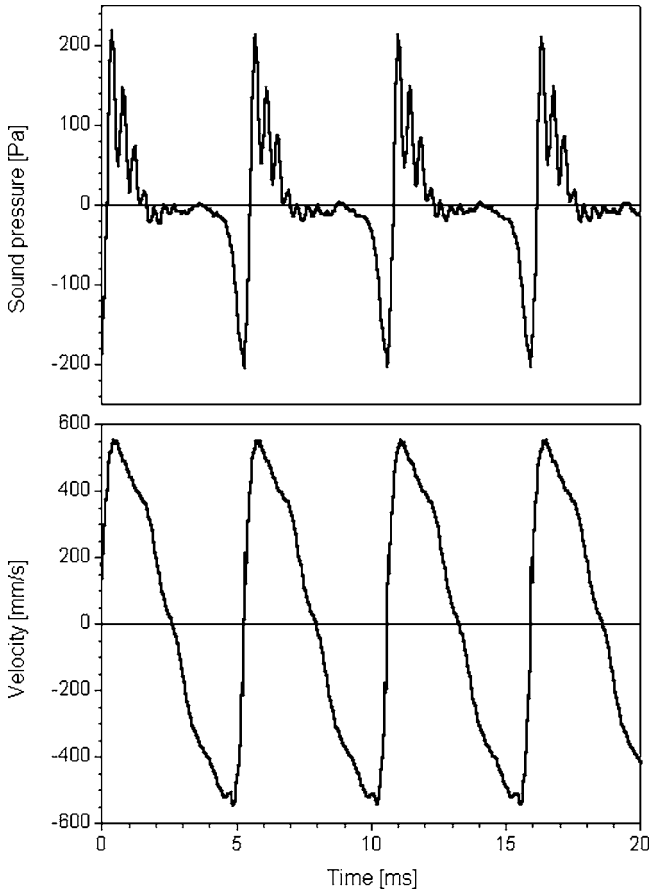


FIG. 2. Sound pressure wave form in the shallot (upper graph) and reed velocity wave form (lower graph) without the resonator attached for a 37.5 mm reed length. The sign of the velocity is negative when the reed moves toward the shallot.

The pressure wave form in the shallot was markedly different from that of the velocity wave form of the reed. Sound pressure and reed velocity are plotted in Fig. 2 for a 37.5-mm reed length. The sign of the velocity is negative when the reed moves towards the shallot. When the reed approaches the shallot, it slows down, turns around, and speeds up very rapidly. As the gap decreases, the pressure in the shallot drops and reaches a minimum value when the velocity is zero, i.e., at the lower turning point of the reed vibration, where the gap is the narrowest.

The sound pressure wave form includes two components: a strong bipolar pressure pulse and a high frequency oscillation. This oscillation begins when the reed moves away from the shallot, and as the gap increases, its frequency shifts upwards, diminishes in amplitude, and dies out. Sound pressure wave forms in the shallot are shown in Fig. 3 for three different reed lengths (31.5, 30.5, and 28.5 mm). The high frequency oscillations show a striking similarity for different reed lengths. Although the period of the sound signal decreases with the reed length, the high frequency oscillations are almost identical.

Spectra of a single period of the bandpass-filtered (1 kHz–8 kHz) sound pressure wave forms in the shallot are shown in Fig. 4 for 31.5, 30.5, and 28.5 mm reed lengths. As expected, broad maxima around 2500 Hz can be observed, corresponding to the ~0.4 ms period of the high frequency

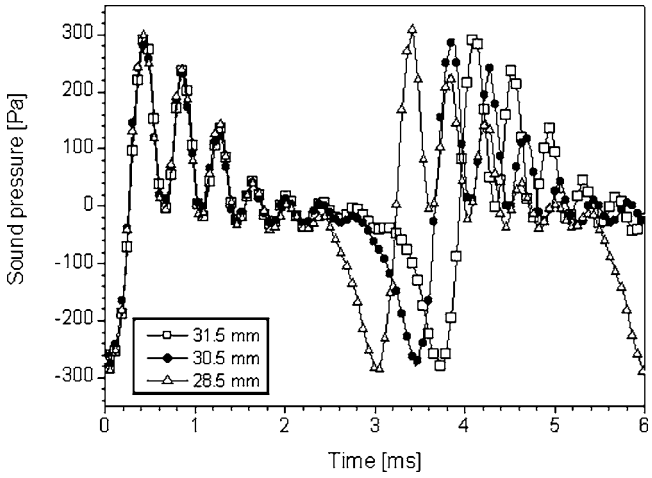


FIG. 3. Sound pressure wave forms in the shallot (without the resonator) for reed lengths of 31.5 mm, 30.5 mm, and 28.5 mm.

oscillations in Fig. 3. The peak amplitude increases with decreasing reed length. This behavior corresponds to the increasing rms vibration amplitude shown in Fig. 9.

The high-frequency oscillation in the shallot will be more and more concealed by the strong bipolar pulse as the vibration frequency increases. The number of the high frequency oscillation periods is gradually reduced as the vibration period decreases. Sound pressure wave forms in the shallot for smaller reed lengths are presented in Fig. 5. Only 2–3 periods of the high frequency oscillation can be recognized, but its frequency is still independent on the vibration frequency of the reed.

B. Reed vibration with resonator

The frequency of the vibrating reed of a blown G4 Trompette pipe with resonator attached was investigated for different reed lengths. The frequencies of the resonator modes were determined by placing a loudspeaker excited with pink noise about 10 cm from the open mouth of the resonator (tilted so as not to cause standing waves) and measuring the sound pressure with a small microphone about

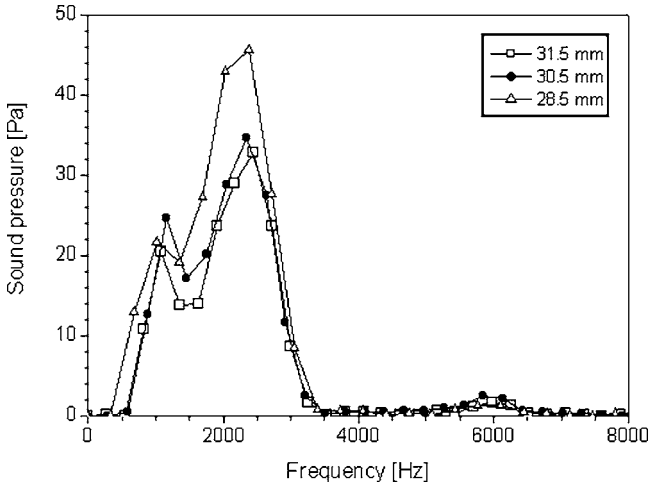


FIG. 4. Spectra of one single period of the sound in the shallot (without the resonator) for reed lengths 31.5 mm, 30.5 mm, and 28.5 mm.

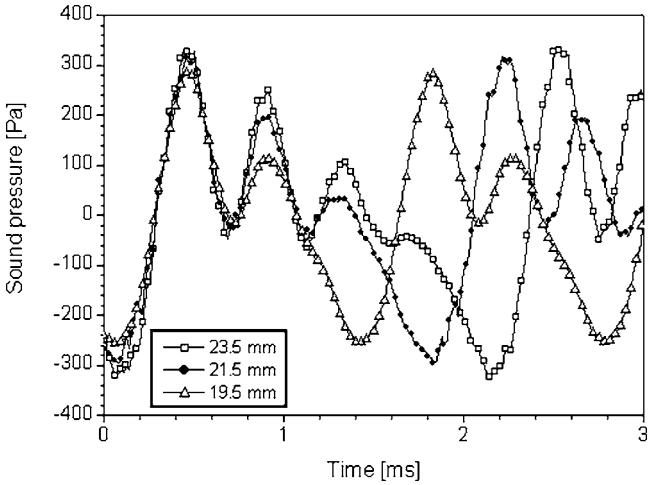


FIG. 5. Sound pressure wave forms in the shallot for reed lengths of 23.5 mm, 21.5 mm, and 19.5 mm. Two or three periods of the higher frequency oscillation in the shallot can still be recognized during each period of the reed vibration.

3 cm from the open throat (the smaller opening of the conical resonator). The eigenfrequencies of the resonator were determined from the frequency response shown in Fig. 6. The first four resonance frequencies are 242.5, 486, 729, and 973 Hz. Note that these frequencies are almost perfectly harmonic.

The effect of the resonator on the reed velocity and pipe sound was investigated for several reed lengths. The measurement was started by the maximum adjustable length (37.5 mm) of the reed with attached resonator. The pallet valve was opened, and was kept open for the entire duration of the measurements. That is, all the measurements were carried out in steady state. Reed velocity, the sound pressure in the shallot and at the open end of the resonator was recorded first, then the resonator was removed and the reed velocity and the sound pressure in the shallot were recorded again. The resonator was attached again, the reed was shortened, and the measurement procedure was repeated. Thus five signals (two reed velocity signals and three sound pressure signals) were recorded at each adjusted reed lengths.

The length to be adjusted was always approached from

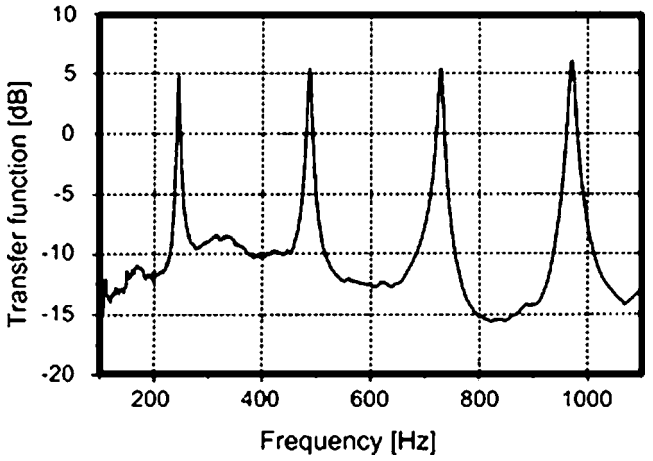


FIG. 6. Resonator transfer function (output spectrum and/or input spectrum) showing the first four acoustic modes of the resonator.

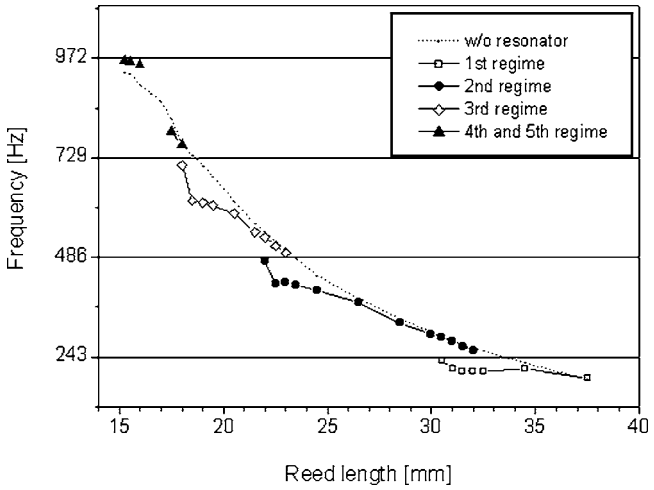


FIG. 7. Reed vibration frequency versus vibrating length with and without the resonator. The first four resonance frequencies (eigenfrequencies) of the resonator are shown as horizontal lines. The nominal frequency of the pipe (207.7 Hz) lies on the almost horizontal part of the first regime of the oscillation. Note the frequency jump and hysteresis near each resonator frequency.

both directions; i.e., by shortening and by lengthening the vibrating part of the reed. If the sound of the pipe was different for the two cases, the five measurements were carried out for both cases.

The frequency of the vibrating reed is shown as a function of reed length in Fig. 7, both with and without the pipe resonator attached. The blowing pressure was 80 mm water in this experiment. The reed frequency varies smoothly with reed length without the resonator, as would be expected,[9] but attaching the resonator lowers the frequency of the reed and also produces five distinct regimes of sounding, separated by four discontinuities, as the system jumps from one regime to the next one. Between the sounding regimes the sound is unstable, and between the second and third regimes there is a small range over which the pipe did not sound.

A hysteresis can also be observed: the reed length of jumping depends on the direction of tuning. This phenomenon occurs when the pipe is kept sounding and the tuning wire is shortened or lengthened in very fine steps. In this way the sound frequency can be brought into the regime of hysteresis, where two stable oscillations are possible at each reed length. In this regime (a domain of about 1.5–2 mm), two sounds with different pitch and timbre occur quite randomly for a given reed length, when the pipe is switched on and off. Certainly, this reed length domain cannot be used for musical purposes.

The first four resonance frequencies of the resonator (marked as dotted horizontal lines in Fig. 7) lie in the frequency domain of the observed discontinuities. Note that the nominal frequency of this pipe (207.7 Hz) is located on the almost horizontal part of the first regime at ~31.5 mm reed length.

Reed velocity wave forms with and without the resonator attached are shown in Fig. 8. The two wave forms are quite similar for 37.5 mm reed length, when the frequencies are close to each other [Fig. 8(a)]. When the frequency with the resonator is smaller than without it, the wave form is also

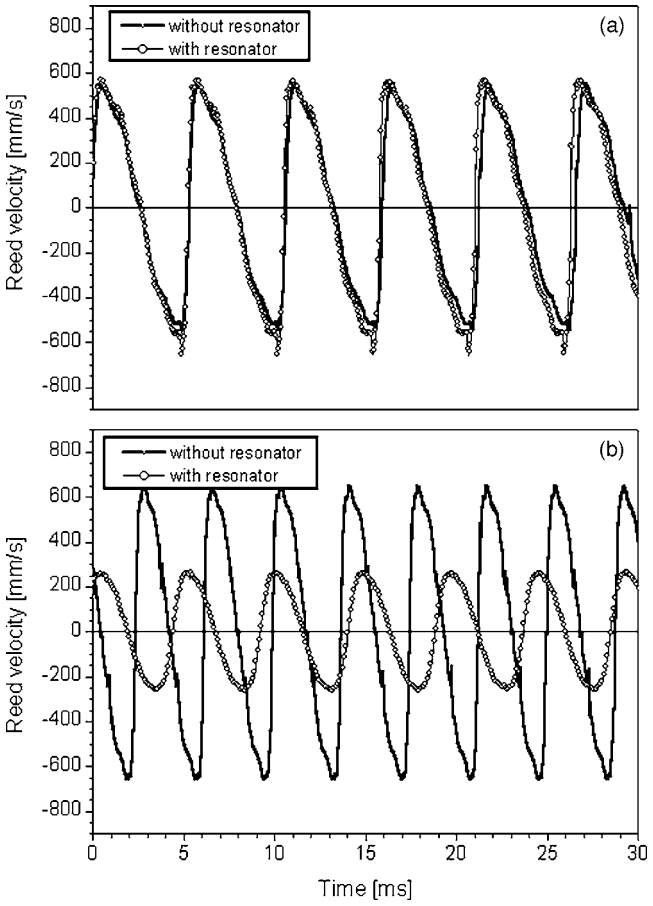


FIG. 8. Reed velocity wave forms for 37.5 mm (upper curve) and 32.0 mm (lower curve) reed lengths with and without the resonator. (a) Wave forms are similar when frequencies with (f=190.1 Hz) and without (f=187.8 Hz) the resonator are close to each other; (b) amplitude is smaller and wave form smoother when the frequency with (f=207.1 Hz) the resonator is smaller than without (f=259.1 Hz) it.

different. The amplitude of the vibration for 32.0 mm reed length [Fig. 8(b)] is much smaller, and the curve is much smoother, i.e., the high frequency content is reduced. At certain reed lengths (for example, at 30 mm and 21.5 mm) the vibration completely dies out.

The reed is unstable in such cases. Although it is possible to adjust the reed length to stop the vibration, the pipe cannot maintain this state indefinitely; it will start to sound after a while. The silent periods can be maintained for up to about 1 minute.

The amplitude of the vibration can be characterized by the rms values calculated from the velocity wave forms. They were plotted against reed frequency in Fig. 9, where the rms amplitudes without the resonator are also shown. Attaching the resonator, the vibration amplitude decreases, and it dies out at certain reed lengths. Around the first three eigenresonance frequencies of the pipe resonator (marked as vertical grid lines) the "forbidden" frequency domains can be clearly seen. The pipe cannot be tuned to a frequency in a "forbidden" domain. The frequency jumps over these domains, when the length of the reed is changed continuously by the tuning wire. The width of the forbidden domains was around 3–4 semitones (minor or major fifths). In order to get a strong, bright sound at a given note, the resonator should

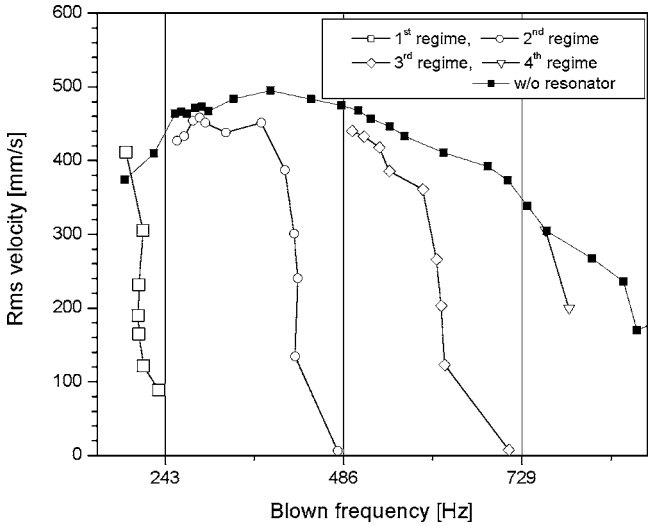


FIG. 9. Rms values of reed velocity without and with the resonator versus frequency. In the first case the rms curve is rather smooth with a maximum around 26.5 mm reed length (387 Hz). With resonator the rms amplitude of the vibration decreases rapidly as the frequency approaches the eigenfrequencies of the resonator from below.

be tuned about 1 semitone below the frequency of that note, so that the required frequency lays over the upper boundary of the "forbidden" domain. On the other hand, the resonator should be tuned about 3 semitones above the required frequency for a soft sound. For example, the resonator of the investigated reed pipe was tuned to 243 Hz, about 3 semitones over the nominal frequency of 207.7 Hz of the pipe. These results contradict the opinions given in Chaps. 13.4 and 17.11 of Ref. 10.

C. Sound in the shallot and at the mouth of the resonator

When the resonator is attached, the sound in the shallot is strongly influenced by the resonator. Sound pressure wave forms are presented in Fig. 10, where wave forms measured in the shallot with and without resonator are shown for the 37.5 mm reed length. The main changes are the significant

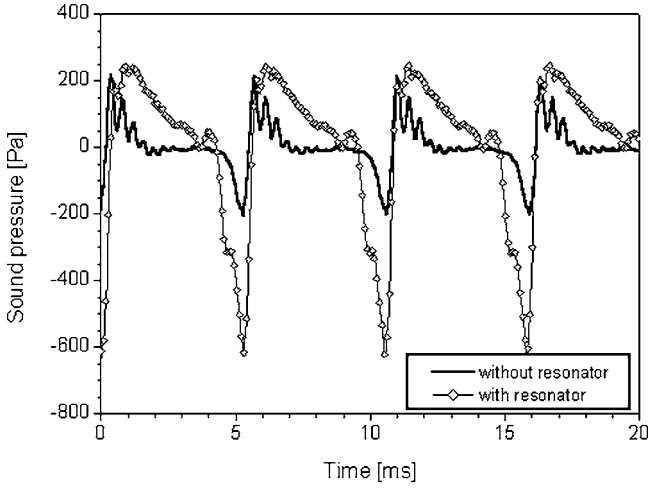


FIG. 10. Sound pressure in the shallot with (dashed curve) and without (solid curve) the resonator for 37.5 mm reed length.

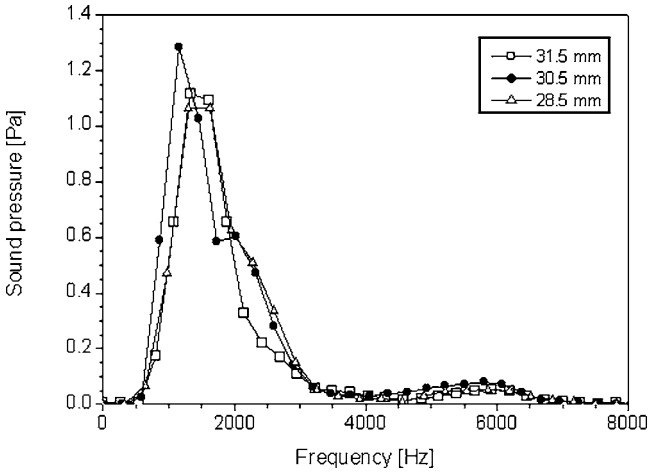


FIG. 11. Sound spectra of a single period at the open end of the resonator for the same three reed lengths (31.5 mm, 30.5 mm, 27.5 mm) as in Fig. 4.

increase of the negative part of the bipolar pressure pulse and the reduction of the amplitude and frequency of the high frequency oscillation.

Spectra of one single period of the bandpass-filtered (1 kHz–8 kHz) sound at the mouth of the resonator are shown in Fig. 11 for three (31.5, 30.5, and 28.5 mm) reed lengths. Broad peaks occur around 1500 Hz; their shape and amplitude is quite similar, in spite of the different sound frequencies (268.9 Hz, 290.0 Hz, and 326.0 Hz). These pipe frequencies can be found in the second regime shown in Figs. 7 and 9. The high frequency oscillation within a single sound period is shifted down to about 1.5 kHz from the 2.5 kHz range observed without the resonator, but it is still quite independent on the reed vibration (and sound) frequency.

Sound pressure wave forms measured in the shallot and close to the open mouth of the resonator are shown in Fig. 12 for 37.5 mm reed length. The wave form at the open mouth (upper curve) shows the characteristic pattern of reed pipe sound: a periodic train of a rapidly developing oscillation followed by a very rapid decay. Note that the sound pressure at the mouth is 20 to 50 times smaller than in the shallot (lower curve). Since the cross sectional area of the open mouth is about 50 times larger than that of the throat inserted into the block, such a reduction of the pressure can be expected.

IV. EVALUATION OF THE RESULTS

A. Sound in the shallot

A possible explanation of the observed sound pressure wave form in the shallot can be given as follows: The strong bipolar pressure pulse (see Fig. 2) is generated by the flow through the gap, which is modulated by the vibration of the reed. This bipolar pressure pulse is produced by aerodynamic forces and it provides the feedback for reed oscillation, as discussed in Ref. 9. Simultaneously, it excites the fundamental acoustic resonance of the shallot. Thus, sound generation in the shallot can be regarded as a by-product of the reed oscillation.

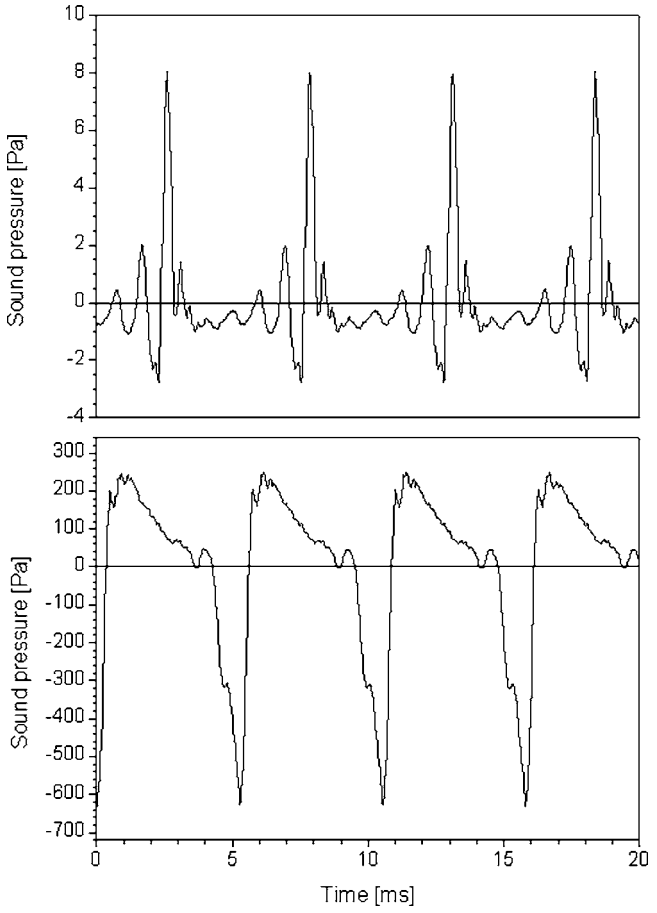


FIG. 12. Sound pressure wave forms in the shallot (lower curve) and close to the pipe mouth (upper curve) for 37.5 mm reed length.

The shallot is essentially a brass tube with open ends (an open end in the block and a side opening covered by the reed with a gap between the shallot and the reed) as shown in Fig. 1. The length of the shallot is 56.5 mm and it is inserted into the bore of the block. The inner diameter of the shallot is 6.5 mm. The length of the bore in the block above the shallot is about 23.5 mm, thus the entire length of the acoustic resonator tube is 80 mm with a jump of the cross sectional area with a factor of about 2 at 56.5 mm. The estimations of the fundamental frequency of such an acoustic system lead to values in the range of 2–3 kHz. The peak frequencies in Fig. 4 lay in this domain around 2500 Hz; thus it is reasonable to assume that the observed high frequency sound in the shallot corresponds to the fundamental acoustic resonance frequency of the shallot tube.

The frequency of the sound generated in the shallot is not constant as shown in Fig. 3 and in our earlier experiments,[4,5] it is shifted slightly upwards as the reed length increases. This effect can be explained by the reduction of the effective length of the shallot. As the reed rolls up from the shallot the length of the closed part of the shallot will be smaller; moreover, the end correction will also be smaller (it is inversely proportional to the square root of the area[10] of the opening). Both effects increase the eigenfrequency of the shallot.

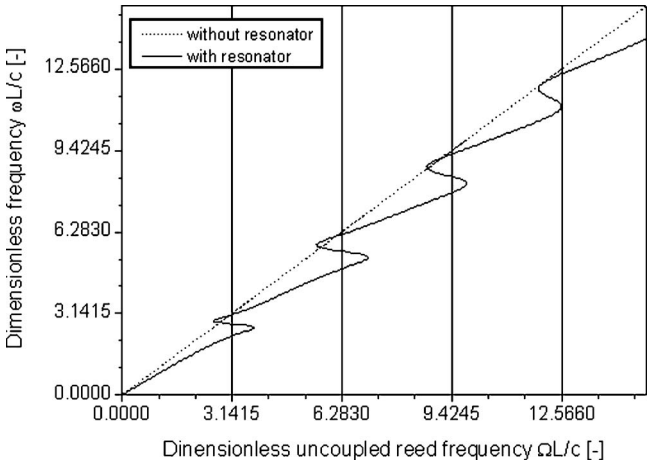


FIG. 13. Calculated dimensionless resonance frequency ($\omega L/c$) of the coupled reed and resonator as a function of the dimensionless reed frequency ($\Omega L/c$) without the resonator. Note the triple-valued function when the reed frequency is close to the eigenfrequencies of the resonator at the multiples of Pi.

B. Reed vibration with the resonator

The effects of reed-resonator coupling were very similar to those observed previously.[4] At the first occasion this behavior came as somewhat of a surprise. Although no experimental data have been published previously, to our knowledge, it has generally been felt that both the reed and the pipe resonator are tuned to the desired sounding frequency, which is mainly determined by the natural frequency of the reed.[10] The results in Figs. 7 and 9 show, however, that the resonator can lower the reed frequency by as much as 20% below its blown frequency without the resonator, when the reed frequency approaches an acoustic resonance from below. The vibration of the reed became more and more damped until the oscillation ceases. By further shortening the length of the reed, it starts to vibrate again with about the same frequency as the blown frequency without resonator. The reed pipe cannot sound in a frequency range of about 3–4 semitones around each acoustic mode frequency of the resonator ("forbidden" frequency domains).

This behavior can be explained by assuming that the reed and the acoustic resonator form a coupled vibrating system. A simple physical model of the coupling is discussed in the Appendix . In this model it is assumed that the particle velocity at the throat of the resonator equals the velocity of the reed, thus the input acoustic flow of the resonator can be given by the product of the area A of the throat and the reed velocity u. On the other hand, the force acting on the reed can be given as the product of the sound pressure p at the throat and the surface S of the reed. These assumptions are formulated in Eq. (A1) of the Appendix. The resonance frequencies of the coupled system and the damping of the reed are calculated from the dispersion equations (A8) as functions of the uncoupled reed frequency. The results of the calculations are shown in Figs. 13 and 14, while the measurement results of Fig. 7 are replotted for comparison in Fig. 15. The calculated (dimensionless) frequencies $\omega L/c$ of the coupled system are plotted versus the (dimensionless) reed frequency $\Omega L/c$ in Fig. 13, where ω, Ω, L, and c are the

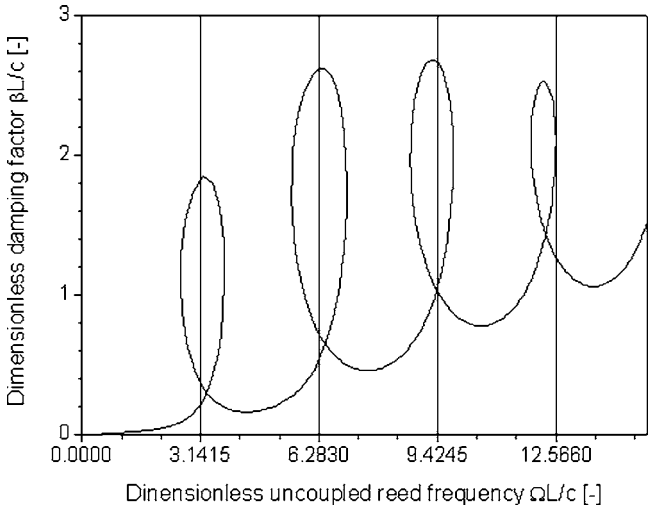


FIG. 14. Dimensionless damping constant of the reed vibration of the coupled reed-resonator system as a function of the dimensionless reed frequency ($\Omega L/c$) without the resonator.

angular frequency of the sounding reed pipe, the angular frequency of the sound and reed vibration without the resonator, the length of the acoustic resonator and the velocity of sound, respectively. The dispersion curve in Fig. 13 contains an instability region around each acoustic eigenfrequency. In these regions the curve is three valued. The lower and upper parts of the dispersion curve have positive slopes; these parts describe stable solutions of the dispersion equations. The middle part has a negative slope; in that domain the solution is unstable.

The (dimensionless) damping coefficient $\beta L/c$ of the reed (where β is the attenuation coefficient of the reed in s^{-1}) is shown as a function of the (dimensionless) uncoupled reed frequency in Fig. 14. The damping of the reed vibration increases significantly as the uncoupled vibration frequency approaches the instability region.

The calculated curve in Fig. 13 helps to understand the measured one in Fig. 15. Physically, the sound pressure at the narrower end of the conical resonator will oppose the vibration of the reed, as its frequency approaches an acoustic resonance. This effect results in a lowering of the reed frequency and an increase of the damping of the reed.

However, the comparison of Figs. 13 and 15 also show that the simple model of coupling can only qualitatively explain the observed phenomena. The calculated and measured frequencies differ quite significantly, but the main features of the dispersion curves are similar.

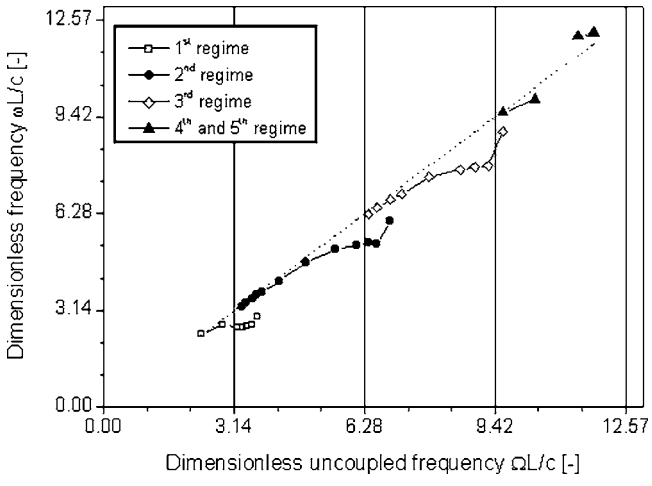


FIG. 15. Measured dimensionless frequency of the coupled reed and resonator as a function of the dimensionless reed frequency ($\Omega L/c$) without the resonator (Fig. 7 replotted). Note the similarity in overall appearance and the differences in details between the calculated curve in Fig. 13 and the measured curve in this figure.

C. Sound in the shallot and at the mouth of the resonator

Without the resonator the sound is dominated by the fundamental acoustic mode of the shallot. After attaching the resonator the acoustic system becomes much more complicated. If there is no change in cross section between the shallot and the resonator tube, the system can be treated as a single acoustic resonator. In practice, however, there is a cross section mismatch between the shallot and the resonator which reflects the acoustic wave. The reflection coefficient, calculated from the diameters of resonator and shallot is about 0.33. In such cases the shallot acts further on as an acoustic resonator, but its resonances will be shifted down because of the acoustic load of the resonator tube. The fundamental shallot mode can couple to those higher eigenmodes of the conical resonator tube whose frequencies are close to the shallot's frequency. However, the eigenfrequencies of the conical tube are more stable than that of the shallot; thus they can stabilize the frequency of the shallot mode. This effect can be observed by comparing the single period spectra in Figs. 4 and 11. The broad peaks are shifted down in frequency, and become narrower, when the resonator is attached.

It is well known, that a periodic signal can be regarded as a convolution of a periodic train of Dirac δ pulses and of the wave form of one single period. Correspondingly, the spectrum of a periodic signal can be regarded as the product of the spectrum of the Dirac δ-pulse train and the spectrum of a single period. The former contains an infinite series of harmonic lines; the latter determines the envelope of the resulting spectrum. Therefore, the broad peaks observed in the single period sound spectra can influence the formant structure of the sound of reed pipes. This may be the physical ground for the empirical experience of organ builders about the significant influence of the size and shape of the shallot on the timbre of the pipe sound.

Since the frequency domain of the formant established by the shallot mode does not change when the pipe is tuned, the formant maximum is shifted continuously over the overtones in the course of tuning. This is one reason why the timbre of reed pipes changes when tuned. The other reason is the change in the damping of the reed vibration. As the reed vibrates more smoothly, the bipolar pressure pulse will be broader, and less high frequency sound can be excited in the shallot and resonator. This effect can be observed in Fig. 8.

V. DISCUSSION OF THE RESULTS

In a clarinet, the reed resonance is much (6 to 10 times) higher than the blown note. The oscillation condition (that the negative of the real part of the reed admittance must be greater than the real part of the pipe admittance; see Eq. 13.12 in Ref. 10) can be satisfied by any of the lower pipe modes. However, the lowest mode has the highest impedance, so that the clarinet prefers to vibrate in this mode, although overblowing can coax it to vibrate in a higher mode (change to a higher register). In the Trompette organ pipe, on the other hand, the frequency appears to be stable to change in the wind pressure, and the jump to a higher frequency regime requires a shortening of the vibrating reed length. The observed frequency jumps, however, cannot be regarded as mode transitions. The reed pipe cannot sound at the mode frequencies of the resonator.

The rather surprising effects of reed-resonator coupling can be qualitatively explained by the simple physical model presented in the appendix . The coupling parameters depend on the density and thickness of the reed and on the length, the average diameter, the conical angle and the attenuation coefficient of the acoustic resonator.

The high-frequency oscillations observed inside the shallot were also somewhat of a surprise. Although they are less prominent with the pipe resonator attached than without it, they could strengthen certain high harmonics in the pipe sound, contributing a formantlike effect. That is, the timbre of a reed pipe may be influenced by the size and shape of the shallot. This statement corresponds to the experience of organ building. Reed pipe makers and voicers use a broad variation of shallot shapes and sizes in their work to produce the desired timbre and speech of the reed pipes.

The data reported in this paper are from a single G4 Trompette pipe, but similar effects are observed in other reed pipes as well.[11] In a rankett pipe, having a closed cylindrical resonator, the frequency jumps are not as well defined, as shown in Fig. 7.

VI. CONCLUSION

Contrary to the opinion often expressed, coupling between the vibrating reed and the pipe resonator appears to be a very important factor in the acoustical behavior of reed organ pipes. In this respect, they resemble the woodwind instruments. Without the pipe resonator, the reed vibration frequency is a continuous function of the reed vibrating length, but when the resonator is added the frequency jumps abruptly from below to above each resonance frequency of the resonator.

Although design of reed organ pipes is a well-developed art, it has developed as an empirical art and is very time consuming. It is believed that a good understanding of the acoustical and mechanical aspects of sound generation in these pipes will allow organ builders and voicers to more easily and more accurately voice them so as to obtain the desired sound.

ACKNOWLEDGMENTS

The research was financially supported by the Förderverein Orgelforschung (FOF). The authors express their thanks to Neville Fletcher for his helpful comments and to the organ builder B. Welde for providing the reed pipes for the measurements. One of the authors (T.D.R.) expresses thanks to the Fraunhofer-Institut Bauphysik, Stuttgart for supporting his research visit.

APPENDIX

It is assumed that the reed and the conical resonator form a coupled system. The acoustic pressure at the smaller opening of the resonator provides a force on the reed, while the velocity of the reed provides the input acoustic flow for the resonator,

$$Z_m u + Sp = 0, \quad Au + Y_a p = 0, \tag{A1}$$

where u, Z_m, S, p, A, and Y_a are reed velocity, the mechanical impedance of the reed, the surface area of the reed, acoustic pressure, area of the resonator opening and acoustic input admittance of the resonator, respectively.

Nontrivial solution of Eq. (A1) exists only if the determinant is zero,

$$Z_m Y_a - AS = 0. \tag{A2}$$

Dividing by the product AS the following dispersion equation can be derived:

$$z_m y_a = 1, \tag{A3}$$

where $z_m = Z_m/S$ and $y_a = Y_a/A$ are the specific impedance of the reed and the specific input admittance of the resonator, respectively.

The specific acoustic input admittance of an open conical resonator is given as[10]

$$y_a = \frac{1}{\rho c}\left(\frac{1}{jKR_1} + \frac{1}{j\tan KL}\right), \tag{A4}$$

where ρ, c, R_1, L, $K = k - j\kappa = (\omega/c)(1 - j\alpha)$, ω and α are the air density, sound velocity, the distance of the input opening from the origin of the cone, the length of the resonator, the complex wave number, the angular frequency and the dimensionless loss factor of the resonator, respectively. The dimensionless loss factor α can be expressed by the quality factor of the resonator as $\alpha = 1/2Q$. The quantity R_1 can be determined from the input radius r_1 and the half-angle of the cone $\varphi/2$ as $R_1 = r_1/\sin(\varphi/2)$.

By separating the real and imaginary parts the specific acoustic input admittance can be written as follows:

$$\begin{aligned} y_a &= \frac{-1}{\rho c}(R_a - jI_a) \\ &= \frac{-1}{\rho c}\left(\frac{\alpha E}{x} + \frac{\sinh(\alpha x)\cosh(\alpha x)}{\sin^2 x + \sinh^2(\alpha x)}\right) \\ &\quad + \frac{j}{\rho c}\left(\frac{E}{x} + \frac{\sin x \cos x}{\sin^2 x + \sinh^2(\alpha x)}\right), \end{aligned} \tag{A5}$$

where $x=\omega L/c$ is the dimensionless frequency variable and $E=L\sin(\varphi/2)/r_1(1+\alpha^2)$.

The reed is regarded as a simple harmonic oscillator with the mechanical impedance,

$$Z_m = -2\beta m - j\frac{m(\Omega^2-\omega^2)}{\omega}, \tag{A6}$$

where m, Ω, and β are the effective mass of the reed, the resonance frequency and the damping constant of the reed, respectively.

In the vicinity of the eigenfrequency of the reed, its specific impedance can be approximated as follows:

$$z_m = \frac{-\rho_r d_r c}{L}\left(2\gamma + j\frac{X^2-x^2}{x}\right), \tag{A7}$$

where ρ_r, d_r, $\gamma=\beta L/c$, and $X=\Omega L/c$ are the density, the thickness, the dimensionless damping constant and resonance frequency of the reed, respectively.

By substituting Eqs. (A5) and (A7) to (A3) and separating the real and imaginary part the following pair of equations can be derived for the unknown quantities γ and X^2:

$$2xR_a\gamma + I_a X^2 = x^2 I_a + Cx, \quad 2xI_a\gamma - R_a X^2 = -x^2 R_a, \tag{A8}$$

where $C=\rho L/\rho_r d_r$.

The solution of (A8) is given by the following expressions:

$$\gamma = \frac{R_a C}{2(R_a^2+I_a^2)}, \quad X^2 = x^2 + \frac{xI_a C}{(R_a^2+I_a^2)}. \tag{A9}$$

The solution of the dispersion equation (A3) for parameter values of $\gamma=0.06$, $E=6.0$, and $C=4$ are presented in Figs. 13 and 14.

[1]N. H. Fletcher, "Sound production by organ flue pipes," J. Acoust. Soc. Am. **60**, 926–936 (1976).
[2]A. Miklos and J. Angster, "Properties of the sound of labial organ pipes," Acust. Acta Acust. 86, 611–622 (2000).
[3]J. W. Coltman, "Jet drive mechanisms in edge tones and organ pipes," J. Acoust. Soc. Am. **60**, 725–733 (1976).
[4]T. D. Rossing, J. Angster, and A. Miklos, "Reed vibration and sound generation in lingual organ pipes," J. Acoust. Soc. Am. **104**, 1767–1768 (1998).
[5]T. D. Rossing, J. Angster, and A. Miklos, "Reed vibration and sound generation in lingual organ pipes," *Proceedings of the International Symposium on Music and Acoustics*, Perugia, Italy, 2001, pp. 313–316.
[6]J. Braasch, J. Angster, and A. Miklos, "The influence of the shallot leather facing on the sound of lingual organ pipes," *Proceedings of the International Symposium on Music and Acoustics*, Perugia, Italy, 2001, pp. 325–328.
[7]J. Angster, G. Paál, W. Garen, and A. Miklos, "Effect of voicing steps on the stationary spectrum and attack transient of a flue organ pipe," Proc. Inst. Acous. **19**, 285–294 (1997).
[8]R. Janke (organ builder) http://www.orgel-info.de.
[9]A. Miklos, J. Angster, S. Pitsch, and T. D. Rossing, "Reed vibration in lingual organ pipes without the resonators," J. Acoust. Soc. Am. **113**, 1081–1091 (2003).
[10]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), Chap. 16.
[11]E. Cox and T. D. Rossing, "Regimes of oscillation and reed vibrations in lingual organ pipes," J. Acoust. Soc. Am. **111**, 2395 (2002).

Measurements of the anisotropy of ultrasonic attenuation in freshly excised myocardium

Steven L. Baldwin, Karen R. Marutyan, Min Yang, Kirk D. Wallace, Mark R. Holland, and James G. Miller
Department of Physics, Washington University, One Brookings Drive, Campus Box 1105, St. Louis, Missouri 63130

(Received 24 August 2005; revised 18 February 2006; accepted 22 February 2006)

Echocardiography requires imaging of the heart with sound propagating at varying angles relative to the predominant direction of the myofibers. The degree of anisotropy of attenuation can significantly influence ultrasonic imaging and tissue characterization measurements *in vivo*. This study quantifies the anisotropy of attenuation of freshly excised myocardium at frequencies typical of echocardiographic imaging. Results show a significantly larger anisotropy than previously reported in specimens of locally unidirectional myofibers. Through-transmission radio frequency-based measurements were performed on specimens from 12 ovine and 12 bovine hearts. Although ovine hearts are closer in size to human, the larger size of bovine hearts offers the potential for specimens in which myofibers are more nearly unidirectionally aligned. The attenuation coefficient increased approximately linearly with frequency. The mean slope of attenuation with frequency was 3–4 times larger for propagation parallel than for perpendicular to the myofibers. At perpendicular insonification, slopes between ovine and bovine myocardium were approximately equal. However, attenuation in bovine specimens was larger for angles approaching parallel. The difference in results for parallel appears consistent with what might be expected from increased myofiber curvature associated with smaller lamb hearts. Quantitative knowledge of anisotropy of attenuation may be useful in understanding mechanisms underlying the interaction of ultrasound with myocardium. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2188333]

PACS number(s): 43.80.Cs, 43.80.Ev [FD] Pages: 3130–3139

I. INTRODUCTION

Attenuation plays a significant role in ultrasonic imaging and tissue characterization. Ideally, measurements of the anisotropy of attenuation in myocardium would be representative of results for myofibers oriented along a single direction. However, such measurements are precluded by the regional myofiber architecture of the heart which is complex (Greenbaum *et al.*, 1981; Harrington *et al.*, 2005; LeGrice *et al.*, 1995; Streeter *et al.*, 1969). The purpose of this investigation was to quantify the anisotropy of attenuation in freshly excised myocardial tissue in which the myofibers are locally unidirectional and to compare those results with earlier measurements of attenuation on freshly excised myocardium. Results presented in the current study show a larger anisotropy of attenuation in myocardium than previously reported. A quantitative knowledge of myocardial attenuation and its dependence on angle of insonification with respect to the local predominant direction of myofibers may be important in echocardiographic modeling and the proper interpretation of diagnostic results. For example, by accounting for myocardial attenuation, the well-known phenomenon of drop out in the lateral and septal walls in parasternal short-axis echocardiographic views of the heart can be mitigated in studies of wall motion and estimates of ventricular volume (Aygen and Popp, 1987; Holland *et al.*, 1998; Melton and Skorton, 1983; Perez *et al.*, 1992; Recchia *et al.*, 1993). In addition, quantitative characterization of cardiac tissue through analyses of the apparent backscatter are influenced by attenuation. The cyclic variation of myocardial backscatter has shown promise as a diagnostic tool (Barzilai *et al.*, 1984; Castaldo *et al.*, 2000; Ciliberto *et al.*, 1996; Colonna *et al.*, 1999; D'Hooge *et al.*, 2000; Di Bello *et al.*, 2001; Finch-Johnston *et al.*, 2000; Goens *et al.*, 1996; Hancock *et al.*, 2002; Holland *et al.*, 2004; Kovacs *et al.*, 2004; Mottley *et al.*, 1984; Naito *et al.*, 1994; Ohara *et al.*, 2005; Pasquet *et al.*, 1998; Pislaru *et al.*, 2001; Rijsterborgh *et al.*, 1993; Sagar *et al.*, 1990; van der Steen *et al.*, 1997; Vandenberg *et al.*, 1991; Vered *et al.*, 1989; Wickline *et al.*, 1986; Yuda *et al.*, 2002). However, measurements of the cyclic variation of "apparent" integrated backscatter (uncompensated for the effects of attenuation) are influenced by both the attenuation of overlying tissue as well as that occurring in the region-of-interest. Recent steps have been taken with excised, formalin-fixed tissue toward the goal of measurement of myocardial attenuation *in vivo* using two-dimensional B-mode images (Baldwin *et al.*, 2005; Sosnovik *et al.*, 2001a; Sosnovik *et al.*, 2001b) and M-mode (Baldwin *et al.*, 2005; Baldwin *et al.*, 2004). The ability to separate the contributions of attenuation from the cyclic variation of the "apparent" integrated backscatter may lead to a better understanding of the underlying physiological or pathophysiological mechanisms of observed results. In addition, myocardial attenuation measured over the heart cycle may prove to be a useful marker of disease in its own right (van der Steen *et al.*, 1991; Wear *et al.*, 1989).

Previous studies have reported the attenuation of freshly excised myocardial tissue in the low frequency range (typical

 0001-4966/2006/119(5)/3130/10/$22.50

for adult echocardiographic imaging) at insonification perpendicular and/or parallel to the predominant orientation of the myofibers (Brandenburger *et al.*, 1981; Dwyer *et al.*, 2000; Fei and Shung, 1985; Fei *et al.*, 1987; Klepper *et al.*, 1981; Mimbs *et al.*, 1980). Early studies from this laboratory examined the anisotropy of attenuation for angles of insonification (over a complete rotation) in freshly excised canine heart (Mottley and Miller, 1990). These investigations observed a clear anisotropy of attenuation in myocardium and demonstrated that the net attenuation through a length of myocardial tissue depends upon the angle of the myofibers encountered along the path of the ultrasonic beam. However, measurements from that study were performed on specimens excised in a different orientation which may have included some off-angle components. The purpose of this study is to measure the anisotropy of attenuation of freshly excised myocardium, at frequencies typical of those used in clinical echocardiography, in specimens in which individual myofibers were nearly unidirectionally aligned. This was accomplished by careful excision of specimens in specific anatomical orientations. Broadband measurements of the anisotropy of attenuation were performed on tissues from standardized regions of 12 ovine and 12 bovine freshly excised hearts. These measurements represent a further step towards the eventual goal of measurements of attenuation *in vivo*.

II. METHODS

A. Preparation of specimens and acquisition of ultrasonic data

A total of 36 tissue specimens from 12 freshly excised lamb hearts and 45 tissue specimens from 12 freshly excised cow hearts were investigated in this study. The hearts were obtained within 30 min of slaughter from a local commercial slaughterhouse and immersed in 0.9% saline solution at room temperature prior to specimen coring. From each ovine heart, three cylindrical plugs were cored, two from the left ventricular free wall and one from the septum using a 14.5 mm inner-diameter coring tool. Specimens were cored from endocardium to epicardium such that the axis of rotation of the cylindrical specimens was oriented orthogonal to the epicardial surface of the heart. The two left ventricular free wall specimens were cored in approximately the same plane perpendicular to the apex-to-base axis of the heart and superior to the anterior and posterior papillary muscles. The septal specimens were cored approximately halfway between the plane of the left ventricular specimens and the apex.

The larger size of bovine hearts permitted a total of four cylindrical plugs to be cored. Three were prepared in an identical manner to that for the ovine specimens, and an additional left ventricular specimen was obtained in the same plane as the other left ventricular specimens as described above, but from the region between the anterior and posterior papillary muscles. In addition, specimens were cored from bovine anterior papillary muscles. However, because of the sometimes irregular shapes of the anterior papillary muscles and the resulting difficulty of coring specimens in which myofibers should be approximately unidirectional, specimens oriented at angles near parallel insonification may involve myofibers that were off angle in elevation. This standardized method of coring provided for a regional comparison of the anisotropy of the myocardial specimens.

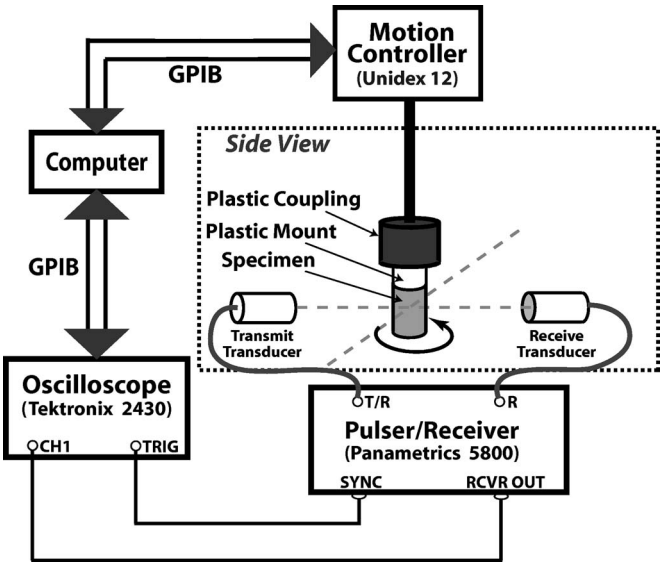


FIG. 1. Experimental setup for performing measurements of the anisotropic properties of myocardial attenuation.

Each myocardial specimen was cored from an intact left ventricle at room temperature that had been stored in the saline solution to reduce the potential for osmotic effects. Immediately after coring, the specimen was mounted and then immersed in a water bath (also at room temperature) approximately 5 to 10 min prior to measurements. The total time from death of the animal to completion of all measurements on that particular heart was approximately three to five hours. Temperature changes during any one data acquisition were no more than 0.1 °C. Figure 1 shows the equipment set up for measurements. The through-transmission radiofrequency data were acquired using a matched pair of 5 MHz focused piezoelectric transducers (Panametrics V309, 1/2 in. diameter; Panametrics, Waltham, MA) separated by twice the nominal focal length of 2 in.. Each tissue specimen was positioned symmetrically at the focus of both transducers and the ultrasonic beam was centered on the midmyocardial region of each specimen. An electric pulse was generated with a Panametrics 5800 Pulser/Receiver operating in pitch/catch mode and sent to the transmit transducer. The through-transmitted signal was received by the second transducer, amplified by the Panametrics 5800, and digitized with a Tektronix 2430A oscilloscope (Tektronix, Beaverton, OR).

For data acquisition, the plastic mount of each myocardial specimen was friction fit into a cylindrical plastic coupling that was, in turn, attached to the metal shaft of the rotational axis. A Unidex 12 motion controller (Aerotech, Pittsburgh, PA) was utilized to permit the measurement assembly to have x, y, and z translational movements as well as rotational freedom. A computer (Apple Computer, Cupertino, CA) served as a data acquisition system controller and off-line storage device for digitized data. Each myocardial specimen was rotated under computer control by a measurement angle, θ (incremented in 5° steps). Substitution through-transmission data were obtained by averaging 32 radio-

frequency traces at each incremented angle over a complete rotation. In addition, immediately after through-transmission data acquisition, pulse-echo data were acquired for specimen thickness measurements at each incremented angle. Corresponding water path reference data were obtained to within ±0.1 °C of the corresponding specimen data.

B. Inherent curvature of myofibers in myocardial specimens

In order to measure the anisotropy of attenuation inherent to the myofibers in myocardial tissue, a specimen would ideally be obtained in which the myofibers are perfectly unidirectional. Visual inspection and histological studies of postmortem human hearts by Greenbaum *et al.* (1981) and studies of canine hearts by Streeter *et al.* (1969) have shown that myofibers are predominantly circumferential in the midwall of the left ventricle, but rotate into and out of the short axis (transverse cross-sectional) plane near the endocardial and epicardial surfaces. Myocardial specimens were cored radially in a plane parallel to the standard short axis echo view of the heart as shown in Fig. 2 (upper panel). The specimen core diameter was chosen in order to be large enough to permit approximately perpendicular incidence of the ultrasonic beam with the surface of the myocardial tissue. (In the focal zone, calculated −6 dB beamwidths were 2.0, 1.2, and 0.9 mm at 3, 5, and 7 MHz, respectively). In addition, in our experience, the core diameter needed to be of sufficient size to yield specimens of uniform shape. As illustrated in Fig. 2 (lower panel), although measurements were performed by rotating a measurement angle θ about the axis of symmetry of the cylindrical specimen, a certain amount of circumferential curvature of the myofibers was unavoidable and remained approximately constant throughout rotation upon data acquisition. This curvature was most pronounced for angles of insonification near parallel to the predominant direction of the myofibers ($\theta \approx 0°$) and, in principle, should have no effect when myofibers are primarily perpendicular to the direction of insonification ($\theta \approx 90°$).

When the tangent to myofibers on the axis of rotation is approximately parallel to the direction of insonification, the tangent to myofibers at the edge of the specimen can be approximated to be at an angle ϕ_{maximum}, where

$$\sin \phi_{\text{maximum}} = \frac{r_{\text{specimen}}}{R_{\text{left ventricle}}} \tag{1}$$

with r_{specimen} the radius of the specimen and $R_{\text{left ventricle}}$ the radius of the left ventricle (see Fig. 2, upper panel). We estimate the angle ϕ_{maximum} for bovine myocardial specimens to be approximately 4° to 7°, and approximately 18° to 28° for ovine myocardial specimens in this study. The larger size of bovine hearts offers the possibility of obtaining myocardial specimens in which the circumferential curvature of the myofibers is less and myofibers are more nearly unidirectionally (uniaxially) aligned than for the case of the smaller ovine hearts.

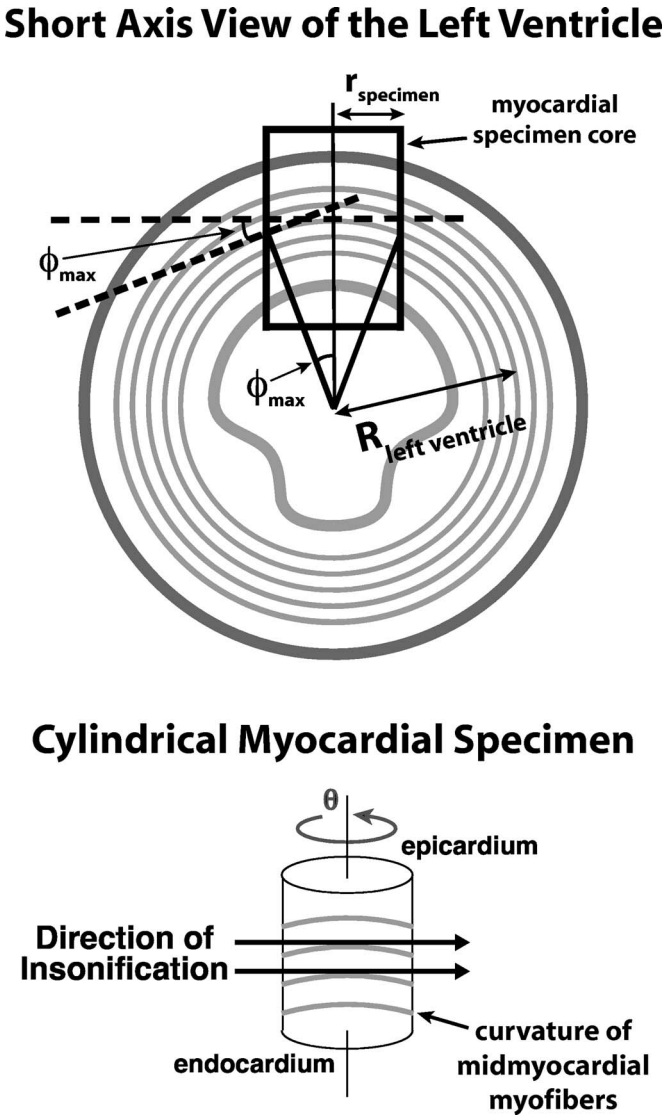


FIG. 2. Approach to coring myocardial specimens used in this study (upper panel). Midmyocardial myofibers are expected to exhibit a circumferential curvature such that myofibers near the edges are at an angle ϕ compared to myofibers in the center that are parallel to the direction of insonification (lower panel).

C. Minimizing potential contributions to phase cancellation at the face of a piezoelectric transducer

Because of the cylindrical shape of the myocardial specimens, the potential exists for a refractive alteration of the beam direction. This could result in a variation in the portion of the through-transmitted beam captured by the receiving transducer as well as phase cancellation at the face of the receiving piezoelectric transducer. Both of these effects could result in potential errors in the measured attenuation coefficient and the frequency dependence of the attenuation (Busse and Miller, 1981a; Busse and Miller, 1981b; Miller *et al.*, 1976; Mimbs *et al.*, 1980; Verdonk *et al.*, 1996). Several steps were taken to minimize the potential for errors associated with phase cancellation. Focused transducers were used for data acquisition because previous measurements in freshly excised, intact myocardium using focused transducers exhibited no significant phase cancellation effects (O'Donnell *et al.*, 1977). Care was taken to align the scan-

ning apparatus and minimize precession of the specimen upon data acquisition (Mimbs *et al.*, 1980; Verdonk *et al.*, 1996). In addition, myocardial specimens were excised with the largest possible diameter for study (14 mm) in order to permit as close as possible to normal insonification at the surface of the myocardial specimens. As a check of the potential extent of phase cancellation artifact, measurements of the anisotropy of attenuation in eight formalin-fixed myocardial specimens were compared with two-dimensional pseudoarray scan measurements of the same specimens using a 0.5 mm needle hydrophone. Mean measurements of the frequency dependence of attenuation as a function of angle of insonification showed no significant differences between the two methods of measurement.

D. Analysis of ultrasonic attenuation data

The experimental procedure involved a through-transmission substitution technique where the specimen to be measured replaces a portion of the water whose attenuation was small enough to be neglected for the frequencies and path lengths used in this study. In this approach, the measured power spectra with and without the specimen inserted are compared in order to determine the attenuation coefficient. Without the specimen, a reference through-transmitted radiofrequency signal is obtained whose power spectrum is given by

$$|V_{\text{reference}}(f)|^2 = \Gamma(f)|P(f)|^2, \tag{2}$$

where f is frequency, $\Gamma(f)$ represents measurement system dependent effects that are the same for both reference and specimen, and $|P(f)|^2$ is the frequency-dependent power spectrum of the transmitted pulse. A myocardial specimen is then inserted into the path of the insonifying beam. The power spectrum of the through-transmitted signal is given by

$$|V_{\text{specimen}}(f,\theta)|^2 = \Gamma(f)|P(f)|^2(T_I^2)e^{-2\alpha(f,\theta)L(\theta)}, \tag{3}$$

where T_I^2 is the combined intensity transmission coefficient associated with the crossing of a single plane wave at a water-to-tissue interface and a tissue-to-water interface, $\alpha(f,\theta)$ is the amplitude attenuation coefficient as a function of frequency and angle of insonification relative to the predominant direction of the myofibers, and $L(\theta)$ is the thickness of the tissue. The factor of 2 in the exponential is present because power, rather than amplitude, is being measured. The signal loss, as a function of f and θ, in units of decibels (dB), is given by

$$\text{signal loss}_{\text{decibels}}(f,\theta) = 10\log_{10}\frac{|V_{\text{reference}}(f)|^2}{|V_{\text{specimen}}(f)|^2} \tag{4}$$

$$= -10\log_{10}(T_I^2) + (20\log_{10}e)\alpha(f,\theta)L(\theta). \tag{5}$$

The first term on the right-hand side of Eq. (5) is an "insertion loss" resulting from nonunity transmission at interfaces.

Previous studies have demonstrated a linear with frequency dependence of the attenuation coefficient over a band-limited frequency range in many soft tissues (Mimbs *et al.*, 1980; Mottley and Miller, 1990; Tu *et al.*, 2003; Verdonk *et al.*, 1996). We consider a Taylor series expansion of the attenuation coefficient about the center frequency of the useful bandwidth: $\alpha(f,\theta) = \alpha_0(\theta) + \beta(\theta)\cdot(f) + \cdots$. The slope of attenuation then becomes

$$\beta_{\text{decibels}}(\theta) = \frac{\frac{\partial}{\partial f}\text{signal loss}_{\text{decibels}}(f,\theta)}{L(\theta)}. \tag{6}$$

Because of the derivative, the slope is not influenced by frequency-independent transmission losses at the boundaries. To calculate the slope of attenuation at any given angle of insonification, the power spectrum (expressed in decibels, dB) of the through-transmitted specimen pulse was subtracted from the power spectrum of the reference pulse. Any differences in equipment gains used to capture the two signals were properly taken into account in the subtraction of the two spectra. A line was fit to the resulting signal loss as a function of frequency using a least-squares fit over the useful bandwidth (3 to 7 MHz). The slope of the best fit line, in units of dB/MHz was divided by the thickness of the specimen $L(\theta)$ to obtain the slope of attenuation in units of dB/(cm MHz) as a function of angle θ. Individual specimen thickness was determined with ultrasonic pulse echo measurements in an approach that has been described previously (Baldwin *et al.*, 2005).

III. RESULTS

The anisotropy of the slope of attenuation in freshly excised ovine and bovine myocardium is shown in the top panels of Figs. 3 and 4, respectively (mean ± one standard deviation). In both cases, peaks near parallel insonification are more narrow than troughs near perpendicular insonification. For ovine myocardium, slopes of attenuation range from 0.52 ± 0.07 dB/(cm MHz) for perpendicular to 1.88 ± 0.17 dB/(cm MHz) for parallel insonification. In the case of bovine myocardium, slopes of attenuation range from 0.55 ± 0.16 dB/(cm MHz) for perpendicular to 2.22 ± 0.23 dB/(cm MHz) for parallel insonification. The bottom panels of Figs. 3 and 4 show corresponding values of the attenuation coefficient at 3 MHz, typical of frequencies employed in echocardiographic imaging, expressed in dB/cm (using $\alpha \cong \beta \cdot f$).

For 27 out of 36 freshly excised ovine specimens measured, the temperature was maintained between 19.1 and 20.1 °C. For the remaining nine ovine specimens, temperature extremes were 17.6 to 19.1 ° C. Mean results excluding these specimens were not significantly different. For 38 out of 45 freshly excised bovine specimens measured, the temperature was maintained between 19.3 and 20.4 °C. For the remaining seven bovine specimens, temperature extremes were 16.2 to 18.8 °C. As with the lamb hearts, mean results excluding these specimens were not significantly different.

As shown in Fig. 5, a comparison of results from different regions of the left ventricle and septum show relatively close agreement for all angles of insonification, both for ovine and bovine myocardium. Left ventricular free wall regional averages include specimens taken from just superior to the anterior papillary muscle (12 for lamb and 11 for cow), from just superior to the posterior papillary muscle (12 for

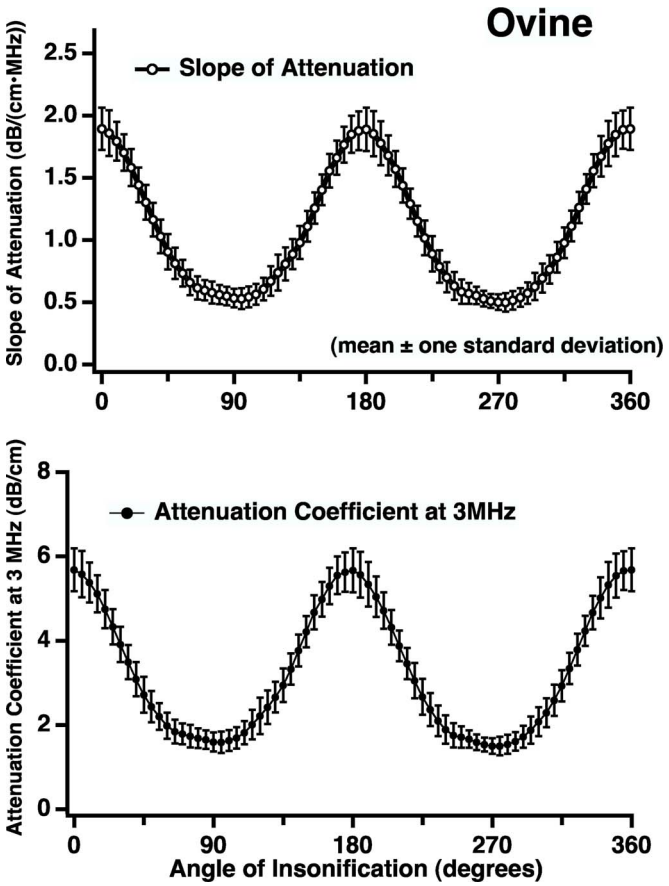


FIG. 3. Slope of attenuation [in dB/(cm MHz)] in freshly excised ovine myocardium (mean±one standard deviation) (upper panel). Corresponding attenuation coefficient at 3 MHz (dB/cm) (lower panel).

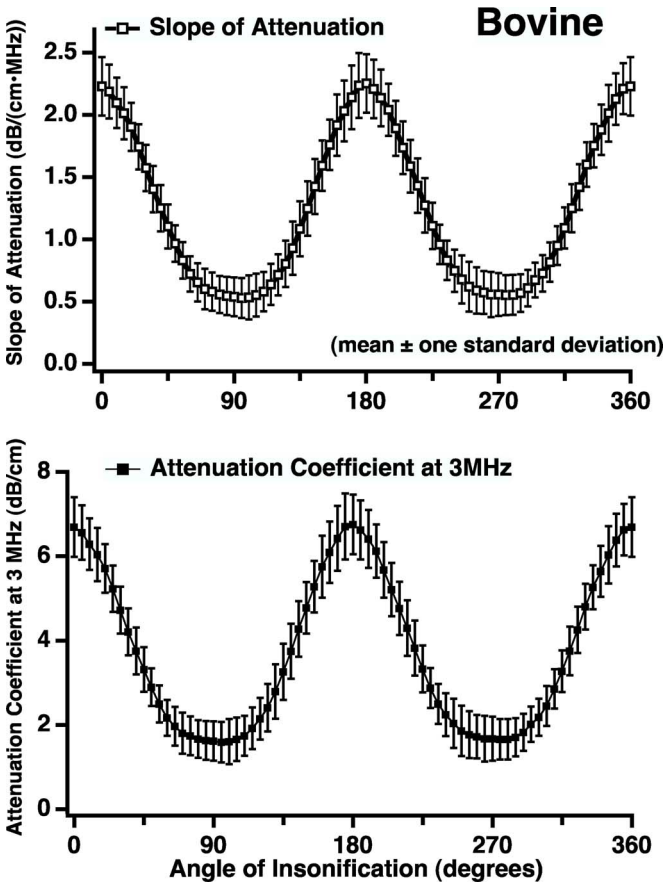


FIG. 4. Slope of attenuation [in dB/(cm MHz)] in freshly excised bovine myocardium (mean±one standard deviation) (upper panel). Corresponding attenuation coefficient at 3 MHz (dB/cm) (lower panel).

lamb and 11 for cow), and from between the two papillary muscles (12 for cow). Septal regional averages include specimens taken approximately halfway between the plane of the left ventricular specimens and the apex (12 for lamb and 11 for cow).

For angles of insonification near perpendicular to the myofiber direction, there is good agreement between ovine and bovine myocardial results, as seen in Fig. 6 (top panel). Values of attenuation are less near parallel insonification for lamb specimens in which the circumferential curvature of the myofibers is expected to be larger than for the more nearly unidirectional myofibers of the cow specimens. Uncertainties (shown in Fig. 6 as ±one standard deviation) are in general larger for angles near parallel insonification than for those near perpendicular insonification, and larger for bovine than for ovine results. The mean slope of attenuation in bovine anterior papillary muscles is shown in Fig. 6 (bottom panel) for comparison. Good agreement between papillary results and those from left ventricular and septal bovine specimens is seen at perpendicular insonification where papillary myofibers are expected to be aligned primarily perpendicular to the direction of insonification. Uncertainties become relatively large for slopes of attenuation in papillary muscles at angles near parallel insonification.

IV. DISCUSSION

Results of this study of freshly excised myocardium indicate a larger anisotropy of attenuation than previously reported. The ratio of the mean slope of attenuation for myocardium with myofibers oriented at parallel insonification to that for perpendicular insonification was measured to be 3.6 in ovine myocardium and 4.0 in bovine myocardium. Maxima appear more sharply peaked compared to minima (as can more clearly be seen by viewing Figs. 3 and 4 upside down). This suggests a greater sensitivity of measured slope of attenuation for slight angular deviations from insonification parallel than from insonification perpendicular to the myofibers. Table I shows that in all cases, for both current measurements and previous ones, uncertainties are larger for parallel insonification than for perpendicular. This behavior might be expected as a consequence of slight deviations in specimen fiber angle and measurement alignment as examined below.

In measurements of anisotropy in a specimen exhibiting uniaxial symmetry (unidirectional fiber alignment), the specimen is ideally prepared such that fibers are aligned relative to a rotational axis that is perpendicular to the fiber direction. In this way, the anisotropy can be measured by rotating the specimen about a measurement angle θ. Hence the direction of insonification ranges from rigorously perpendicular to rigorously parallel to the direction of the fibers. This ideal situation is depicted in the spherical coordinate system of Fig. 7(a) where the polar angle $\phi=90°$ and the angle between the direction of insonification and the direction of the myofibers Γ ranges between 0° and 90° as the

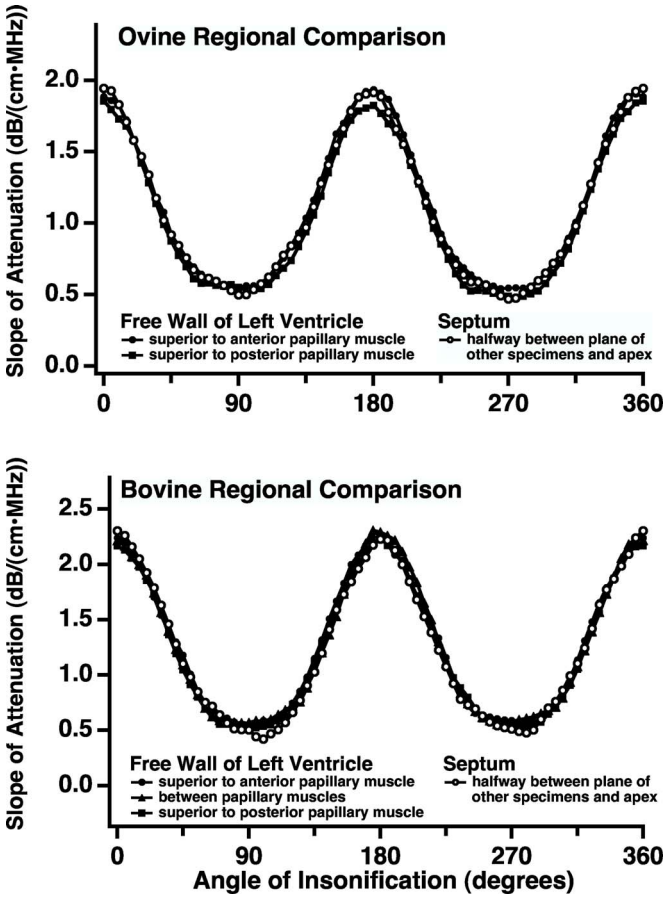


FIG. 5. Comparison of results from different regions of the left ventricle and septum for ovine (upper panel) and bovine (lower panel) myocardial specimens. Left ventricular free wall regional averages include specimens taken from just superior to the anterior papillary muscle and just superior to the posterior papillary muscle (and additionally for bovine myocardium from between the two papillary muscles). Septal regional averages include specimens taken approximately halfway between the plane of the left ventricular specimens and the apex.

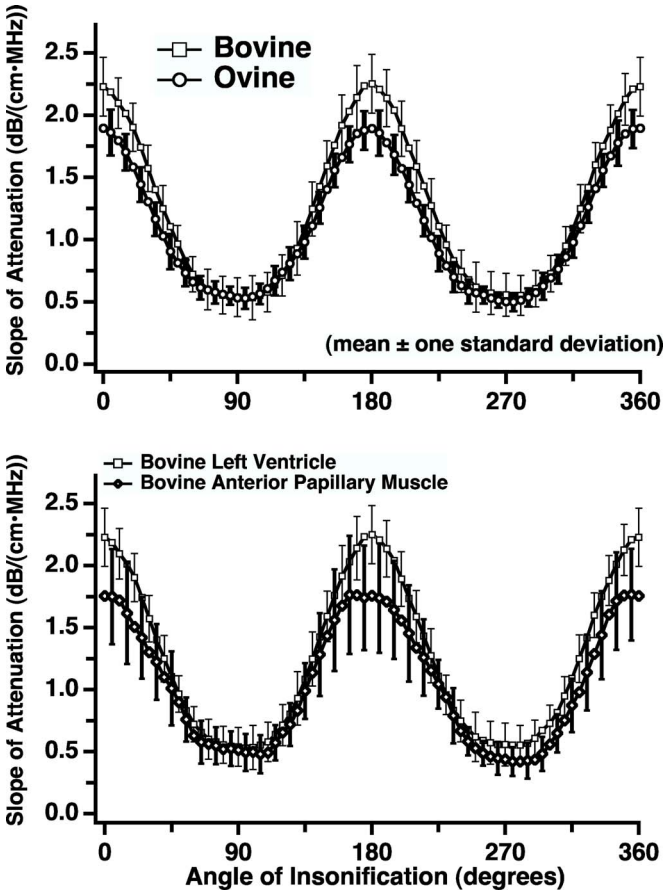


FIG. 6. Comparison of slopes of attenuation for cow and lamb myocardium (mean±one standard deviation) (upper panel). Comparison of average results for bovine anterior papillary muscles with average left ventricular results (lower panel).

measurement angle θ is varied from 0° to 360° during the course of measurement (a result of the symmetry of the uniaxial specimen). The angle Γ is computed according to the equation (Mottley and Miller, 1988)

$$\Gamma = \cos^{-1}[|\cos(\theta)|\sin(\phi)]. \quad (7)$$

In practice, however, because of the complex three-dimensional fiber architecture of the heart and the possibility for slight angular deviations in specimen preparation and measurement, small deviations from the ideal situation are likely. Because acoustic properties are determined by the angle between the direction of insonification and the direction of the myofibers, it is important to examine how this angle Γ differs from the measurement angle θ in the presence of a polar component ϕ as depicted in Fig. 7(b). This makes it possible to know better how to separate the actual angular dependence of ultrasonic properties of the specimen (reflected in changing angle Γ) from the effects of experimental angular deviations in individual fiber orientation or measurement alignment as the specimen is rotated by a measurement angle θ when interpreting measured data.

The upper right-hand panel of Fig. 7(b) shows how deviations of the polar angle ϕ from 90° affects the actual angle Γ between the direction of insonification and the direction of the myofibers. Figure 7(c) shows how the actual angle Γ differs from the measurement angle θ (the ideal case) as the polar angle is progressively altered away from perpendicular. At a measurement angle of θ=90° or 270°, the angle Γ=90° regardless of the value of the polar angle ϕ. However, as the measurement angle θ approaches parallel, larger deviations of the polar angle ϕ from perpendicular result in increasingly higher values of Γ (larger deviations from the ideal value of Γ=0° when ϕ=90°). Attenuation has been observed to vary systematically as a function of the angle between the direction of insonification and the predominant myofiber orientation, with a minimum perpendicular and a maximum parallel to the direction of the myofibers. Thus, because the measured attenuation actually depends on the angle Γ, any polar component (resulting from slight specimen misalignment or deviations of the myofibers from perfectly parallel) as the measurement angle θ is incremented should result in a progressively larger decrease from the actual values (in the ideal measurement case) as the measurement angle θ approaches parallel insonification. Based on these geometric considerations alone, this might be expected to lead to slightly lower and broader peaks in the measured anisotropy of attenuation, and reported results may represent lower limits of the magnitude and angular dependence for angles of insonification near parallel. Furthermore, as a con-

TABLE I. Comparison of measured ultrasonic myocardial slopes of attenuation [in dB/(cm MHz)] of the current study with previous studies (mean±one standard deviation).

	Parallel to myofibers	Perpendicular to myofibers	Difference between parallel and perpendicular	Ratio of mean parallel to mean perpendicular
Results of the current study				
Freshly excised bovine myocardium	2.22±0.23	0.55±0.16	1.67±0.28	4.0
Freshly excised ovine myocardium	1.88±0.17	0.52±0.07	1.36±0.18	3.6
Results of previous studies				
Freshly excised canine myocardium (Mottley and Miller 1990) (*with tomographic methods of data reduction*)	1.20±0.20	0.55±0.10	0.65±0.22	2.2
Freshly excised canine myocardium (Klepper *et al.* 1981 and Brandenburger *et al.* 1981) (*with tomographic methods of data reduction*)	1.65±0.17	0.63±0.15	1.02±0.23	2.6
Freshly excised canine myocardium (O'Donnell *et al.* 1977) (*comparable regions, superior to papillary muscles*)	—	0.56±0.08	—	—

sequence of slight experimental variations that could lead to slightly varying polar components ϕ from specimen to specimen, the behavior of the angle Γ described above would be expected to lead to larger uncertainties in measured attenuation for measurement angles θ near parallel than near perpendicular. This behavior is consistent with current as well as previous measurements of myocardial anisotropy in Table I.

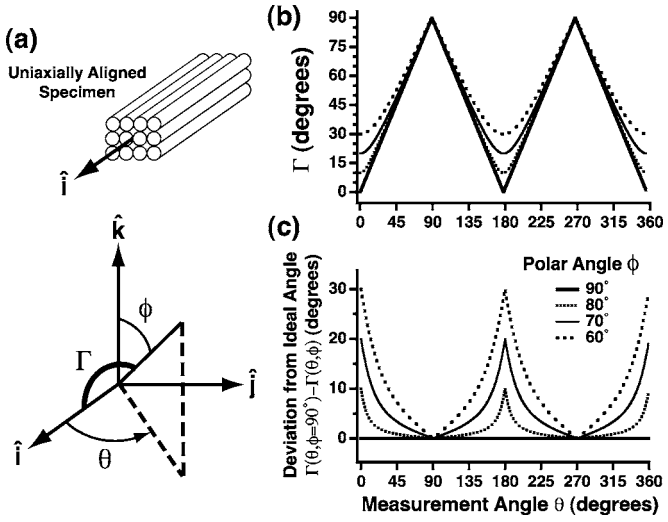


FIG. 7. (a) Spherical coordinate system in which θ is the measurement angle, ϕ is the polar component that may result from slight deviations in specimen fiber angle and measurement alignment, and Γ is the actual angle between the direction of insonification and the direction of the myofibers. (b) Γ as a function of the measurement angle θ for deviations of the polar angle ϕ from 90°. (c) Difference between the actual angle Γ in the ideal case of polar angle ϕ=90° and that in the presence of specific polar components.

A. Comparison of results to previous studies

When comparing results from different sets of specimens or studies, it is of interest to observe from Fig. 7(c) that the difference in $\Gamma(\theta, \phi=90°)$, the ideal measurement of anisotropy and $\Gamma(\theta, \phi)$, the measurement in practice, becomes less sensitive to small deviations of the measurement angle θ from perpendicular as the polar angle ϕ approaches perpendicular. Furthermore, Γ becomes identical to θ for perpendicular insonification regardless of the value of the polar component ϕ. Thus, based on geometric considerations alone, better agreement might be expected among different studies with slightly different measurement configurations and specimen preparations for angles near perpendicular than for angles near parallel. As seen in Table I, values for freshly excised ovine and bovine myocardium reported in this study show very good agreement with previous measurements of freshly excised canine myocardium for perpendicular insonification (Brandenburger *et al.*, 1981; Klepper *et al.*, 1981; Mottley and Miller, 1990; O'Donnell *et al.*, 1977). The measurements by O'Donnell, *et al.* (1977) were performed using phase sensitive piezoelectric transducers like those employed in this study, whereas those of Klepper *et al.* (1981), Brandenburger *et al.* (1981), and Mottley and Miller (1990) were performed using a phase insensitive acoustoelectric receiving transducer.

At parallel insonification, Klepper *et al.* (1981) and Brandenburger *et al.* (1981) reported a slope of attenuation of 1.65±0.17 dB/(cm MHz). This value is slightly smaller, but in agreement with the mean value of 1.88±0.17 dB/(cm MHz) measured in the current study in

similarly sized ovine hearts (within one standard deviation). In the previous study, tomographic methods of data reconstruction were employed that required a set of measurements sampled perpendicular to the beam direction at small intervals in a cross-sectional plane across the entire myocardial specimen for each angle of rotation (Brandenburger *et al.*, 1981; Klepper *et al.*, 1981). It has been shown that in an object such as a section of myocardium that can be modeled as a superposition of anisotropic media whose anisotropy axes are parallel to a plane in space (a superposition of anisotropic media with varying polar components ϕ), then the ensemble will exhibit an anisotropic angular variation whose functional dependence is identical to that of the constituent media, but whose anisotropy magnitude is bounded by (less than or equal to) the anisotropy magnitude exhibited by the individual components of the media (Brandenburger *et al.*, 1981). Thus, in order to measure the magnitude of anisotropy of the individual components, in this case the myofibers, it would need to be possible to obtain specimens in which individual myofibers are aligned uniformly along a single direction and such that they can be measured over a full range of angles from strictly parallel to strictly perpendicular to the direction of insonification. Any deviation from achieving strictly parallel insonification would result in a lower measured maximum value of attenuation. The authors state that "Because the muscle fibers in the heart do not all lie in the plane in which the measurements were made, the 2.6 to 1 anisotropy may reflect a lower bound on the value that would be observed if all fibers were confined to the plane of the measurement" (Brandenburger *et al.*, 1981).

Results for parallel insonification reported by Mottley and Miller (1990) of 1.20 ± 0.20 dB/(cm MHz) are significantly smaller than values reported in the other studies. This behavior may be a consequence of a difference in the way that the canine specimens were cored in that investigation. In the present study, specimens were cored from endocardium to epicardium such that the axis of rotation of the cylindrical specimens was oriented orthogonal to the epicardial surface of the heart (and consequently, orthogonal to the midmyocardial myofibers) as depicted in the upper and lower panels of Fig. 2. This approach permits a straightforward way to minimize the potential for off-angle coring of specimens relative to the direction of the midmyocardial myofibers and allows for insonification at all angles of rotation to be strictly in the midmyocardial region where myofibers are circumferential (Fig. 2, lower panel). Specimens were cored in the earlier study by Mottley and Miller (1990) "such that their axes were parallel to the long axis of the heart" (normal to the short axis plane of the heart shown in Fig. 2 (upper panel)) and "took in all of the heart wall from endo- to epicardium." Any deviation of the myofibers from perfectly parallel to the direction of insonification would cause the actual angle Γ to differ from the measurement angle θ for angles approaching parallel and would be expected to decrease the observed attenuation. Furthermore, based on considerations from previous histologic studies (Greenbaum *et al.*, 1981; Harrington *et al.*, 2005; LeGrice *et al.*, 1995; Streeter *et al.*, 1969), such a coring included some unavoidable transmural portions approaching subendocardium and subepicardium at the edges of the cylindrical specimens where myofibers are oriented progressively into and out of the short axis plane of Fig. 2 (upper panel) (Mottley and Miller, 1990). The Mottley and Miller (1990) study also employed transmission tomographic techniques to eliminate the need for separate specimen thickness measurements by averaging a series of ultrasonic scan lines acquired laterally across the entire myocardial specimen for each angle of insonification (Mottley and Miller, 1990). For angles of insonification near parallel to the circumferential myofibers of the midmyocardium, such measurements would also include in their averages scan lines near the subendocardial and subepicardial lateral edges of the specimen which may lower the average measured attenuation. This effect would not be expected to occur at perpendicular angles of insonification, because when the sound propagates perpendicular to the circumferential direction, it is primarily perpendicular to all of the myofibers (both subendocardial and subepicardial as well as midmyocardial myofibers). These trends in expected differences are consistent with the differences between the measurements of Mottley and Miller and the results of the current study.

B. Potential effects of inherent curvature of myofibers in myocardial specimens

The difference in attenuation between ovine and bovine myocardial specimens is maximum at parallel insonification. As the angle approaches perpendicular insonification, the agreement between lamb and cow results improves, until they are in complete agreement at angles near perpendicular [see Fig. 6 (upper panel)]. These characteristics of the measurements may result, in part, from increased circumferential myofiber curvature over the cylindrical specimens of the lamb hearts. As seen in Fig. 2 (lower panel), such deviation between lamb and cow results might be expected to be largest at angles near parallel insonification and least at angles near perpendicular where insonification is primarily perpendicular to the myofibers. The effect of myofiber curvature may be to introduce a net polar component ϕ as the measurement angle θ is varied. One approach for estimating the expected difference between ovine and bovine results at parallel insonification is as follows. The expected decrease in attenuation at parallel incidence might be inferred from results of this study by examining the measured angular dependence of the attenuation upon rotation of the bovine myocardial specimens about measurement angle θ (assuming the curvature in that case does not have a significant effect), shown in Fig. 4 (upper panel). As the insonifying beam passes through the specimen, the angle between it and the local tangent to the myofibers ϕ encounters a range of values from the maximum at the near edge of the specimen, to zero at the center of the specimen, and again to the maximum at the far edge of the specimen. The expected value of slope of attenuation might be estimated by averaging the measured bovine values in Fig. 4 (upper panel) corresponding to the range of angles ϕ encountered. For ovine specimens, values of ϕ are estimated to range between $\pm(\phi_{\text{maximum}} \approx 23^\circ)$ about parallel insonification. Based on these considerations, the corresponding average would lead to an expected slope of

attenuation of approximately 2.0 dB/(cm MHz). This value is slightly larger than, but within the uncertainties of the average measured value for lamb myocardium at parallel insonification [1.88±0.17 dB/(cm MHz)] as seen in Fig. 6 (upper panel). Although myofibers of bovine specimens are locally more nearly parallel than those of ovine specimens, we note that measurement uncertainties are larger for bovine than for ovine specimens, as can be seen in the upper panel of Fig. 6. Although the same diameter coring tool was used for preparing both the ovine and bovine tissue, the larger epicardial to endocardial thickness of the bovine specimens resulted in a tendency for them to hang in such a way that they were more elliptical in cross section with an average diameter less than that for the more circular ovine specimens. The increased variability of the bovine data may result from the slightly more irregular shape of the bovine specimens that might have led to slightly larger specimen to specimen variation in myofiber angle. These results demonstrate the potential value of taking into account myofiber curvature and angular shifts within a region-of-interest when making attenuation measurements. Factors to consider include heart size as well as region-of-interest size and local fiber orientation within the heart.

C. Considerations regarding papillary muscle results

Good agreement between results of attenuation in papillary muscle specimens compared to those from the left ventricle is seen in Fig. 6 (lower panel) for perpendicular insonification. For reasons indicated in Sec. II, however, specimens oriented at angles near parallel insonification may involve myofibers that are off angle in elevation and have correspondingly larger deviations of the polar angle from 90°. Consequently, measured peaks in attenuation may represent significantly lower limits of expected attenuation in papillary muscles at these angles. The difficulty of coring specimens in which myofibers are approximately unidirectional may be further indicated by the relatively larger standard deviations of papillary muscle results near parallel insonification compared with those near perpendicular insonification and compared with corresponding left ventricle results at parallel insonification.

V. SUMMARY

The degree of anisotropy of attenuation in freshly excised myocardium suggests the potentially significant influence of attenuation on ultrasonic imaging and tissue characterization measurements *in vivo*. The attenuation coefficient was found to increase as a function of frequency in an approximately linear manner and to increase monotonically as a function of angle of insonification from a minimum perpendicular to a maximum parallel relative to the direction of the myofibers. The mean slope of the attenuation coefficient with respect to frequency was three to four times larger for propagation parallel than for propagation perpendicular to the myofibers. Measured attenuation values are less sensitive to slight angular shifts near perpendicular insonification than near parallel. Detailed quantitative knowledge of the anisotropic behavior of attenuation with angle of insonification of the myofibers may be valuable in understanding the mechanisms of observed myocardial attenuation *in vivo*.

ACKNOWLEDGMENTS

This work was supported, in part, by NIH R37 HL40302. The authors gratefully acknowledge Star Packing Company, St. Louis, MO for providing ovine and bovine myocardial tissues.

Aygen, M., and Popp, R. L. (**1987**). "Influence of the orientation of myocardial fibers on echocardiographic images," Am. J. Cardiol. **60**, 147–152.

Baldwin, S. L., Holland, M. R., Sosnovik, D. E., and Miller, J. G. (**2005**). "Effects of region-of-interest length on estimates of myocardial ultrasonic attenuation and backscatter," Med. Phys. **32**, 418–426.

Baldwin, S. L., Marutyan, K. R., Yang, M., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2005**). "Estimating myocardial attenuation properties from m-mode ultrasonic backscatter," Ultrasound Med. Biol. **31**, 477–484.

Baldwin, S. L., Yang, M., Marutyan, K. R., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2004**). "The influence of bright intramural echoes on estimates of ultrasonic attenuation from backscattered ultrasound in excised myocardium," Ultrason. Imaging **26**, 233–249.

Baldwin, S. L., Yang, M., Marutyan, K. R., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2005**). "Measurements of the anisotropy of ultrasonic velocity in freshly excised and formalin-fixed myocardial tissue," J. Acoust. Soc. Am. **118**, 505–513.

Barzilai, B., Madaras, E. I., Sobel, B. E., Miller, J. G., and Perez, J. E. (**1984**). "Effects of myocardial contraction on ultrasonic backscatter before and after ischemia," Am. J. Physiol. Heart Circ. Physiol. **16 247**, H478–H483.

Brandenburger, G. H., Klepper, J. R., Miller, J. G., and Snyder, D. L. (**1981**). "Effects of anisotropy in the ultrasonic attenuation of tissue on computed tomography," Ultrason. Imaging **3**, 113–143.

Busse, L. J., and Miller, J. G. (**1981a**). "Detection of spatially nonuniform ultrasonic radiation with phase sensitive (piezoelectric) and phase insensitive (acoustoelectric) receivers," J. Acoust. Soc. Am. **70**, 1377–1386.

Busse, L. J., and Miller, J. G. (**1981b**). "Response characteristics of a finite aperture, phase insensitive ultrasonic receiver based upon the acoustoelectric effect," J. Acoust. Soc. Am. **70**, 1370–1376.

Castaldo, M., Funaro, S., Veneroso, G., and Agati, L. (**2000**). "Detection of residual tissue viability within the infarct zone in patients with acute myocardial infarction: Ultrasonic integrated backscatter analysis versus dobutamine stress echocardiography," J. Am. Soc. Echocardiogr **13**, 358–367.

Ciliberto, G. R., Pingitore, A., Mangiavacchi, M., Alberti, A., Paterni, M., and Picano, E. (**1996**). "The clinical value of blunting of cyclic gray level variation for the detection of acute cardiac rejection: A two-dimensional, doppler, and videodensitometric ultrasound study," J. Am. Soc. Echocardiogr **9**, 306–313.

Colonna, P., Montisci, R., Galiuto, L., Meloni, L., and Iliceto, S. (**1999**). "Effects of acute myocardial ischemia on intramyocardial contraction heterogeneity: A study performed with ultrasound integrated backscatter during transesophageal atrial pacing," Circulation **100**, 1770–1776.

D'Hooge, J., Bijnens, B., Jamal, F., Pislaru, C., Pislaru, S., Thoen, J., Suetens, P., Van de Werf, F., Angermann, C., Rademakers, F. E., Herregods, M. C., and Sutherland, G. R. (**2000**). "High frame rate myocardial integrated backscatter. Does this change our understanding of this acoustic parameter?" Eur. J. Echocardiogr. **1**, 32–41.

Di Bello, V., Pedrinelli, R., Bertini, A., Giorgi, D., Talini, E., Dell'Omo, G., and Mariani, M. (**2001**). "Cyclic variation of the myocardial integrated backscatter signal in hypertensive cardiopathy: A preliminary study," Coron Artery Dis. **12**, 267–275.

Dwyer, C., Allen, P., and Buckin, V. (**2000**). "Temperature dependence of the ultrasonic parameters of bovine muscle: Effects of muscle anisotropy," *13th Conference of the European Colloid and Interface Society, Trinity College*, Vol. **115**, Dublin, Ireland (Springer-Verlag, Berlin), pp. 282–286.

Fei, D. Y., and Shung, K. K. (**1985**). "Ultrasonic backscatter from mammalian tissues," J. Acoust. Soc. Am. **78**, 871–876.

Fei, D. Y., Shung, K. K., and Wilson, T. M. (**1987**). "Ultrasonic backscatter from bovine tissues: Variation with pathology," J. Acoust. Soc. Am. **81**, 166–172.

Finch-Johnston, A. E., Gussak, H. M., Mobley, J., Holland, M. R., Petrovic,

O., Perez, J. E., and Miller, J. G. (**2000**). "Cyclic variation of integrated backscatter: Dependence of time delay on the echocardiographic view used and the myocardial segment analyzed," J. Am. Soc. Echocardiogr **13**, 9–17.

Goens, M. B., Karr, S. S., and Martin, G. R. (**1996**). "Cyclic variation of integrated ultrasound backscatter: Normal and abnormal myocardial patterns in children," J. Am. Soc. Echocardiogr **9**, 616–621.

Greenbaum, R. A., Ho, S. Y., Gibson, D. G., Becker, A. E., and Anderson, R. H. (**1981**). "Left ventricular fibre architecture in man," Br. Heart J. **45**, 248–263.

Hancock, J. E., Cooke, J. C., Chin, D. T., and Monaghan, M. J. (**2002**). "Determination of successful reperfusion after thrombolysis for acute myocardial infarction: A noninvasive method using ultrasonic tissue characterization that can be applied clinically," Circulation **105**, 157–161.

Harrington, K. B., Rodriguez, F., Cheng, A., Langer, F., Ashikaga, H., Daughters, G. T., Criscione, J. C., Ingels, N. B., and Miller, D. C. (**2005**). "Direct measurement of transmural laminar architecture in the anterolateral wall of the ovine left ventricle: New implications for wall thickening mechanics," Am. J. Physiol. Heart Circ. Physiol. **288**, H1324–H1330.

Holland, M. R., Wallace, K. D., and Miller, J. G. (**2004**). "Potential relationships among myocardial stiffness, the measured level of myocardial backscatter ("image brightness") and the magnitude of the systematic variation of backscatter (cyclic variation) over the heart cycle," J. Am. Soc. Echocardiogr **17**, 1131–1137.

Holland, M. R., Wilkenshoff, U. M., Finch-Johnston, A. E., Handley, S. M., Perez, J. E., and Miller, J. G. (**1998**). "Effects of myocardial fiber orientation in echocardiography: Quantitative measurements and computer simulation of the regional dependence of backscattered ultrasound in the parasternal short-axis view," J. Am. Soc. Echocardiogr **11**, 929–937.

Klepper, J. R., Brandenburger, G. H., Mimbs, J. W., Sobel, B. E., and Miller, J. G. (**1981**). "Application of phase-insensitive detection and frequency-dependent measurements to computed ultrasonic attenuation tomography," IEEE Trans. Biomed. Eng. **BME-28**, 186–201.

Kovacs, A., Courtois, M. R., Weinheimer, C. J., Posdamer, S. H., Wallace, K. D., Holland, M. R., and Miller, J. G. (**2004**). "Ultrasonic tissue characterization of the mouse myocardium: Successful *in vivo* cyclic variation measurements," J. Am. Soc. Echocardiogr **17**, 883–892.

LeGrice, I. J., Smaill, B. H., Chai, L. Z., Edgar, S. G., Gavin, J. B., and Hunter, P. J. (**1995**). "Laminar structure of the heart: Ventricular myocyte arrangement and connective tissue architecture in the dog," Am. J. Physiol. Heart Circ. Physiol. **38 269**, H571–H582.

Melton, H. E., Jr., and Skorton, D. J. (**1983**). "Rational gain compensation for attenuation in cardiac ultrasonography," Ultrason. Imaging **5**, 214–228.

Miller, J. G., Yuhas, D. E., Mimbs, J. W., Dierker, S. B., Busse, L. J., Laterra, J. J., Weiss, A. N., and Sobel, B. E. (**1976**). "Ultrasonic tissue characterization: Correlation between biochemical and ultrasonic indices of myocardial injury," *Proc. IEEE Ultrasonics Symposium*, Anapolis, 76 CH 1120–5SU, pp. 33–43.

Mimbs, J. W., O'Donnell, M., Bauwens, D., Miller, J. G., and Sobel, B. E. (**1980**). "The dependence of ultrasonic attenuation and backscatter on collagen content in dog and rabbit hearts," Circ. Res. **47**, 49–58.

Mottley, J. G., Glueck, R. M., Perez, J. E., Sobel, B. E., and Miller, J. G. (**1984**). "Regional differences in the cyclic variation of myocardial backscatter that parallel regional differences in contractile performance," J. Acoust. Soc. Am. **76**, 1617–1623.

Mottley, J. G., and Miller, J. G. (**1988**). "Anisotropy of the ultrasonic backscatter of myocardial tissue: I. Theory and measurements *in vitro*," J. Acoust. Soc. Am. **83**, 755–761.

Mottley, J. G., and Miller, J. G. (**1990**). "Anisotropy of the ultrasonic attenuation in soft tissues: Measurements *in vitro*," J. Acoust. Soc. Am. **88**, 1203–1210.

Naito, J., Masuyama, T., Tanouchi, J., Mano, T., Kondo, H., Yamamoto, K., Nagano, R., Hori, M., Inoue, M., and Kamada, T. (**1994**). "Analysis of transmural trend of myocardial integrated ultrasound backscatter for differentiation of hypertrophic cardiomyopathy and ventricular hypertrophy due to hypertension," J. Am. Coll. Cardiol. **24**, 517–524.

O'Donnell, M., Mimbs, J. W., Sobel, B. E., and Miller, J. G. (**1977**). "Ultrasonic attenuation in normal and ischemic myocardium," *Proc. Second International Symposium on Ultrasonic Tissue Characterization* (National Bureau of Standards Spec. Publ. 525, U. S. Government Printing Office, Washington DC, **1979**), pp. 63–71.

Ohara, Y., Hiasa, Y., Hosokawa, S., Suzuki, N., Takahashi, T., Kishi, K., and Ohtani, R. (**2005**). "Ultrasonic tissue characterization predicts left ventricular remodeling in patients with acute anterior myocardial infarction after primary coronary angioplasty," J. Am. Soc. Echocardiogr **18**, 638–643.

Pasquet, A., D'Hondt, A. M., Melin, J. A., and Vanoverschelde, J. L. (**1998**). "Relation of ultrasonic tissue characterization with integrated backscatter to contractile reserve in chronic left ventricular ischemic dysfunction," Am. J. Cardiol. **81**, 68–74.

Perez, J. E., Klein, S. C., Prater, D. M., Fraser, C. E., Cardona, H., Waggoner, A. D., Holland, M. R., Miller, J. G., and Sobel, B. E. (**1992**). "Automated, on-line quantification of left ventricular dimensions and function by echocardiography with backscatter imaging and lateral gain compensation," Am. J. Cardiol. **70**, 1200–1205.

Pislaru, C., D'Hooge, J., Pislaru, S. V., Brandt, E., Cipic, R., Angermann, C. E., Van de Werf, F. J., Bijnens, B., Herregods, M., and Sutherland, G. R. (**2001**). "Is there a change in myocardial nonlinearity during the cardiac cycle?" Ultrasound Med. Biol. **27**, 389–398.

Recchia, D., Miller, J. G., and Wickline, S. A. (**1993**). "Quantification of ultrasonic anisotropy in normal myocardium with lateral gain compensation of two-dimensional integrated backscatter images," Ultrasound Med. Biol. **19**, 497–505.

Rijsterborgh, H., Mastik, F., Lancee, C. T., Verdouw, P., Roelandt, J., and Bom, N. (**1993**). "Ultrasound myocardial integrated backscatter signal processing: Frequency domain versus time domain," Ultrasound Med. Biol. **19**, 211–219.

Sagar, K. B., Pelc, L. R., Rhyne, T. L., Komorowski, R. A., Wann, L. S., and Warltier, D. C. (**1990**). "Role of ultrasonic tissue characterization to distinguish reversible from irreversible myocardial injury," J. Am. Soc. Echocardiogr **3**, 471–477.

Sosnovik, D. E., Baldwin, S. L., Holland, M. R., and Miller, J. G. (**2001a**). "Transmural variation of myocardial attenuation and its potential effect on contrast-mediated estimates of regional myocardial perfusion," J. Am. Soc. Echocardiogr **14**, 782–788.

Sosnovik, D. E., Baldwin, S. L., Lewis, S. H., Holland, M. R., and Miller, J. G. (**2001b**). "Transmural variation of myocardial attenuation measured with a clinical imager," Ultrasound Med. Biol. **27**, 1643–1650.

Streeter, D. D., Jr., Spotnitz, H. M., Patel, D. P., Ross, J., Jr., and Sonnenblick, E. H. (**1969**). "Fiber orientation in the canine left ventricle during diastole and systole," Circ. Res. **24**, 339–347.

Tu, H., Varghese, T., Madsen, E. L., Chen, Q., and Zagzebski, J. A. (**2003**). "Ultrasound attenuation imaging using compound acquisition and processing," Ultrason. Imaging **25**, 245–261.

van der Steen, A. F., Rijsterborgh, H., Lancee, C. T., Mastik, F., Krams, R., Verdouw, P. D., Roelandt, J. R., and Bom, N. (**1997**). "Influence of data processing on cyclic variation of integrated backscatter and wall thickness in stunned porcine myocardium," Ultrasound Med. Biol. **23**, 405–414.

van der Steen, A. F., Rijsterborgh, H., Mastik, F., Lancee, C. T., van Hoorn, W. M., and Bom, N. (**1991**). "Influence of attenuation on measurements of ultrasonic myocardial integrated backscatter during cardiac cycle (an *in vitro* study)," Ultrasound Med. Biol. **17**, 869–877.

Vandenberg, B. F., Stuhlmuller, J. E., Rath, L., Kerber, R. E., Collins, S. M., Melton, H. E., and Skorton, D. J. (**1991**). "Diagnosis of recent myocardial infarction with quantitative backscatter imaging: Preliminary studies," J. Am. Soc. Echocardiogr **4**, 10–18.

Verdonk, E. D., Hoffmeister, B. K., Wickline, S. A., and Miller, J. G. (**1996**). "Anisotropy of the slope of ultrasonic attenuation in formalin-fixed human myocardium," J. Acoust. Soc. Am. **99**, 3837–3843.

Vered, Z., Mohr, G. A., Barzilai, B., Gessler, C. J., Jr., Wickline, S. A., Wear, K. A., Shoup, T. A., Weiss, A. N., Sobel, B. E., Miller, J. G., and Perez, J. E. (**1989**). "Ultrasound integrated backscatter tissue characterization of remote myocardial infarction in human subjects," J. Am. Coll. Cardiol. **13**, 84–91.

Wear, K. A., Milunski, M. R., Wickline, S. A., Perez, J. E., Sobel, B. E., and Miller, J. G. (**1989**). "Contraction-related variation in frequency dependence of acoustic properties of canine myocardium," J. Acoust. Soc. Am. **86**, 2067–2072.

Wickline, S. A., Thomas, L. J., III, Miller, J. G., Sobel, B. E., and Perez, J. E. (**1986**). "Sensitive detection of the effects of reperfusion on myocardium by ultrasonic tissue characterization with integrated backscatter," Circulation **74**, 389–400.

Yuda, S., Dart, J., Najos, O., and Marwick, T. H. (**2002**). "Use of cyclic variation of integrated backscatter to assess contractile reserve and myocardial viability in chronic ischemic left ventricular dysfunction," Echocardiogr. **19**, 279–287.

Using self-organizing maps to recognize acoustic units associated with information content in animal vocalizations

John Placer
Department of Computer Science, Northern Arizona University, Flagstaff, Arizona 86011

C. N. Slobodchikoff
Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011

Jason Burns, Jeffrey Placer, and Ryan Middleton
Department of Computer Science, Northern Arizona University, Flagstaff, Arizona 86011
and College of Engineering and Natural Sciences, Northern Arizona University, Flagstaff, Arizona 86011

(Received 27 September 2005; revised 26 January 2006; accepted 10 February 2006)

Kohonen self-organizing neural networks, also called self-organizing maps (SOMs), have been used successfully to recognize human phonemes and in this way to aid in human speech recognition. This paper describes how SOMS also can be used to associate specific information content with animal vocalizations. A SOM was used to identify acoustic units in Gunnison's prairie dog alarm calls that were vocalized in the presence of three different predator species. Some of these acoustic units and their combinations were found exclusively in the alarm calls associated with a particular predator species and were used to associate predator species information with individual alarm calls. This methodology allowed individual alarm calls to be classified by predator species with an average of 91% accuracy. Furthermore, the topological structure of the SOM used in these experiments provided additional insights about the acoustic units and their combinations that were used to classify the target alarm calls. An important benefit of the methodology developed in this paper is that it could be used to search for groups of sounds associated with information content for any animal whose vocalizations are composed of multiple simultaneous frequency components. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184227]

PACS number(s): 43.80.Ev, 43.80.Ka [JAS] Pages: 3140–3146

I. INTRODUCTION

Research on animal vocalizations that contain multiple simultaneous frequency components often focuses on the general, aggregate structure of those vocalizations. Frequency and time parameters are gathered and an entire vocalization is characterized as a collection of frequency-time attributes. In some studies these aggregate properties of sounds have been compared for overlap using spectrogram cross correlation (Clark *et al.*, 1987). Alternatively information has been extracted and analyzed statistically using multivariate methods such as Principal Components Analysis or Discriminant Function Analysis (Eakle *et al.*, 1989; Galeotti and Pavan, 1991; Slobodchikoff *et al.*, 1991; Galeotti *et al.*, 1993; Blackshaw *et al.*, 1996; Hill and Lill, 1998; Slobodchikoff *et al.*, 1991; Ackers and Slobodchikoff, 1999; Lengagne, 2001; Sousa-Lima *et al.*, 2002; Baker and Logue, 2003).

As an example of studies that focus on the general aggregate structure of vocalizations consider Fig. 1, which shows how sound spectrograms can be used to visually distinguish between different types of alarm calls vocalized by Gunnison's prairie dogs. The shape of the spectrogram on the left is typical of dog-elicited alarm calls and the shape of the spectrogram on the right is typical of coyote-elicited alarm calls. In past studies analyses related to the general structure of alarm calls have been performed on data by measuring various macroscopic properties of collections of similar sound spectrograms of alarm calls. Properties such as ascending slope, descending slope, dominant frequency, and subdominant frequency were measured and the data was then subjected to discriminant function analysis (Slobodchikoff *et al.*, 1991). Note that the sound spectrograms in Fig. 1 were created from the same database that provided the data for the experiments discussed in this paper.

In the types of studies referred to above, the macroscopic structure of vocalizations is examined while the detailed internal structure, the acoustic microstructure, of the vocalizations is left unexamined and unappreciated. One striking illustration of this is the case of repetitive vocalizations of animal signals; it is generally assumed that repeated vocalizations from a given animal convey the same information even when the frequency-time parameters of their sound spectrograms vary. However, alarm calls are an example of repetitive vocalizations whose internal variations might be highly significant. Alarm calls are often given as a series or *bout* of repetitive elements, as in the case of alarm calls vocalized by Gunnison's prairie dog (*Cynomys gunnisoni*) in the presence of predators (Slobodchikoff *et al.*, 1991). This paper presents a methodology for investigating the possibility that the variations in the acoustic properties of animal alarm calls are highly significant because these variations indicate the existence of distinct acoustic structures that encode information.

A link between acoustic microstructures and information encoding in alarm calls has been suggested by earlier studies

 0001-4966/2006/119(5)/3140/7/$22.50

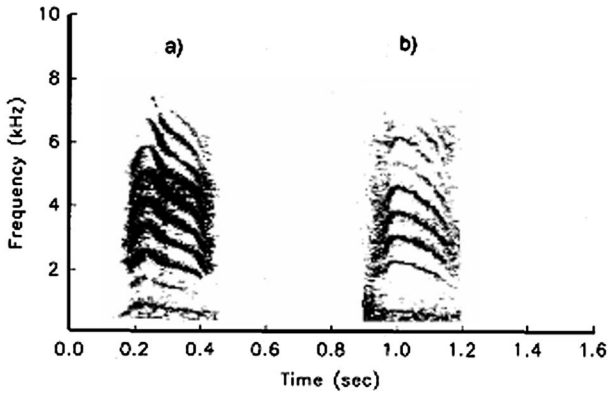


FIG. 1. Sound spectrograms showing a typical dog-elicited alarm call on the left-hand side and a typical coyote-elicited alarm call on the right-hand side.

of Gunnison's prairie dog vocalizations. In two studies (Placer and Slobodchikoff, 2000, 2001), composites of frequency ratio information taken from small partitions of Gunnison's prairie dog alarm calls were used to classify those calls according to the species of predator that was present when the calls were vocalized. These results suggested that information about predator species was encoded in the acoustic microstructure of the alarm calls examined. However, the composite frequency ratio distributions did not themselves reveal the details of the acoustic structure of individual alarm call partitions nor did they provide any clues about how these small-scale acoustics might be organized to encode information in the macroscopic structure of an alarm call.

In a later study (Placer and Slobodchikoff, 2004), Gunnison's prairie dog alarm calls were decomposed into numerous equal-sized partitions and the four most energetic frequency components of each partition were extracted. The two lowest frequencies of the four were then used to identify the acoustic unit associated with any given partition. When the acoustic units created by this process were examined, it was discovered that there were unique acoustic units and combinations of acoustic units that occurred exclusively in vocalizations associated with specific predator species. These exclusive sounds were used to classify alarm calls according to the species of predator that was present when the calls were vocalized. However, the accuracy of the classifications of individual alarm calls were modest and varied significantly according to different criteria that were applied in the study. Nonetheless, the results of this research also suggested a potential link between acoustic microstructure and information encoding and they began to reveal some of the important features of the acoustic microstructure of alarm calls.

In the work described in this paper new, more accurate techniques are developed that are used to recognize small-scale acoustic structures in animal alarm calls that might be used to encode information. These techniques are based on the self-organizing map (SOM) algorithm of Kohonen (1998). A SOM or Kohonen map is an artificial neural network that classifies input data into a number of clusters that are based on the features of the input data itself. In other words, a SOM is not told what the categories of interest are; the SOM uses its input data to discover for itself what the relevant categories are. This is a particularly important attribute for the study discussed in this paper where the relevant categories are acoustic units whose structures or features are not yet well understood.

Self-organizing maps have been shown to be effective in the automatic recognition of human speech. Kohonen used a SOM to create a recognition system for human speech that could detect phonemes from a continuous speech signal (Kohonen, 1988). This "phonetic typewriter," as Kohonen called it, was used to recognize both Finnish and Japanese phonemes. Since then SOMs have been used for the phoneme-based recognition of a number of other languages including Chinese (Wu *et al.*, 1992) and Slovene (Mihelic *et al.*, 1992). SOMs have also been used as more general feature extractors of human speech that can utilize contextual information (Kangas *et al.*, 1992). In sender recognition experiments with six mouse lemur males, Zimmermon and Lerch used a high performance Kohonen map trained for continuous human speech recognition. The success rate in recognizing these six individual males varied from 55.6% to 100% (Zimmerman and Lerch, 1993).

This paper describes how SOMS can be used to characterize and identify classes of acoustic units that are associated with specific information content in animal vocalizations. A SOM was trained to identify clusters of acoustic units in Gunnison's prairie dog alarms where each cluster contained sounds with similar acoustic properties. Individual sounds belonging to specific clusters as well as combinations of these sounds were found to be associated exclusively with alarm calls vocalized in the presence of a specific predator species. These sounds and sound combinations were used with high accuracy to classify the targeted alarm calls according to the species of predator that was present when the alarm calls were vocalized. Furthermore, the topological structure of the output layer of the SOM was shown to reveal interesting characteristics of these sounds and their combinations. The details of these experiments are presented in the following sections of this paper.

II. MATERIALS AND METHODS

This study utilized digitized versions of Gunnison's prairie dog alarm calls that were recorded over a period of 10 years (1988–1997) at two separate prairie dog colonies. Uher Model No. CR160AV and Sony TC-D5PRO II cassette recorders and a Sennheiser ME-88 shotgun microphone were used to record the alarm calls. Additional details about this library of alarm calls can be found in Placer and Slobodchikoff (2004). A total of 300 alarm calls was selected from this database where 100 alarm calls were vocalized in the presence of dogs, 100 in the presence of coyotes, and 100 in the presence of hawks. A system of computer programs was created by one of the authors (J.P.) that performed all the data analysis presented in this paper. Some of these programs were implemented using MATLAB, a high-performance numeric computation software package, and other programs were implemented using the computer language Java.

The experiments discussed in this paper consisted of five distinct steps. First, each of the 300 alarm calls utilized in the

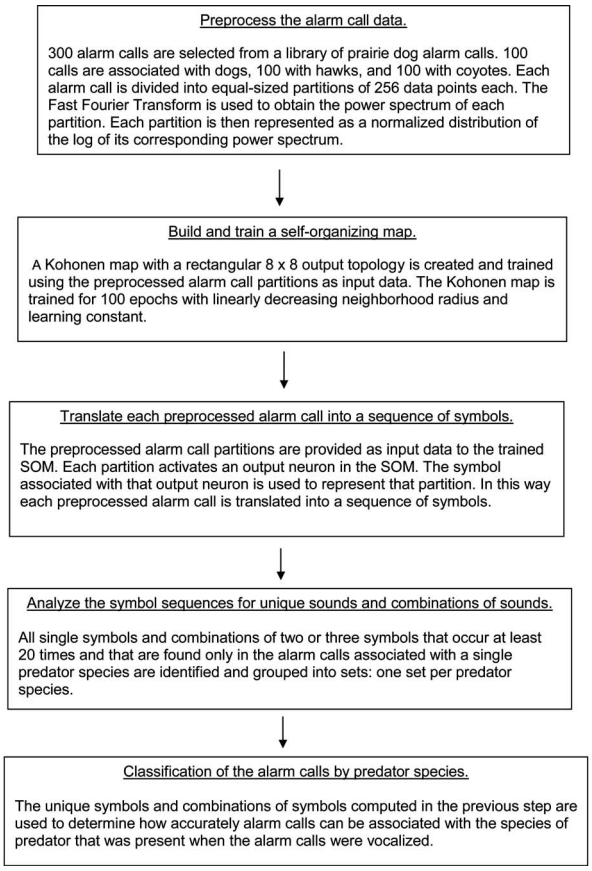


FIG. 2. The experimental protocol used to classify prairie dog alarm calls by the predator species that was present when the calls were vocalized.

experiments was preprocessed using the fast Fourier transform. Second, a self-organizing map was trained using the preprocessed data as input. Third, the trained SOM was used to translate the preprocessed alarm call data into sequences of symbols where each distinct symbol represented a distinct cluster of similar acoustic units. Fourth, the sequences of symbols were examined for single symbols and combinations of two or three symbols that were found exclusively in the alarm calls associated with a specific predator species. Fifth, these unique symbols were then used to classify the alarm calls according to the predator species present when the calls were vocalized. Furthermore, relationships among the unique symbols were examined. A graphical representation of these steps is given in Fig. 2. Each of these steps is described in detail in the sections provided below.

A. Preprocess the alarm call data

Each of the 300 alarm calls utilized in this study was preprocessed. Every alarm call was divided into equal-sized partitions of 256 data points each. The fast Fourier transform (FFT) was then applied to each partition's data points in order to compute a power spectrum of the frequencies contained in that partition. The alarm call data had been recorded at 44 100 samples per second and each partition contained 256 data points so the resolution of the FFT was just over 172 Hertz (44100/256). Since the upper limit of the frequencies that needed to be examined was only 9000 Hertz, each power spectrum contained 52 power values (the floor of 9000/172). A normalized distribution of the log of each power spectrum was then used to represent each partition. When the preprocessing just described was complete, each alarm call had been translated into a sequence of 25 vectors, where each vector contained 52 values.

B. Build and train a self-organizing map

A Kohonen map (SOM) was implemented for this experiment using the language Java. A SOM has two layers: an input layer and an output layer. Each neuron in the input layer is connected to each neuron in the output layer where each connection between an input neuron and an output neuron contains a weight. The input layer simply holds the values of the current input vector. The Kohonen map used in the experiments discussed in this paper had 52 input neurons, one neuron for each value in the normalized distribution of the log of the power spectrum of an alarm call partition. Since each input neuron was connected to each output neuron, every output neuron had 52 weights associated with it.

The output neurons of a SOM can be arranged in a number of different spatial or topological structures. The SOM used in these experiments had an output layer that consisted of a simple 8×8 rectangular grid of neurons. During training, when an input vector is provided to a SOM, the output neuron whose weight vector is closest to the value of the input vector (i.e., the winning neuron) has its weights adjusted so that they more closely resemble the values of the input vector. Furthermore, not only are the weights of the winning output neuron adjusted, but the neurons in a neighborhood surrounding that output neuron also have their weights adjusted to more closely resemble the input vector. The closeness between two vectors was determined by a simple Euclidean distance measure.

The neighborhood radius for the SOM used in this experiment began at 4 neurons and was decreased linearly as training progressed. The degree of adjustment of weights in a SOM is determined by a learning constant. When the weight vector on the winning output neuron is adjusted to align more closely with the current input vector, the learning constant limits the amount of this alignment. With a learning constant of 1.0, the weight vector would be moved so that it is exactly aligned with the input vector. A learning constant of 0.5 would cause the weight vector to be moved half of the distance to complete alignment. In this experiment the learning constant began with the value 0.7 and was decreased linearly to the value 0.05 as training progressed. Both the neighborhood radius and the learning constant reached their final values of 0 and 0.05, respectively, one-half of the way through the SOM training session. In the second half of the session, called the convergence phase, the neighborhood vanishes so that only the winning neuron itself is adjusted with each input. Furthermore, the learning constant is quite small so that the SOM slowly refines and adjusts the weights on its winning output neurons.

A training session of a SOM consists of repeatedly presenting the vectors in the input data set (the preprocessed alarm calls in our case) to the SOM and allowing the SOM to adjust the weights on its connections. The SOM used in this

TABLE I. The mapping of the Kohonen map's output neurons to symbol names. Each cell represents an output neuron and the text in each cell is the name of the symbol associated with that cell.

S1	S2	S3	S4	S5	S6	S7	S8
S9	S10	S11	S12	S13	S14	S15	S16
S17	S18	S19	S20	S21	S22	S23	S24
S25	S26	S27	S28	S28	S30	S31	S32
S33	S34	S35	S36	S37	S38	S39	S40
S41	S42	S43	S44	S45	S46	S47	S48
S49	S50	S51	S52	S53	S54	S55	S56
S57	S58	S59	S60	S61	S62	S63	S64

experiment was trained for 100 epochs. One epoch is completed when all of the vectors in the input data set have been presented to the SOM. One major result of training a SOM is that the weights of its output neurons approach values that represent groups or clusters of similar vectors in the input data set. In effect, each output neuron becomes an exemplar of a class of similar vectors in the input data set. Recall that each vector in the input data set used in this experiment represents the acoustic unit found in a partition of one of the 300 target alarm calls. Therefore, the output nodes of the SOM used in this experiment represented clusters of similar acoustic units found in the partitions of the alarm calls utilized for the experiment.

Another major result of training a SOM is found in the topology or spatial ordering of its output layer. Since neighborhoods of output neurons have their weights adjusted together, similar input patterns will be recognized by output neurons that are spatially close to each other in the output layer of the SOM. In the experiments discussed here, this means that similar acoustic units will be found in neurons that lie close to each other in the 8×8 output grid of the SOM.

C. Translate each preprocessed alarm call into a sequence of symbols

After the SOM was trained in this experiment, it was then used to translate each preprocessed alarm call into a sequence of symbols. To understand how this was done consider that the output layer of the SOM was an 8×8 matrix of neurons. Each neuron had a unique symbol associated with it according to its position in the matrix. The mapping of symbols to output neurons is shown in Table I. Each cell in the matrix in Table I represents an output neuron in the SOM used in these experiments. The text contained in any given cell is the name of the symbol associated with that cell. After training the SOM, the preprocessed partitions composing each alarm call were once again provided as input to the trained SOM. Each input vector (i.e., preprocessed partition) provided to the SOM was replaced by the symbol associated with the winning neuron produced by that input vector; thus, each partition in each alarm call was replaced by a specific symbol. In this way, each preprocessed alarm call was translated into a sequence of symbols. For example, the sequence of symbols

S9S46S40S40S23S35S35S35S42S35S23S35S42S35S35S35S8S40

was produced when the trained SOM was used to translate one of the preprocessed alarm calls that had been vocalized in the presence of a dog.

D. Analyze the symbol sequences for unique sounds and combinations of sounds

After this translation process in which each preprocessed alarm call was expressed as a sequence of symbols, the symbols contained in the translated alarm calls were analyzed in order to identify any single symbols (monograms), neighboring pairs of symbols (digrams) and sequences of three symbols (trigrams) that were found only in the alarm calls associated with a particular species of predator. During this part of the analysis only monograms, digrams, and trigrams that occurred at least 20 times in the target alarm calls were examined. This precaution was meant to guard against the use of acoustic units that rarely occurred and that might be artifacts of the arbitrary way in which alarm calls were partitioned. Three sets of symbols were constructed: one for monograms, digrams, and trigrams that occurred only in alarm calls associated with coyotes, another set for monograms, digrams, and trigrams that occurred only in alarm calls associated with dogs, and a third set for monograms, digrams, and trigrams that occurred only in alarm calls associated with hawks.

E. Classification of the alarm calls by predator species

In principle, any alarm call that could be shown to contain a monogram, digram, or trigram that was uniquely associated with a given predator species could itself be unambiguously associated with that same predator species. Thus, the three sets of unique monograms, digrams, and trigrams described above were used to classify the alarm calls by predator species. The percentage of alarm calls that could be identified unambiguously in this way for a given predator species was considered the percentage of correctly identified alarm calls for that predator species.

III. RESULTS

Kohonen maps are initialized with random connection weights before training. One consequence of this is that the final connection weights that result from any given training session will vary. This, in turn, means that the results of translating alarm calls into symbol sequences and the associated classification accuracies will vary among a number of separately trained SOMs. Therefore, in order to determine average classification accuracies, 10 experimental runs, each performed with a separately trained Kohonen map (SOM), were executed. The data resulting from these 10 experimental runs are given in Table II.

Each row in Table II except the last gives the data associated with a separate experimental run. The last row provides the means of all the data generated in the 10 experiments with standard deviation values given in parentheses. The first value of each row simply gives an identifying num-

TABLE II. Statistics for the alarm calls correctly classified by predator species with the standard deviations for the final averages given as parenthesized values.

Run #	Dog %Correct	Hawk %Correct	Coyote %Correct	Average %Correct
1	82	100	91	91
2	90	100	88	92.67
3	75	100	91	88.67
4	91	100	95	95.33
5	95	100	91	95.33
6	81	100	90	90.33
7	92	100	83	91.67
8	82	100	88	90.00
9	92	100	89	93.67
10	81	100	83	88.00
Average %Correct	86.1 (6.64)	100 (0)	88.9 (3.70)	91.67 (2.57)

ber for that experimental run. The second, third, and fourth values give the percentage of alarm calls classified correctly for the predator species dog, hawk, and coyote, respectively. The last value provides the average correct classification for that run across all species of predators.

Table I, which maps symbols to output neurons, is expanded in Table III to include information about the monograms, digrams, and trigrams that were found to be unique to a particular predator species. The name in bold at the top center of each cell (output neuron) is the name of the symbol associated with that cell. If that symbol is found at the start of one or more monograms, digrams or trigrams found exclusively in alarm calls associated with only one predator species, then all of those monograms, digrams, and trigrams are listed below the symbol. Cells that contain such lists also specify the relevant predator species in parentheses next to the cell's symbol name at the top of the cell. Symbols associated with cells that contain no lists below their symbol name simply do not occur at the beginning of a monogram, digram or trigram that is unique to a particular predator species.

To better explain Table III some examples are needed. Consider the cell named **S24**. This cell contains a list with only one digram (*S24S21*), which begins with symbol *S24*. This digram is found only in the alarm calls associated with the species coyote; the species name is shown in parentheses next to the cell's symbol name. Now consider cell **S62**. The list in this cell contains two digrams, *S62S62* and *S62S37*, both of which begin with symbol *S62*. The digram *S62S62* indicates that the sound associated with symbol *S62* is held across two partitions of an alarm call. Notice again that the predator species coyote is specified in parentheses next to the cell's symbol name. Finally, consider cell **S4**. The predator species specified in this cell is dog and the list provided in the cell simply contains *S4*. This means that the monogram *S4* occurs only in alarm calls associated with dogs; therefore, no digrams or trigrams associated exclusively with dogs begin with the symbol *S4*.

Other types of information can also be gained from Table III. For example, notice that the cells associated with dogs and coyotes each tend to have short lists of unique monograms, digrams, and trigrams. However, the cells associated with hawks tend to contain long lists. A closer look at these long lists reveals that combinations of a small group of symbols (S35, S51, S53, S58, and S60) are used to create most of the digrams and trigrams in all of the lists associated with hawks. These same symbols are used in differing orders and combinations throughout the lists.

IV. DISCUSSION

Kohonen maps were used to associate individual alarm calls with specific predator species with average accuracies of greater than 91%; these results demonstrate that Kohonen maps are useful tools for exploring and identifying sounds associated with specific information content in animal vocalizations. Furthermore, analysis of the digrams and trigrams specific to a particular predator species revealed certain types of symbol patterns associated with specific predator species. For example, in Table III, digrams and trigrams associated with hawks were regrouped into new lists where the same two symbols were used in each digram or trigram. This regrouping revealed how particular symbols can occur in different orders or be expressed for varying lengths of time or both. As an example, consider the list below that is composed of digrams and trigrams associated only with hawks, where each digram and trigram contains only the symbols S53 and S60:

S53S60, S53S53S60, S60S53, S60S53S53.

The symbol sequences in the list above indicate that the symbol S53 can occur either before or after the symbol S60. In other words, the relative order of the two sounds is not fixed. Furthermore, recall that S53 represents a sound (acoustic unit) expressed in a single partition of an alarm call whereas S53S53 represents the sound S53 expressed across two partitions of an alarm call. Thus, in the list above the sound represented by symbol S53 can be expressed for a single partition of time (S53S60, S60S53) or it can be held across two partitions of time (S53S53S60, S60S53S53). Similar kinds of relationships can be found in digrams and trigrams containing different pairs of symbols. Whether these patterns of symbols represent a first glimpse of lexical structures or a first glimpse of syntactic structures in prairie dog alarm calls is yet to be determined. However, the patterns are clearly evident and they are seen to occur in vocalizations restricted to a specific predator species.

It should also be mentioned that the variety of symbols (sounds) associated with specific predator species in Table III is probably less diverse than it might seem at first. Symbols that lie close to each other in the Kohonen map output layer and that are unique to the same predator species might well represent slight variations of the same basic sound. For example, consider cells **S25** and **S26**, which list digrams unique to alarm calls associated with dogs. These cells are located next to each other and the digram in each cell only contains the symbol named by that cell. Thus, it is likely that symbols *S25* and *S26* refer to the same basic sound expressed in slightly different ways. In human speech, where the pho-

TABLE III. Lists of monograms, digrams, and trigrams that are unique to a specific predator species. The list below the bold heading in each cell gives all of the monograms, digrams, and trigrams that begin with the symbol named in the heading and that are unique to the particular predator species also named in the heading of the cell.

S1	**S2(Dog)** S2S2	**S3**	**S4(Dog)** S4	**S5**	**S6**	**S7**	**S8**
S9	**S10(Coyote)** S10S10	**S11**	**S12(Coyote)** S12S12	**S13**	**S14(Dog)** S14S14	**S15**	**S16**
S17	**S18**	**S19(Coyote)** S19S37	**S20**	**S21(Coyote)** S21S21 S21S10	**S22(Coyote)** S22S22	**S23**	**S24(Coyote)** S24S21
S25(Dog) S25S25	**S26(Dog)** S26S26	**S27**	**S28**	**S29**	**S30**	**S31(Coyote)** S31S31	**S32**
S33	**S34**	**S35(Hawk)** S35S35 S35S60 S35S35S60 S35S35S58 S35S35S53 S35S60S53 S35S58S11 S35S53 S35S53S51	**S36**	**S37(Coyote)** S37S37 S37S62 S37S19 S37S12	**S38(Coyote)** S38S38	**S39**	**S40**
S41	**S42**	**S43(Dog)** S43S43	**S44(Dog)** S44S63	**S45**	**S46**	**S47**	**S48**
S49(Dog) S49	**S50**	**S51(Hawk)** S51S51 S51S53 S51S51S53 S51S53S53 S51-53-60H	**S52**	**S53(Hawk)** S53 S53S60 S53S53S60 S53S60S60 S53S60S58 S53S51 S53S51S53	**S54(Dog)** S54S54	**S55**	**S56**
S57	**S58(Hawk)** S58S35S35 S58S11S32H	**S59**	**S60(Hawk)** S60S58 S60S60S58 S60S58S58 S60S58S35 S60S53 S60S53S53 S60S35	**S61**	**S62(Coyote)** S62S62 S62S37	**S63(Dog)** S63S63	**S64(Dog)** S64S64

nemes are known, several different output neurons located close together in a given SOM will recognize variations of the same phoneme. It seems reasonable to assume that a particular meaningful sound found in a nonhuman animal vocalization also would be expressed with some variation.

One area for future research would be to combine the recent progress made in using formants to identify acoustic structures in alarm calls (Slobodchikoff and Placer, 2006) with the knowledge gained from the experiments discussed in this paper. Specific formant information could be combined with the general spectral information utilized by the Kohonen maps to see if an improvement in the performance of the SOMs is achieved. Improvement in SOM performance, reflected by improved classification accuracies, would result in more accurate clustering of like sounds and would provide more detailed information about the small-scale acoustic structure of the vocalizations.

Another fruitful area for future research would be to apply the techniques discussed in this paper to recordings of animal vocalizations that allow the exploration of new types of information encoding. For example, we are now preparing to use Kohonen mapping techniques to study a newly recorded set of alarm calls that will allow us to search not only for predator species information but also for descriptive information about predators and for information about the individuals that are vocalizing. In future work, as the Kohonen methodology is applied to recordings of alarm calls vocalized in a wide variety of situations, it might be possible to

develop an acoustic dictionary for prairie dog communications. In such a dictionary each acoustic pattern in a large collection of such patterns would be associated with specific information content. It would then be possible to examine all the patterns found in the dictionary in order to attempt to develop a general theory ofthe lexical structure of prairie dog communications. The development of this type of rudimentary lexicon for prairie dog communications would also make it possible to attempt a higher level examination of the complete structure of alarm calls to see if a syntactic structure could be determined for these calls. Information about a possible syntactic structure of alarm calls might even allow the synthetic creation of alarm calls for playback experiments.

Most of the work done by researchers on animal vocalizations has been focused on the macrostructure and general acoustic characteristics of those vocalizations. Consequently, not many tools have been developed with which to study the acoustic microstructure of animal vocalizations. Yet, the vocalizations of a growing number of animals are being shown to contain referential communication which conveys information about types of predators, degrees of danger, and other details about the world in which these animals live. These discoveries suggest that understanding the small-scale structure of animal vocalizations might be vitally important to determining the level of complexity of those communications and the actual information they contain. New methodologies and software tools for studying the acoustic microstructure of animal vocalizations need to be created so that detailed models and analyses related to these communications can be developed. The experiments described in this paper show how Kohonen maps might well provide one class of these needed tools.

Ackers, S. H., and Slobodchikoff, C. N. (**1999**). "Communication of stimulus size and shape in alarm calls of Gunnison's prairie dogs," Ethology **105**, 149–162.

Baker, M. C., and Logue, D. M. (**2003**). "Population differentiation in a complex bird sound: A comparison of three bioacoustical analysis procedures," Ethology **109**, 223–242.

Blackshaw, J. K., Jones, D. N., and Thomas, F. J. (**1996**). "Vocal individuality during suckling in the intensively housed domestic pig," Appl. Anim. Behav. Sci. **50**, 33–41.

Clark, C. W., Marler, P., and Beeman, K. (**1987**). "Quantitative analysis of animal vocal phonology: an application to swamp sparrow song," Ethology **76**, 101–115.

Eakle, W. L., Mannan, R. W., and Grubb, T. G. (**1989**). "Identification of individual breeding bald eagles by voice analysis," J. Wildlife Management **53**, 450–455.

Galeotti, P., Paladin, M., and Pavan, G. (**1993**). "Individually distinct hooting in male pygmy owls Glaucidium passerinum: a multivariate approach," Ornis Scandi. **24**, 15–20.

Galeotti, P., and Pavan, G. (**1991**). "Individual recognition of male tawny owls (Strix aluco) using spectrograms of their territorial calls," Ethol. Ecol. Evol. **3**, 113–126.

Hill, F. A. R., and Lill, A. (**1998**). "Vocalizations of the Christmas Island hawk-owl Ninox natalis: Individual variation in advertisement calls," Emu **98**, 221–226.

Kangas, J., Torkkol, K., and Kokknen, M. (**1992**). "Using SOMs as feature extractors for speech recognition," ICASSP-92 **2**, 341–344.

Kohonen, T. (**1988**). "The 'neural' phonetic typewriter," Computer **21**, 11–22.

Kohonen, T. (**1998**). "The self-organizing map," Neurocomputing **21**, 1–6.

Lengagne, T. (**2001**). "Temporal stability in the individual features in the calls of eagle owls (Bubo bubo)," Behaviour **138**, 1407–1419.

Mihelic, F., Ipsic, I., Dobrisek, S., and Pavesic, N. (**1992**). "Feature representations and classification procedures for Slovene phoneme recognition," Pattern Recogn. Lett. **13**, 879–891.

Placer, J., and Slobodchikoff, C. N. (**2000**). "A fuzzy-neural system for identification of species-specific alarm calls of Gunnison's prairie dogs," Behav. Processes **52**, 1–9.

Placer, J., and Slobodchikoff, C. N. (**2001**). "Developing new metrics for the investigation of animal vocalizations," Intell. Autom. Soft Comput. **7**, 249–259.

Placer, J., and Slobodchikoff, C. N. (**2004**). "A method for identifying sounds used in the classification of alarm calls," Behav. Processes **67**, 87–98.

Slobodchikoff, C. N., Ackers, S. H., and Van Ert, M. (**1998**). "Geographical variation in the alarm calls of Gunnison's prairie dogs," J. Mammal. **79**, 1265–1272.

Slobodchikoff, C. N., Kiriazis, J., Fischer, C., and Creef, E. (**1991**). "Semantic information distinguishing individual predators in the alarm calls of Gunnison's prairie dogs," Anim. Behav. **42**, 713–719.

Slobodchikoff, C. N., and Placer, J. (**2006**). "Acoustic structures in the alarm calls of Gunnison's prairie dogs," J. Acoust. Soc. Am. (to be published).

Sousa-Lima, R. S., Paglia, A. P., and Da Fonseca, G. A. B. (**2002**). "Signature information and individual recognition in the isolation calls of Amazonian manatees, Trichechus inunguis (Mammalia: Serenia)," Anim. Behav. **63**, 301–310.

Wu, P., Warwick, K., and Koska, M. (**1992**). "Neural network feature maps for Chinese phonemes," Neurocomputing **4**, 109–112.

Zimmerman, A., and Lerch, C. (**1993**). "The complex acoustic design of an advertisement call in male mouse lemurs (*Microcebus murinus*) and sources of its variation," Ethology **93**, 211–224.

Tissue deformation induced by radiation force from Gaussian transducers

Matthew R. Myers
Center for Devices and Radiological Health, HFZ-170, U. S. Food and Drug Administration, Rockville, Maryland 20852

(Received 29 August 2005; revised 21 February 2006; accepted 21 February 2006)

Imaging techniques based upon the tissue mechanical response to an acoustic radiation force are being actively researched. In this paper a model for predicting steady-state tissue displacement induced by a radiation force arising from the absorption of Gaussian ultrasound beams is presented. A simple analytic expression is derived that agrees closely with the numerical quadrature of the displacement convolution integrals. The analytic result reveals the dependence of the steady-state axial displacement upon the operational parameters, e.g., an inverse proportional relationship to the tissue shear modulus. The derivation requires that the transducer radius be small compared to the focal length, but accurate results were obtained for transducer radii comparable to the focal length. Favorable comparisons with displacement predictions for non-Gaussian transducers indicate that the theory is also useful for a broader range of transducer intensity profiles. [DOI: 10.1121/1.2187087]

PACS number(s): 43.80.Gx, 43.35.Wa [CCC] Pages: 3147–3152

I. INTRODUCTION

Imaging applications based upon the mechanical response of tissue to acoustic radiation force show promising clinical potential (Nightingale *et al.*, 2001; Palmeri and Nightingale, 2004; Walker 1999). An important metric used to assess the feasibility of these applications is the tissue deformation field resulting from the absorption of an ultrasound beam of specified characteristics. Mathematical models are valuable for predicting deformation as a function of tissue properties, transducer characteristics, and acoustic power levels.

Sarvazyan *et al.* (1998) derived a solution to the displacement equation in terms of an inverse Hankel transform, to study the strain induced during shear-wave elasticity imaging. Walker (1999) computed the deformation due to a focused ultrasound beam modeled as a series of absorbing disks. The intensity was assumed constant within each disk. The sum over the disk forcing functions was performed numerically. Nightingale *et al.* (2001) estimated the strain and displacement for radiation forces typical of remote palpation. Their model assumed a point load for the localized force. Lizzi *et al.* (2003) performed finite-difference simulations of tissue displacement, and analyzed how the induced motion is affected by the stiffness of the lesion during focused ultrasound surgery. Palmeri *et al.* (2005) utilized the finite-element technique to develop a general model for computing the tissue response to acoustic radiation force.

In this paper, an analytic expression for the steady-state deformation generated by an acoustic radiation force is developed. It is assumed that the amplitude distribution on the transducer face has a Gaussian amplitude shading. The Gaussian stipulation, along with the assumption that the focal length is long compared to the transducer radius, allows the final solution for the steady-state axial displacement to be given in terms of a simple, closed-form expression. This expression is advantageous in elucidating the role of the various operational parameters, and for providing rapid estimates of tissue response. Additionally, the simple functional form is useful for inclusion in standards and guidelines for the safe use of imaging modalities.

In the following section the asymptotic model is developed. In Sec. III, calculations are performed to test the accuracy of the model relative to a full, two-dimensional quadrature over the acoustic intensity distribution. The parameter values provided by Walker (1999) are used in the calculations. In Sec. IV some of the features of the displacement field elucidated by the model are discussed, and the limits of validity of the asymptotic approach are provided.

II. METHOD

We first consider a point force located at the origin in an elastic medium, acting in the axial (z) direction. The axial displacement due to the force of magnitude F is given by Love (1944):

$$w = D\left(\frac{z^2}{R^3} + \frac{\lambda + 3\mu}{\lambda + \mu}\frac{1}{R}\right), \tag{1}$$

where

$$D = \frac{F(\lambda + \mu)}{8\pi\mu(\lambda + 2\mu)}. \tag{2}$$

Here λ and μ are the Lame' constants for the medium, z is the axial coordinate of the observation point, and $R=(x^2+y^2+z^2)^{1/2}$ is the distance from the origin to the observation point. For a volumetric force with density (force/volume) f, the total displacement may be obtained by convolving (1) with f:

$$w(x,y,z)=\int \frac{f(x',y',z')(\lambda+\mu)}{8\pi\mu(\lambda+\mu)} \times\left(\frac{(z-z')^2}{[(x-x')^2+(y-y')^2+(z-z')^2]^{3/2}} +\frac{\lambda+3\mu}{\lambda+\mu}\frac{1}{[(x-x')^2+(y-y')^2+(z-z')^2]^{1/2}}\right)dV'. \tag{3}$$

The quantity multiplying the force density f is the Green's function for an infinite medium, and hence (3) will not, in general, satisfy the exact conditions at the boundary, e.g., at the transducer face. This issue is discussed further in Sec. IV.

We consider next the case where the force density arises due to the absorption of a propagating ultrasound beam. The density f is given by the rate of attenuation of ultrasound intensity divided by the speed of sound. Within the plane-wave approximation for the beam,

$$f=2\alpha I/c \tag{4}$$

(Nightingale *et al.*, 2001), where α is the amplitude absorption coefficient. We suppose that the ultrasound beam is focused and propagating in the $+z$ direction, and that it possesses Gaussian amplitude shading. The intensity profile for the beam is given by (Wu and Du, 1990),

$$I=I_0B^{-1}A_1(z)\exp[-2A_1(r/a)^2]\exp(-2\alpha z), \tag{5}$$

where

$$A_1(z)=\frac{B}{\frac{B^2z^2}{r_0^2}+\left(1-\frac{z}{z_f}\right)^2}. \tag{6}$$

Here B is the Gaussian coefficient at the transducer, z_f is the geometric focal length of the transducer, and $r_0=\frac{1}{2}ka^2$, k being the acoustic wavenumber and a the transducer radius. Upon using (4)–(6) in (3) and switching to cylindrical coordinates, we obtain the following expression for the axial displacement at a point located along the transducer axis:

$$w(x=0,\ y=0,\ z)=w_1+w_2, \tag{7}$$

$$w_1=\frac{\alpha(\lambda+\mu)I_0}{2Bc\mu(\lambda+2\mu)}\int_0^\infty dr'\int_0^\infty dz'\,A_1(z') \times\exp[-2A_1(r'/a)^2] \times\exp(-2\alpha z')\frac{r'(z-z')^2}{[r'^2+(z-z')^2]^{3/2}}, \tag{8}$$

$$w_2=\frac{\alpha(\lambda+3\mu)I_0}{2Bc\mu(\lambda+2\mu)}\int_0^\infty dr'\int_0^\infty dz'\,A_1(z') \times\exp[-2A_1(r'/a)^2] \times\exp(-2\alpha z')r'[r'^2+(z-z')^2]^{-1/2}. \tag{9}$$

Expressions (8) and (9) require the properties of the medium to be uniform. We first consider displacement at the focus $z=z_f$. After making the transformations

$$\rho=\frac{(2A_1)^{1/2}r'}{a},\quad \zeta=\frac{z'-z_f}{z_f}, \tag{10}$$

we obtain for (9),

$$w_2=\frac{\alpha(\lambda+3\mu)I_0a^2}{4Bc\mu(\lambda+2\mu)}e^{-2\bar\alpha}\int_0^\infty d\rho\,\rho e^{-\rho^2} \times\int_{-1}^\infty d\zeta\frac{e^{-2\bar\alpha\zeta}}{\{\delta^2\rho^2[\epsilon^2(\zeta+1)^2+\zeta^2]+\zeta^2\}^{1/2}}, \tag{11}$$

where

$$\delta=\frac{a}{(2B)^{1/2}z_f},\quad \bar\alpha=\alpha z_f,\quad \epsilon=\frac{Bz_f}{r_0}=\frac{2Bz_f}{ka^2}. \tag{12}$$

The parameter ϵ is the reciprocal of the transducer gain. The parameter δ, proportional to the reciprocal of the transducer f number, measures the transducer radius relative to the transducer focal length. When δ is small, asymptotic methods can be employed to perform the integral in (11), in the following way.

We employ the method of matched asymptotic expansions (Van Dyke, 1975) to construct an integrand that approximates that in (11) when $\delta\ll1$. We first consider just the ζ integral, and break it up into the intervals $(-1,0)$ and $(0,\infty)$. The corresponding integrals over these intervals we denote I_- and I_+. When ζ is of magnitude 1 (outer integrand), the integrand of I_+ is approximately equal to

$$\frac{e^{-2\bar\alpha\zeta}}{\zeta}. \tag{13}$$

For the inner integrand where ζ is small (of magnitude $\epsilon\delta$), the integrand of I_+ may be approximated by

$$\frac{1}{(\delta^2\rho^2\epsilon^2+\zeta^2)^{1/2}}. \tag{14}$$

The common representation of the integrand of I_+, obtained by expanding the outer integrand for small ζ or the inner integrand for large ζ, is equal to $1/\zeta$. The composite representation for the integrand of I_+, equal to the inner+outer-common, is given by

$$\frac{e^{-2\bar\alpha\zeta}}{\zeta}+\frac{1}{(\delta^2\rho^2\epsilon^2+\zeta^2)^{1/2}}-\frac{1}{\zeta}. \tag{15}$$

The composite representation is integrable over $(0,\infty)$ and may be integrated to

$$I_+=-\log(\epsilon\,\delta\rho)-C_e-\log(\bar\alpha). \tag{16}$$

Here C_e is Euler's constant (≈0.577). Repeating the same process for the interval $(-1,0)$ yields

$$I_-=\log(2)-\log(\epsilon\,\delta\rho)+\int_0^1\frac{e^{2\bar\alpha t}-1}{t}dt. \tag{17}$$

The integral in (17) may be expressed in terms of an exponential integral $Ei(x)$ (Gradshteyn and Ryzhik, 1980). However, in regions of maximum displacement (see Sec. III), this term is typically less important than the logarithmic term, and a less accurate expression suffices. We ap-

proximate the integrand by a line connecting the integration end points (trapezoidal rule), resulting in an integral equal to $\bar{\alpha}+(\exp 2\bar{\alpha}-1)/2$. The total ζ integral then becomes

$$I_- + I_+ = \log(2) - 2\log(\epsilon\delta\rho) - \log(\bar{\alpha}) - C_e + \bar{\alpha} + (\exp 2\bar{\alpha} - 1)/2. \tag{18}$$

Inserting (18) into (11) and using

$$\int_0^\infty \rho \log(\rho) e^{-\rho^2}\, d\rho = -C_e/4 \tag{19}$$

(Gradshteyn and Ryzhik, 1980) yields the asymptotic approximation for w_2:

$$w_2 = \frac{\alpha(\lambda+3\mu)I_0 a^2}{8Bc\mu(\lambda+2\mu)} e^{-2\bar{\alpha}} \left[\log\left(\frac{2}{\bar{\alpha}}\right) - 2\log(\epsilon\delta) + \bar{\alpha} + \frac{e^{2\bar{\alpha}}-1}{2} \right]. \tag{20}$$

The other contribution to the axial displacement, w_1 [Eq. (7)], may also be estimated using the method of matched asymptotic expansions. The process is similar to that for w_2. When the result is combined with (20), we have the complete approximation to the axial displacement at the focus:

$$w(0,0,z_f) = \frac{\alpha I_0 a^2 e^{-2\bar{\alpha}}}{4Bc\mu} \left\{ -\frac{(\lambda+\mu)}{(\lambda+2\mu)} + \left[\log\left(\frac{2}{\bar{\alpha}}\right) - 2\log(\epsilon\delta) + \bar{\alpha} + \frac{e^{2\bar{\alpha}}-1}{2} \right] \right\}. \tag{21}$$

For tissues it is typically the case that $\lambda \gg \mu$. (See Sec. III.) Using this approximation, along with the definitions of α, δ, and ϵ (12), we obtain

$$w(0,0,z_f) \approx \frac{\alpha I_0 a^2 e^{-2\alpha z_f}}{4Bc\mu} \left[2\log\left(\frac{ka}{\sqrt{B}}\right) - \log(\alpha z_f) + \alpha z_f + \frac{e^{2\alpha z_f}-3}{2} \right]. \tag{22}$$

Equation (22) represents the axial displacement at the focus ($z=z_f$). To determine the axial displacement at points on the z axis different from the focus, we return to Eq. (10). We use the coordinate z in the ζ transformation instead of z_f. Upon making this replacement and continuing the analysis, we find that the denominator in the ζ integral is

$$\{\delta^2\rho^2[\epsilon^2(\zeta+1)^2 + (z_f/z - 1 + \zeta)^2] + \zeta^2\}^{1/2}. \tag{23}$$

For $\delta \ll 1$ and ζ and z_f/z values of order unity, this expression is approximately equal to $|\zeta|$, as is the case for the denominator of the ζ integral in (11). When ζ is small, (23) can be approximated by $[\delta^2\rho^2(\epsilon^2+(z_f/z-1)^2)+\zeta^2]^{1/2}$, compared with $[\delta^2\rho^2\epsilon^2+\zeta^2]^{1/2}$ for the corresponding term of (11), which appears in (14). By comparing these two small-ζ approximations, it can be seen that if we make the assignment

$$\epsilon' = [\epsilon^2 + (z_f/z - 1)^2]^{1/2}, \tag{24}$$

and define the parameters of (12) in terms of z rather than z_f, the analysis beginning with (14) is identical to the case of $z=z_f$. The final result for the displacement in the z direction at location z can be obtained directly from (21):

$$w(0,0,z) = \frac{\alpha I_0 a^2 e^{-2\overline{\alpha'}}}{4Bc\mu} \left\{ -\frac{(\lambda+\mu)}{(\lambda+2\mu)} + \left[\log\left(\frac{2}{\overline{\alpha'}}\right) - 2\log(\epsilon'\delta') + \bar{\alpha}' + \frac{e^{2\overline{\alpha'}}-1}{2} \right] \right\}. \tag{25}$$

Here

$$\delta' = \frac{a}{(2B)^{1/2}z}, \quad \overline{\alpha'} = \alpha z, \quad \epsilon' = \left[\left(\frac{2Bz}{ka^2}\right)^2 + \left(\frac{z_f}{z} - 1\right)^2 \right]^{1/2}. \tag{26}$$

If we insert the expressions of (26) directly into (25), and as in (22) assume $\lambda \gg \mu$, we arrive at the simplified result for the axial displacement at location z:

$$w(0,0,z) \approx \frac{\alpha I_0 a^2 e^{-2\alpha z}}{4Bc\mu} \left[2\log\left(\frac{ka\gamma}{\sqrt{B}}\right) - \log(\alpha z) + \alpha z + \frac{e^{2\alpha z}-3}{2} \right], \tag{27}$$

where

$$\gamma(z) = \left(1 + \frac{k^2a^4(z_f - z)^2}{4B^2z^4}\right)^{-1/2}. \tag{28}$$

Using (12) and remembering that ϵ is the reciprocal of the gain G, (28) can be expressed as

$$\gamma(z) = \left[1 + G^2\left(\frac{z_f}{z}\right)^2\left(1 - \frac{z_f}{z}\right)^2\right]^{-1/2}. \tag{29}$$

III. RESULTS

To test the accuracy of the asymptotic approximation (27), the displacement field was calculated by both (27) and the double integrals in (7)–(9). The tissues considered by Walker (1999)—vitreous, breast, and liver—were used in the calculations. The material properties used in the calculations are shown in Table I. The attenuation and speed of sound are the values quoted by Walker (1999), and the Lame' constants are derived from Walker's values of the Elastic modulus and Poisson's ratio (Love 1944). It can be seen that the approximation $\lambda \gg \mu$ is appropriate for the organs considered.

TABLE I. Mechanical properties of tissues considered in calculations.

	α (dB/cm/MHz)	c (m/s)	λ (N/m^2)	μ (N/m^2)
Breast	0.75	1510	1.67×10^6	3330
Liver	1.17	1540	5.8×10^5	1170
Vitreous	0.1	1513	16.7	0.033

Also following Walker (1999), a focal length of 1.0 cm, a transducer radius of 0.25 cm, and a frequency of 10 MHz were used in the calculations. The peak intensity I_0 for the Gaussian beam was set to 0.0123 W/cm^2, corresponding to an average intensity of 0.0062 W/cm^2 at the transducer. This

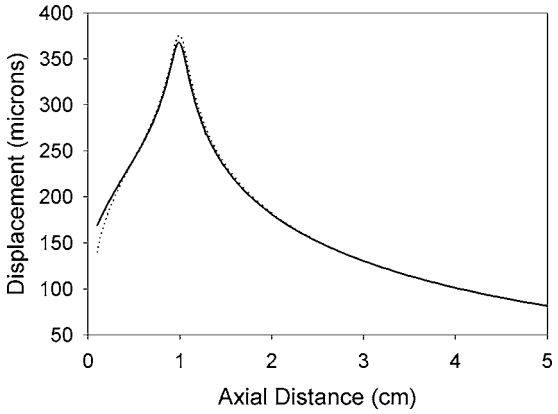


FIG. 1. Axial displacement as a function of axial distance, for vitreous tissue. Intensity=1 W/cm^2 at focus, in water. Transducer diameter =0.5 cm, focal length=1.0 cm, frequency=10 MHz. Solid line—numerical integration of convolution integral. Dotted line—asymptotic model.

value corresponds to an average intensity of 1.0 W/cm^2 at the focus, in water. The Gaussian coefficient B was set to 1.0.

The double integrals in (7)–(9) were performed using a numerical implementation of Simpson's rule on an IBM RS6000/44P workstation. The upper limits of integration on r' and z' [infinite in (8) and (9)] were increased until doubling the values resulted in less than a 0.1% change in the integral. Final upper limits used were $10a$ for the radial integral and $40z_f$ for the axial integral.

Figure 1 depicts the axial displacement as a function of distance for the vitreous tissue. A sharp peak in the displacement occurs at the beam focus ($z=1$ cm). The displacement at the focus is approximately 0.037 cm for the ultrasound intensity considered. The difference between the numerical result and that predicted by (27) is less than a few percent, except at very small axial distances, where the analytical result is lower.

The corresponding curves for breast tissue are shown in Fig. 2. The magnitude of the displacement is seen to be much smaller than for vitreous, at all locations. Two slight peaks, one at the focal length and one at slightly less than half the focal length, can be seen in a broad region of maximum displacement. The analytical prediction is slightly higher in

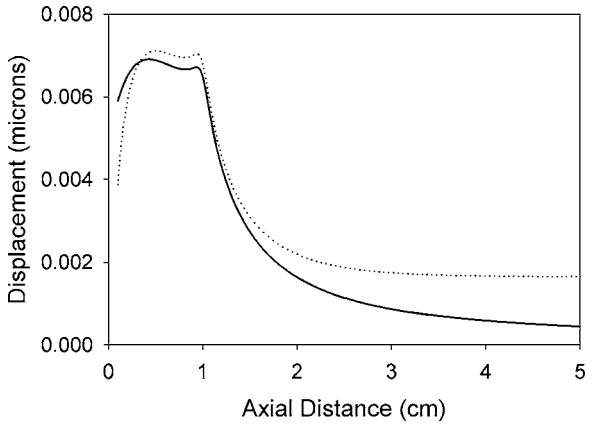


FIG. 2. Axial displacement as a function of axial distance, for breast tissue. Intensity=1 W/cm^2 at focus, in water. Transducer diameter=0.5 cm, focal length=1.0 cm, frequency=10 MHz. Solid line—numerical integration of convolution integral. Dotted line—asymptotic model.

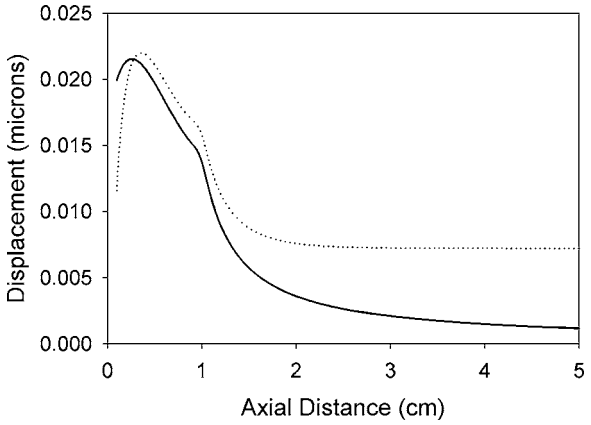


FIG. 3. Axial displacement as a function of axial distance, for liver tissue. Intensity=1 W/cm^2 at focus, in water. Transducer diameter=0.5 cm, focal length=1.0 cm, frequency=10 MHz. Solid line—numerical integration of convolution integral. Dotted line—asymptotic model.

this plateau region. As with the vitreous tissue, the analytical result drops off rapidly for very small axial distances. Beyond about 1.5 focal lengths, the analytical result significantly exceeds the numerical one.

For the liver (Fig. 3), there is a sharp peak in the displacement curve occurring well ahead ($z\approx 0.25$ cm) of the focal zone. Again the analytical result slightly exceeds the numerical computation in the region of maximum displacement, as well as for axial distances up to a focal length. Beyond a focal length, the discrepancy becomes significantly larger. As with the breast and vitreous, the analytical approach considerably underestimates the displacement very close to the transducer. Displacement magnitudes are comparable to those for the breast, and much smaller than those for the vitreous.

A second set of parameters was used in a study to quantify the dependence of axial deformation upon attenuation, for different transducer sizes. Three different transducer radii were considered—0.25, 0.5, and 1.0 cm—with the total radiated power held constant at 0.12 W. The frequency was 5 MHz, the shear modulus μ 1170 N/m^2, and the focal length 2.0 cm. The displacement at the focus was computed using (27) and plotted in Fig. 4. The attenuation of maximum displacement decreases slightly with increasing transducer radius; for $a=1$ cm it is about 2.3 dB/cm. Over the entire range of attenuations, doubling the transducer radius (while decreasing the intensity to hold the power constant) from 0.25 to 0.5 cm provides the same increment in displacement as doubling the radius from 0.5 to 1.0 cm.

IV. DISCUSSION

The requirement that the focal length be long compared with the transducer radius (small δ) allows the displacement at a given axial location to be split into contributions due to ultrasound absorption at that location and absorption over the rest of the beam. In the derivation using matched asymptotic expansions [Eqs. (13)–(20)], the region near $\zeta=0$ represents the volume close to the observation point, while larger values of ζ represent the rest of the beam. Roughly speaking, of the bracketed terms in (27), the single term not containing the

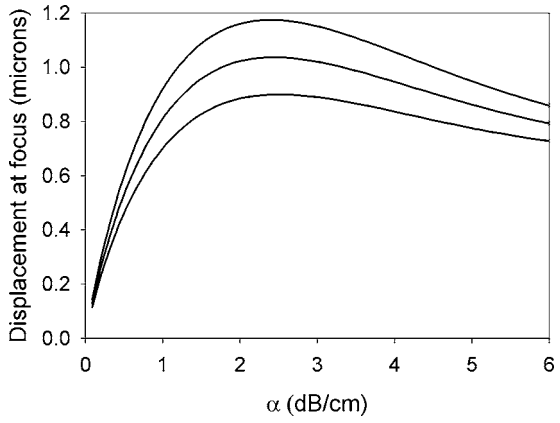


FIG. 4. Axial displacement at the focus as a function of attenuation, for transducer radii of 1.0 cm (top curve), 0.5 cm (middle), and 0.25 cm (bottom). Power is held constant at 0.12 W. Frequency is 5 MHz, the shear modulus 1170 $\mathrm{N/m^2}$, and the focal length 2.0 cm.

attenuation is the local contribution, while the other terms represent the contribution from the remote sections of the beam. For a slightly attenuating medium such as the vitreous, the first term in brackets in (27) dominates the others near $z=z_f$, and displacement in the focal region is essentially a local effect. This is consistent with the peaked profile of Fig. 1. For a highly attenuating medium such as liver, the first bracketed term is smaller than the others, even at the focus, and the displacement at the focus is due mostly to distributed forces throughout the beam. Hence, only a slight deviation is seen in the displacement curve at the focus in Fig. 3.

Near the transducer, local effects cannot be separated from the boundary contribution, and the approximation breaks down. The analytical result is thus seen to deviate from the numerical one for small axial distances in Figs. 1–3. For the calculations performed, accuracy is better than 20% if the observation point z is kept greater than $0.15z_f$.

The deviation between the analytic predictions and numerical results for larger values of αz (larger axial distances in Figs. 2 and 3) is due partly to the limitation of the inner approximation to the exponential term [$\exp(2\alpha\zeta)\approx 1$ for small ζ], leading to (17), and partly to the approximation to the exponential integral in (17). A numerical implementation of $Ei(x)$ would result in closer agreement between the asymptotic theory and numerical integration of (7)–(9), for all values of αz. However, since the affected region is considerably beyond the location where maximum displacement occurs, the added complexity was not deemed worthwhile. The model as presented in this paper is accurate (error $<20\%$) for $\alpha z<1.5$, that is, the focal length cannot significantly exceed the ultrasound attenuation length.

The results of Fig. 4 follow the behavior predicted analytically by (27). For constant power (i.e., constant I_0a^2) and large ka, the increase in displacement with increasing radius is determined by the logarithmic term in (27), and, hence, doubling the radius results in an increase proportional to log 2, independent of radius. Also, (27) predicts a maximum displacement for $\alpha=1/(2z_f)$ when ka is large. For the parameters chosen, this attentuation producing maximum displacement (in the limit of large ka) is 2.2 dB/cm, close to what is observed in Fig. 4.

For a Gaussian beam, the average intensity across the transducer face is $I_0/2B$ (Wu and Du, 1990). The acoustic power as a function of the axial location is given by

$$W(z)=\frac{1}{2}I_0\pi a_{\mathrm{eff}}^2\exp(-2\alpha z), \tag{30}$$

where $a_{\mathrm{eff}}=a/\sqrt{B}$ is the effective transducer radius. The acoustic power (per unit length) dissipated at location z is $-W'(z)=2\alpha W(z)$. Expressing (27) in terms of $W'(z)$, we obtain

$$w(0,0,z)\approx\frac{-W'(z)}{4\pi\mu c}\left(2\log(ka_{\mathrm{eff}}\gamma)-\log(\alpha z)+\alpha z+\frac{e^{2\alpha z}-3}{2}\right). \tag{31}$$

Equation (30) may be interpreted as the displacement for a general focused transducer of radius a_{eff} and power $W(z)$. Since the transducer power level and effective radius considered in this paper match those of Walker (1999), as do the tissue properties, a comparison with the results of Walker (1999) based upon a transducer without Gaussian shading can be meaningfully performed. Indeed, the plots of Walker closely match those in Figs. 1–3 in the locations where the asymptotic theory is valid, roughly $0.15\ \mathrm{cm}<z<1.5$ cm. This region includes the focal zone and region of maximum displacement. The agreement provides evidence that the asymptotic theory can be applied to non-Gaussian transducers via (31).

In the calculations of Sec. III, the bracketed quantity in (31) divided by 4π was a constant of magnitude 1 over a wide range of parameter values. Thus, for very quick estimates, the magnitude of the displacement field may be approximated by

$$w(z_f)\approx -W'(z)/\mu c. \tag{32}$$

This combination approximates the axial displacement at the focus to within 20%, for all three tissues considered in Figs. 1–3.

A useful alternative form of (32) may also be obtained by using (30), and expressing I_0 in terms of the focal intensity I_f via (5):

$$w(z_f)\approx\frac{4\pi\alpha I_f z_f^2}{\mu c(ka)^2}. \tag{33}$$

Here z_f is the focal length and k the acoustic wave number. The value of αz_f considered by Lizzi *et al.* (2003) is too large for the full expression (31) to be valid, however, a comparison of sorts can be made on the basis of (33). After inserting the parameters ($f=4.09$ MHz, $\alpha=0.5$ dB/(cm MHz), $\mu=400\ \mathrm{N/m^2}$, $a=2.1$ cm, $I_f=345\ \mathrm{W/cm^2}$) of Lizzi *et al.* (2003) in (33), we obtain a displacement of approximately 110 microns. Lizzi *et al.* measured and computed a displacement of approximately 50 microns at the end of the 5 ms push exposure. Examining the convergence of the displacement curves (Fig. 9 of Lizzi *et al.*) for times during the exposure, we can extrapolate to a steady-state displacement on the order of

100 microns. The empirical observation of Nightingale *et al.* (2001) that displacements on the order of tens of microns require intensities on the order of 300 W/cm^2 can also be verified using (33). Following Nightingale *et al.* (2001) and setting $\mu=400$ kPa, $f=7.2$ MHz, $\alpha=0.432$ cm^{-1}, $a=2.5$ cm, and $z_f=5$ cm (for $F/1$), we arrive at a displacement of 12 microns when the focal intensity is 300 W/cm^2.

Palmeri *et al.* (2005) employ a rectangular array and highly transient conditions, and hence a direct comparison cannot be made between their results and those of this study. However, the functional dependence of displacement upon $1/\mu$ predicted by the present model [e.g. (27) or (33)] can be investigated using the experimental and finite-element results of Palmeri *et al.* The shear moduli for the 100 bloom, 150 bloom, and 200 bloom phantoms of Palmeri *et al.* are 2.83, 6, and 7.67 kPa, respectively, and the maximum displacements for the three phantoms are 10.1, 4.2, and 3.0 μm. Assuming an inverse proportional dependence upon shear modulus, the maximum displacement for the 200 bloom phantom relative to the maximum displacement for the 150 bloom is 0.78. The experimental value is 0.71 (±0.15), while the finite-element simulations give 0.80. The ratio of the maximum displacement for the 100 bloom phantom relative to the maximum displacement for the 150 bloom is 2.12 based upon a $1/\mu$ dependence, while the experiments and finite-element simulations give 2.40 (±0.35) and 1.85, respectively. These results provide some confirmation of the simple $1/\mu$ dependence, though it must be noted that the maximum displacement may occur at different axial locations for the different phantoms, while the displacement ratios calculated strictly upon a $1/\mu$ dependence assume that z is the same in all cases.

The asymptotic approach formally requires that the transducer radius a be small compared to the focal length z_f. The calculations of Sec. III were extended to larger transducer radii, maintaining the focal length at 1.0 cm. The accuracy of the analytic expression degraded with increasing a/z_f, however, the error in the displacement at the focus was only 8% for $a=z_f/2$ (transducer diameter equal to focal length), and 23% for $a=z_f$. The accuracy of the asymptotic model is likely to suffice for transducer geometries applicable to radiation-force imaging.

As noted in Sec. II, an infinite-medium Green's function was employed to derive the convolution-integral solution (3) for the axial displacement. This approximation, equivalent to assuming that the region $z<0$ is filled with a medium possessing identical elastic properties to the region $z>0$, but with no absorbed energy, precludes the enforcement of a boundary condition on the transducer face. In the important case where $\delta<1$, $\alpha z_f<1$, $\epsilon\ll 1$ (focal length larger than transducer radius, moderate attenuation, high gain), the displacement near the focus is dominated by the local radiation force, and the boundary effects are likely to be ignorable. For the analogous case of temperature rise computed using the infinite-medium Green's function, Wu and Du (1990) found that the temperature rise computed via an expression similar to (3) underestimated the solution imposing a zero-temperature boundary condition on $z=0$. The difference at the focus was less than 10%. The relative error compared to the solution imposing a zero-derivative condition at $z=0$ is about 20%. For high levels of attenuation, where the maximum displacement is achieved close to the transducer, the error associated with ignoring the boundary condition on $z=0$ is likely to be higher. As noted by Walker (1999), a more sophisticated Green's function could be employed in these cases.

V. CONCLUSION

The asymptotic model presented in this paper helps analyze the mechanical response of tissue to acoustic radiation force. The primary result of the model is given by (31). For tissue and transducer properties typical of radiation-force imaging, (31) accurately predicts the axial tissue displacement in critical areas such as the focal zone and region of maximum displacement. For quantitative accuracy, the axial distance of interest must be less than about 1.5 attenuation lengths, i.e., $\alpha z<1.5$. The model formally requires that the transducer radius be small compared to the focal length, but accurate results were obtained for transducer radii comparable to the focal length. Similarly, the model is developed for a transducer with Gaussian amplitude shading, but calculations showed that when expressed in terms of average power dissipation, the model appears to be applicable to more general intensity profiles.

Gradshteyn, I. S., and Ryzhik, I. M. (**1980**). *Table of Integrals, Series, and Products* (Academic Press, Orlando, FL).

Lizzi, F. L., Muratore, R., Deng, C. X., Ketterling, J. A., Alam, S. K., Mikaelian, S., and Kalisz, A. (**2003**). "Radiation-force technique to monitor lesions during ultrasonic therapy," Ultrasound Med. Biol. **29**, 1593–1605.

Love, A. E. H. (**1944**). *A Treatise on the Mathematical Theory of Elasticity* (Dover, New York).

Nightingale, K. R., Palmeri, M. L., Nightingale, R. W., and Trahey, G. E. (**2001**). "On the feasibility of remote palpation using acoustic radiation force," J. Acoust. Soc. Am. **110**, 625–634.

Palmeri, M. L., and Nightingale, R. W. (**2004**). "On the thermal effects associated with radiation force imaging of soft tissue," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 551–565.

Palmeri, M. L., Sharma, A. C., Bouchard, R. R., Nightingale, R. W., and Nightingale, R. W. (**2005**). "A finite-element method of soft tissue response to impulsive acoustic radiation force," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **52**, 1699–1712.

Sarvazyan, A. P., Rudenko, O. V., Swanson, S. D., Fowlkes, J. B., and Emelianov, S. Y. (**1998**). "Shear wave elasticity imaging: A new ultrasonic technology of medical diagnostics," Ultrasound Med. Biol. **24**, 1419–1435.

Van Dyke, M. (**1975**). *Perturbation Methods in Fluid Mechanics* (Parabolic Press, Stanford, CA).

Walker, W. F. (**1999**). "Internal deformation of a uniform elastic solid by acoustic radiation force," J. Acoust. Soc. Am. **105**, 2508–2518.

Wu, J., and Du, G (**1990**). "Temperature elevation generated by a focused Gaussian beam of ultrasound," Ultrasound Med. Biol. **16**, 489–498.

Acoustic structures in the alarm calls of Gunnison's prairie dogs

C. N. Slobodchikoff[a)]
Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona 86011

J. Placer
Department of Computer Sciences, Northern Arizona University, Flagstaff, Arizona 86011

(Received 29 September 2005; revised 27 January 2006; accepted 15 February 2006)

Acoustic structures of sound in Gunnison's prairie dog alarm calls are described, showing how these acoustic structures may encode information about three different predator species (red-tailed hawk—*Buteo jamaicensis*; domestic dog—*Canis familaris*; and coyote—*Canis latrans*). By dividing each alarm call into 25 equal-sized partitions and using resonant frequencies within each partition, commonly occurring acoustic structures were identified as components of alarm calls for the three predators. Although most of the acoustic structures appeared in alarm calls elicited by all three predator species, the frequency of occurrence of these acoustic structures varied among the alarm calls for the different predators, suggesting that these structures encode identifying information for each of the predators. A classification analysis of alarm calls elicited by each of the three predators showed that acoustic structures could correctly classify 67% of the calls elicited by domestic dogs, 73% of the calls elicited by coyotes, and 99% of the calls elicited by red-tailed hawks. The different distributions of acoustic structures associated with alarm calls for the three predator species suggest a duality of function, one of the design elements of language listed by Hockett [in *Animal Sounds and Communication*, edited by W. E. Lanyon and W. N. Tavolga (American Institute of Biological Sciences, Washington, DC, 1960), pp. 392–430]. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2185489]

PACS number(s): 43.80.Jz, 43.80.Ka [JAS] Pages: 3153–3160

I. INTRODUCTION

The alarm calls of some animals have been shown to have referential meaning. These calls refer to something in the external environment of the animal, such as a predator, that elicits an alarm response. A number of animals have calls for terrestrial and for aerial predators. These animals include many ground squirrels (*Spermophilus spp*) (Owings and Hennessy, 1984), chickens (*Gallus gallus domesticus*) (Evans and Evans, 1999), tree squirrels (*Tamiasciurus hudsonicus*) (Greene and Meagher, 1998), dwarf mongooses (*Helogale undulata*) (Beynon and Rasa, 1989), and suricates (*Suricata suricatta*) (Manser, 2001; Manser *et al.*, 2001). In each of these species, a terrestrial predator elicits an acoustically different alarm call from that elicited by an aerial predator. A few animal species have been shown to have specific vocalizations for different predator species or categories of predators. Such referential meaning has been documented for vervet monkeys (*Cercopithecus aethiops*), Diana monkeys (*Cercopithecus diana*), Campbell's monkeys (*Cercopithecus campbelli*), and prairie dogs (*Cynomys gunnisoni*). Vervet monkeys have calls for three different types of predators: snake or python, leopard, and eagle (Cheney and Seyfarth, 1990). Diana monkeys and Campbell's monkeys have calls for leopards (*Panthera pardus*) and crowned-hawk eagles (*Stephanoaetus coronatus*) (Zuberbuhler, 2000, 2001).

Gunnison's prairie dogs, which are found in the grasslands of northern Arizona, northern New Mexico, southern Colorado, and eastern Utah, have alarm calls for four different species of predator: hawk (*Buteo jamaicensis*), human (*Homo sapiens*), coyote (*Canis latrans*), and domestic dog (*Canis familiaris*) (Placer and Slobodchikoff, 2000, 2001), as well as vocalizations for objects in their environment that are not known to them but could potentially represent a predator (Ackers and Slobodchikoff, 1999). The escape responses of the prairie dogs differ to the different predators (Kiriazis, 1991; Slobodchikoff, 2002; Kiriazis and Slobodchikoff, 2006). When a human appears at the edge of the colony, all of the prairie dogs run to their burrows and dive inside. When a red-tailed hawk dives toward the colony in an attempt to capture a prairie dog, only the animals in the immediate flight path of the hawk run to their burrows and dive inside. The remaining animals outside the flight path typically stand on their hind legs and watch the progress of the hawk. When a coyote appears, all the animals in the colony run to the lips of their burrows and stand upright in an alert posture, and prairie dogs that are below ground emerge to stand at their burrows so that most of the colony is above ground, watching the coyote. When a domestic dog appears, the prairie dogs stand in an alert posture, and only run to their burrows if the dog comes near. In playback experiments, Kiriazis and Slobodchikoff (2006) showed that the escape responses of the prairie dogs to playbacks of the calls for the different predators when no predator is present are the same as when a predator appears and elicits alarm calls and escape responses from the animals in the colony. Of these four predators, humans represent perhaps the greatest threat,

[a)]Electronic mail: con.slobodchikoff@nau.edu

shooting large numbers of prairie dogs (Vosburgh and Irby, 1998). Coyotes and hawks are responsible for the majority of nonhuman predation, while domestic dogs seem to represent relatively little threat as predators (Verdolin and Slobodchikoff, 2002).

In addition to containing information about the species of predator, Gunnison's prairie dog alarm vocalizations also provide information about specific details of an individual predator. Slobodchikoff *et al.* (1991) found that Gunnison's prairie dogs incorporated information into their calls about specific colors of shirts, such as blue, green, yellow, and grey, worn by different humans, and also found that the calls incorporated details of the size and shape of the individual humans. When Slobodchikoff *et al.* (1991) had humans of different sizes and shapes walking through prairie dog colonies wearing the same white laboratory coat that obscured physical features, the prairie dogs incorporated less information into their alarm calls about the individual humans than when the humans were not wearing the laboratory coat. A recent study has shown that the other four species of prairie dogs in North America (the black-tailed, *Cynomys ludovicianus*, found in the western part of the Midwest; the Mexican, *C. mexicanus*, found in central Mexico; the white-tailed, *C. leucurus*, found in western Colorado and northeastern Utah; and the Utah, *C. parvidens*, found in southwestern Utah) can also incorporate information into their alarm calls about different shirt colors worn by humans (Frederiksen, 2005). The calling patterns of the Gunnison's prairie dogs differ depending on whether a predator is approaching slowly or rapidly, providing the prairie dogs with a wide range of information about the species of predator, the individual characteristics or description, and the rate of approach (Kiriazis, 1991; Slobodchikoff, 2002). This represents a complex communication system that appears to exceed the amount of information encoded in the calls of many other species with referential labeling.

In earlier studies (Placer and Slobodchikoff, 2000, 2001) an artificial neural network was used to classify Gunnison's prairie dog alarm calls by species of predator. Each alarm call was decomposed into several equal-sized time periods and the ratios of the frequencies present in each period were computed. The frequency ratios in all time periods for a given alarm call were then combined to represent that alarm call. This measure of frequency ratios was used as input to a neural network. The classification results clearly showed that frequency ratios could be used to classify predator-specific prairie dog alarm calls. However, although high classification accuracy was achieved in that study, the frequency ratios did not retain time-dependent information nor did they provide information about the underlying acoustic structure of the calls.

Analysis of the acoustic structure of such alarm calls in mammals has most often focused on a description of the sound frequencies and time-dependent parameters of those frequencies. For example, the alarm calls of vervet monkeys (*Cercopithecus aethiops*) and rhesus macaques (*Macaca mulatta*) have been analyzed in terms of the fundamental frequencies (*F*0) (Hauser and Fowler, 1992), and vervet monkeys (Seyfarth and Cheney, 1984; Owren and Bernacki, 1988), Barbary macaques (Macaca *sylvanus*) (Fischer *et al.*, 1995), baboons (*Papio cynocephalus*) (Owren *et al.*, 1997; Fischer *et al.*, 2002), and Diana monkeys (*Cercopithecus diana*) (Riede and Zuberbuhler, 2003) have been analyzed in terms of frequency, time, and amplitude parameters. These analyses have not searched for deeper acoustic structures within the alarm calls.

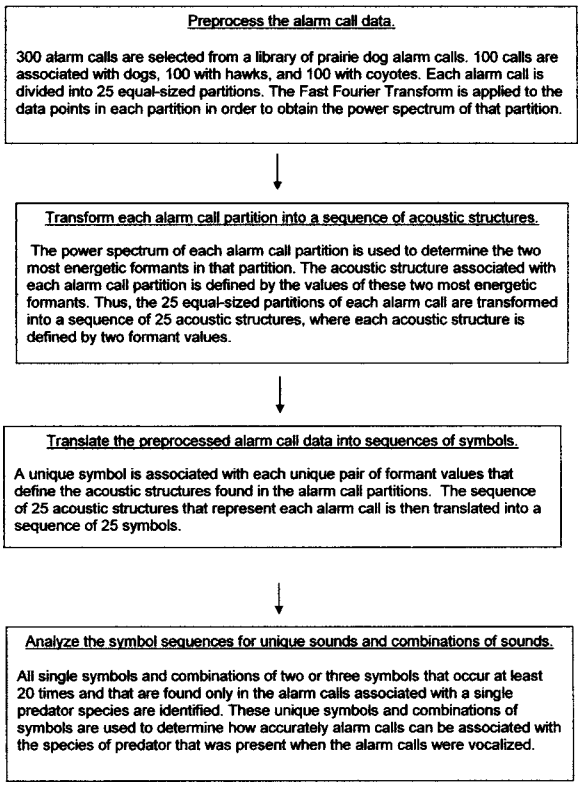


FIG. 1. Summary of the methods used in the analysis.

II. METHODS

The methods that produced the research discussed in this paper are described in detail in Placer and Slobodchikoff (2004) and are summarized in Fig. 1. Three hundred (300) Gunnison's prairie dog alarm calls were selected from a large collection of tape recordings obtained over a 10-year period (1988–1997) at two separate prairie dog colonies described in Slobodchikoff *et al.* (1991). Details about this database can be found in Placer and Slobodchikoff (2004). One hundred (100) calls were vocalized in the presence of dogs, 100 in the presence of hawks, and 100 in the presence of coyotes. The alarm calls were converted into simple digitized time-domain recordings and processed to reduce background noise using the software package Cool Edit 2000. One of the authors (JP) created a system of computer programs to analyze these alarm calls using the high-performance numeric computation software called MATLAB and the computer language Java.

Each of the digitized alarm calls was divided into 25 equal-sized partitions. Figure 2 shows a digitized alarm call that has been divided into 25 partitions. The mean total times of the calls for the different predators were as follows: coyote: 0.193 ± 0.014 s (mean$\pm$SD), $N=100$ calls; dog:

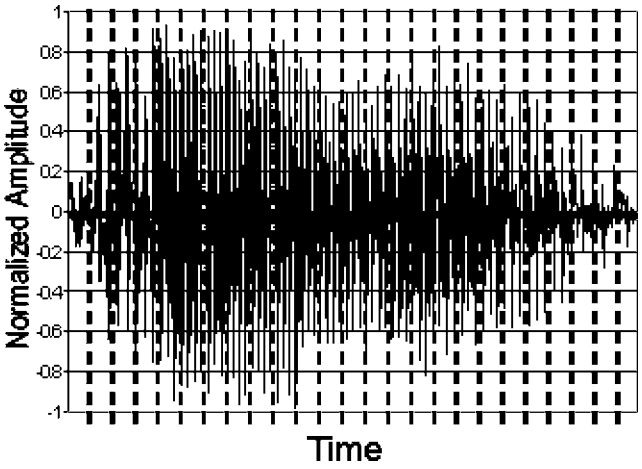


FIG. 2. A prairie dog call divided into 25 partitions.

0.134±0.020 s, N=100 calls; hawk: 0.084±0.007 s, N=100. The mean times of the individual partitions were the following. Coyote: 0.0148±0.003 s (mean±SD), N=2500 partitions; dog: 0.0103±0.0015 s, N=2500 partitions; hawk: 0.0065±0.0005 s, N=2500 partitions.

The fast fourier transform was then applied to the data points in each alarm call partition in order to produce a normalized power spectrum of the sound frequencies contained in that partition. The computer analysis software then used the eight most energetic frequencies of each power spectrum to determine the two most prominent formants (designated $F1$ and $F2$) contained in that alarm call partition, where formant values were rounded to the nearest multiple of 300 Hz. Figure 3 shows a typical power spectrum taken from one partition of a divided alarm call. The eight data points that lie above the horizontal line in the figure are the largest power values that are used to compute the two most energetic formants in this partition. The power values that are clustered together as contiguous data points are treated as part of a single formant. The lowest frequency in such a cluster is used to represent that formant. Formants are vocal tract resonances that appear as hills or raised sections in a power spectrum. A number of techniques for identifying the highest or most prominent formants were experimented with and the simple technique of using only the eight most energetic frequencies in each power spectrum produced the best results (Placer and Slobodchikoff, 2004).

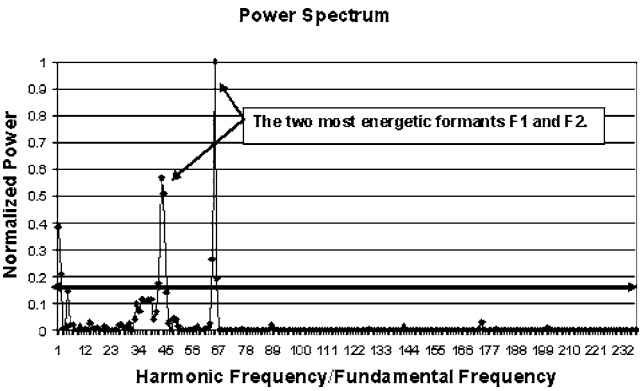


FIG. 3. The power spectrum of a partition of a prairie dog alarm call vocalized in the presence of a hawk. The two most energetic formants were used to identify distinct acoustic units. The values on the vertical axis are normalized because the absolute values of the power are not relevant. The values on the horizontal axis represent an index expressed in simple integers, derived by dividing the harmonic frequencies in the power spectrum of the partition by the fundamental frequency in that partition.

TABLE I. Sound frequency values (Hz) for the most common acoustic structures associated with alarm calls for coyotes.

Symbol No.	F1	F2
S6	0	3600
S15	0	3900
S13	300	3600
S14	300	3900
S31	300	4200
S17	300	3300
S23	300	600
S32	300	900
S62	2700	3600
S61	2700	3300
S16	3000	3900
S60	3000	3600
S5	3300	3900
S3	3300	4200
S11	3300	3600
S7	3600	3900
S8	3600	4500
S4	3600	4200
S12	3900	4200

Each of the two formants computed for each alarm call partition was represented by a single sound frequency. Thus, the acoustic structure associated with each partition was defined by two frequencies that represented that partition's two most energetic formants. Using this procedure, each alarm call would be divided into 25 equal-sized partitions and each partition in turn would be represented by two formant frequencies. In this way, each of the 300 alarm calls used in this study was translated into a sequence of 25 acoustic structures where each acoustic structure was represented by two simple frequencies (see Tables I–III).

Working with these translated alarm calls, the analysis software examined each acoustic structure in each alarm call. Each unique pair of frequencies that defined an acoustic unit

TABLE II. Sound frequency values (Hz) for the most common acoustic structures associated with alarm calls for domestic dogs.

Symbol No.	F1	F2
S47	300	1200
S32	300	900
S53	300	2700
S75	1200	3300
S120	1200	2400
S55	1200	3600
S3	3300	4200
S19	3300	4500
S5	3300	3900
S8	3600	4500
S18	3900	4800
S30	4200	5100

TABLE III. Sound frequency values (Hz) for the most common acoustic structures associated with alarm calls for red-tailed hawks.

Symbol No.	F1	F2
S13	300	3600
S32	300	900
S14	300	3900
S17	300	3300
S23	300	600
S24	600	3600
S39	600	3300
S62	2700	3600
S61	2700	3300
S162	3000	6000
S68	3300	6300
S70	3300	6000
S143	3600	5700
S157	3600	5400
S71	3600	6300
S154	3900	5700
S2	3900	6000
S144	4200	6000
S155	5100	6000
S161	5100	6300
S153	5400	6300

was assigned a unique symbol and that symbol was used to replace the acoustic unit. For example, assume the acoustic unit (300 Hz, 600 Hz) was encountered during this phase of the processing. If that exact pair of frequencies had been encountered previously, then a symbol would already have been used to represent them and that symbol would be used to replace the pair of frequencies. However, if this was the first time that particular pair of frequencies was encountered, a new symbol would be generated and the new symbol would be used to replace the pair. The names of the symbols always began with an "S" that was followed by a simple integer. Thus, if a new symbol was to be created and 25 symbols had already been generated, then the new symbol's name would be "S26." Using these techniques the alarm calls that had been translated into sequences of acoustic units were now translated into sequences of symbols. Shown below is an example of an alarm call vocalized in the presence of a coyote that has been translated into a sequence of 25 symbols:

S14S7S6S17S5S4S9S13S18S9S8S19S4S7S8S4S7

S6S6S4S19S4S19S8S13

Once the alarm calls were translated into sequences of symbols, the sequences were then searched for single symbols (monograms), pairs of contiguous symbols (digrams), and triplets of contiguous symbols (trigrams) that were found exclusively in the alarm calls associated with a particular predator species. Alarm calls found to contain a monogram, digram, or trigram that was uniquely associated with a given predator species could be unambiguously associated with that same predator species. Thus, monograms, digrams, and trigrams unique to a given predator species were used to classify the alarm calls by predator species. Only symbols that occurred at least 20 times in the target alarm calls were used in the classification process in order to guard against using infrequently occurring acoustic units that might simply be artifacts of the sound processing techniques utilized. Only symbols that occurred at least 50 times were considered to be common symbols that are used in the descriptions of the results below.

A similarity index was calculated based on counts of occurrence of the different acoustic symbols within the three types of calls. The index was calculated as

$$\text{similarity index} = \sum \frac{A \text{ or } B \text{ (whichever has the smaller number of occurrences)}}{A + B}$$

where A is the frequency of occurrence of a symbol in the calls elicited by one predator, B is the frequency of occurrence of the same symbol in the calls elicited by the other predator, and the index is summed over all of the symbols that are common to the calls of both predators.

III. RESULTS

Most of the common acoustic structures (>50 occurrences) were found in calls associated with all three predators, but the proportions of acoustic structures varied. In alarm calls elicited by coyotes, there was only a single acoustic structure (S15) that was unique to coyote calls (Fig. 4). In alarm calls elicited by domestic dogs, there was also only one acoustic structure (S120) that was unique (Fig. 5). In alarm calls elicited by red-tailed hawks, there were six acoustic structures (S153, S154, S155, S157, S161, and S162) that were unique to red-tailed hawk calls (Fig. 6). The proportion of occurrence of the symbols in the three types of calls was significantly different (coyote-hawk, $\chi^2 = 1770/2$ (Bonferroni correction) $= 885$, df $= 23$, $p < 0.001$; hawk-dog, $\chi^2 = 1374/2$ (Bonferroni correction) $= 687$, df $= 22$, $p < 0.001$; coyote-dog, $\chi^2 = 679/2$ (Bonferroni correction) $= 340$, df $= 19$, $p < 0.001$; χ^2 values were calculated on the basis of counts of occurrences). Based on the occurrence of acoustic structures in the calls, hawk calls and coyote calls had a similarity of 0.159, hawk calls and dog calls had a similarity of 0.134, and coyote calls and dog calls had a similarity of 0.218.

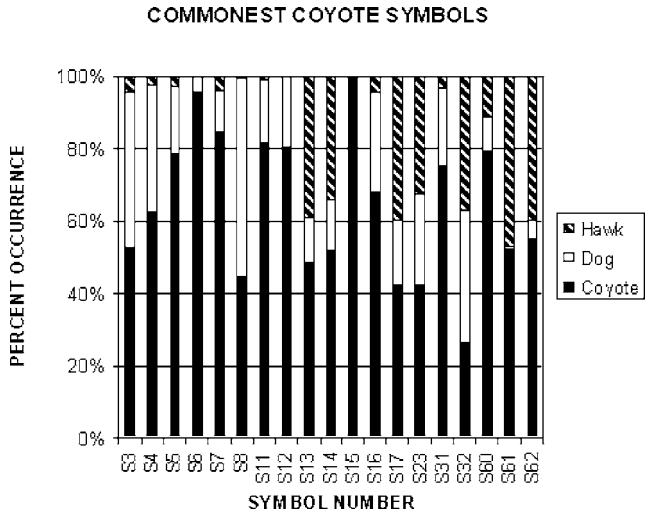


FIG. 4. The most common acoustic structures found within calls elicited by coyotes. Only one acoustic structure (S15) is unique to coyote calls.

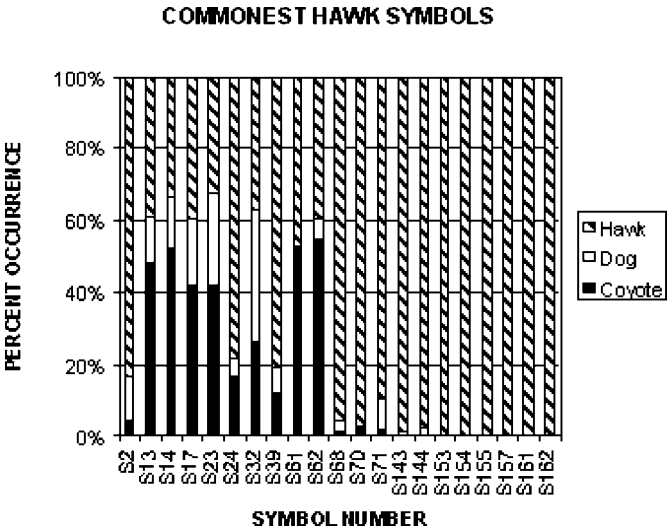


FIG. 6. The most common acoustic structures found within calls elicited by red-tailed hawks. Six acoustic structures (S153, S154, S155, S157, S161, S162) are unique to red-tailed hawk calls.

When digrams and trigrams were considered, there were more unique combinations. A number of these consisted of repetition of the same acoustic structure in contiguous partitions. Dog elicited calls had a unique trigram consisting of the same acoustic structure (S120) repeated in three adjacent partitions (S120S120S120), and another unique trigram consisting of an acoustic structure (S47) also repeated in three adjacent partitions (S47S47S47). Coyote elicited calls had three unique digrams and two unique trigrams. Of the digrams, one consisted of repetition of the same acoustic structure in two adjacent partitions (S6S6), and two consisted of two different acoustic structures in adjacent partitions (S16S60; S7S6). The two unique trigrams in coyote elicited calls were repetitions of the same acoustic structures in three adjacent partitions (S5S5S5; S7S7S7). Hawk elicited calls had eight unique digrams and one unique trigram. Of these, three consisted of repetitions of the same acoustic structures in adjacent partitions (S68S68; S70S70; S154S154), and five consisted of associations of different acoustic structures in adjacent partitions (S2S154; S68S70; S70S68; S154S157; S154S143). The unique trigram in hawk elicited calls consisted of a repetition of the same acoustic structure (S68) in three adjacent partitions (S68S68S68).

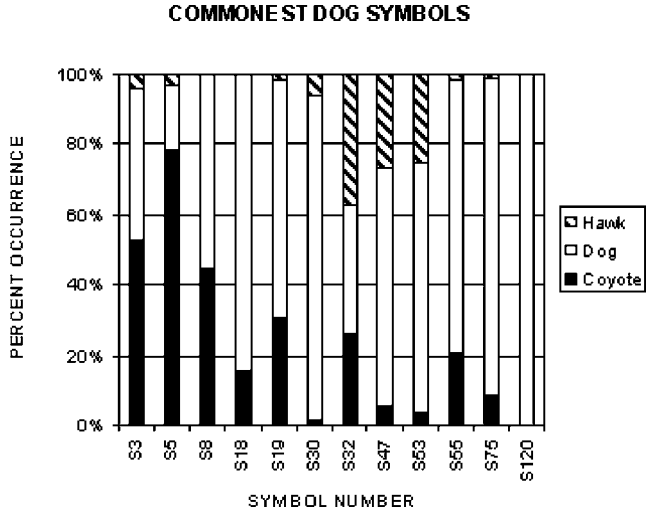


FIG. 5. The most common acoustic structures found within calls elicited by domestic dogs. Only one acoustic structure (S120) is unique to domestic dog calls.

Using the unique monograms, digrams, and trigrams, we subjected the original data to a classification analysis to determine how well these unique acoustic structures can identify the calls to the correct predator species. We found that 67% of the dog elicited calls were correctly classified as dog, 73% of the coyote elicited calls were correctly classified as coyote, and 99% of red-tailed hawk elicited calls were correctly classified as hawk, giving an overall classification average of 80% correct classification. When we plotted all of the acoustic structures that made up the unique monograms, digrams, and trigrams, we found that the acoustic structures associated with these unique grams for each predator species seemed to cluster in different regions (Fig. 7). Hawk-associated acoustic structures generally had higher formant frequencies for both $F1$ and $F2$ than the coyote- or domestic dog-associated frequencies, which had lower formant frequencies.

When we ranked the acoustic structures by ascending $F1$ frequency (Figs. 8–10), the acoustic structures produced a pattern in which the higher formant frequency increased in one symbol and then decreased in the following symbol for many of the symbols. This is a pattern similar to that seen in the structure of English vowels when the two lowest formants associated with each vowel are plotted (Fig. 11) [source of frequencies for English vowels (Kim, 1991)].

IV. DISCUSSION AND CONCLUSIONS

The same acoustic structures were found in calls associated with all three species of predator. However, the distribution of these acoustic structures was different for each of the three predators. This suggests two points. One is that the acoustic structures could be used as building blocks for larger units that might have meaning, in a way analogous to how phonemes are used in human languages to build words

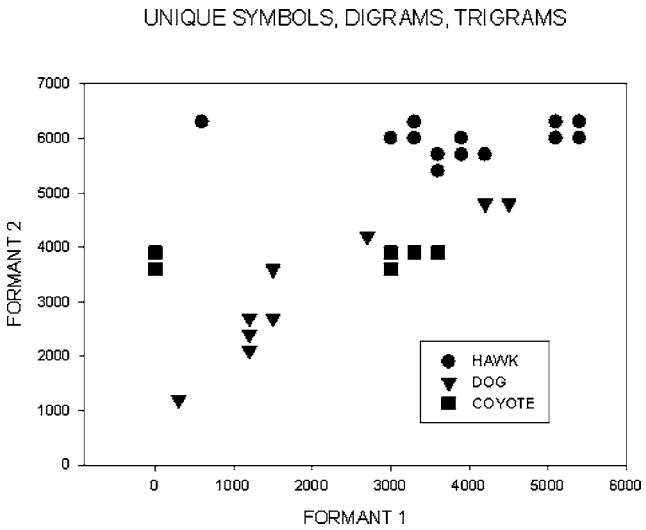


FIG. 7. Formant frequencies found in unique monograms, digrams, and trigrams associated with red-tailed hawk, coyote, and domestic dog alarm calls. The F1 and F2 frequency (Hz) values for each symbol that is part of a unique combination have been plotted, even though that symbol may be part of a unique digram or a unique trigram.

and sentences. This suggests that prairie dog alarm vocalizations might contain duality, i.e., that larger units of semantic significance are made up of smaller units that can be combined in different ways to produce many different larger units. This is one of the design features of language listed by Hockett (1960) for human and animal languages and has not been shown previously in animal vocalizations. The other point is that perhaps some or many acoustic structures might be used to express additional information beyond the species of predator present when an alarm call is vocalized. This hypothesis is consistent with the work of Slobodchikoff *et al.* (1991), who demonstrated that prairie dog calls contain information associated with size, shape, and color of different individuals within a predator species, and it is consistent with the work of Kiriazis (1991), who showed that prairie dog calls contain information related to direction and speed of approach of predators.

The vowellike acoustic structures in prairie dog alarm calls are perhaps not surprising, given that at least one non-

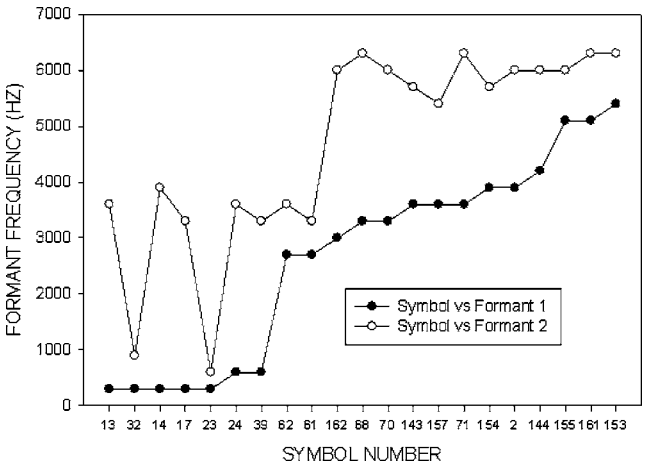


FIG. 8. Formant frequencies for calls elicited by red-tailed hawks, with the lower frequency ascending.

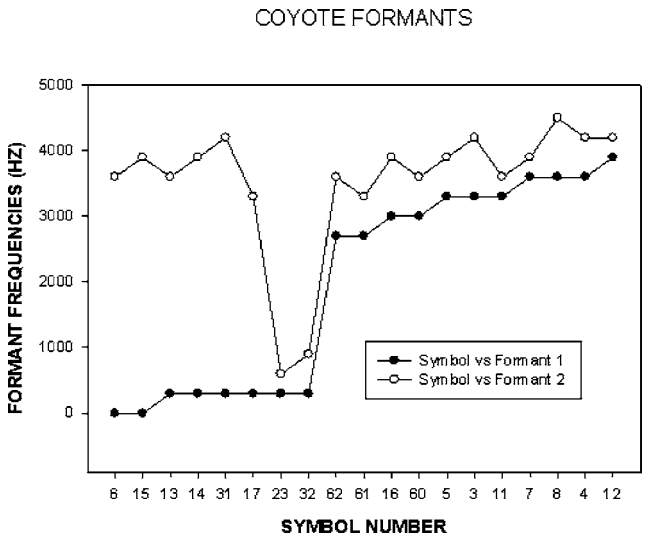


FIG. 9. Formant frequencies for calls elicited by coyotes, with the lower frequency ascending.

human primate, the baboon (*Papio cyncephalus*), is known to produce grunts that have acoustic properties similar to those of human vowels (Zhinkin, 1963; Andrew, 1976; Owren *et al.*, 1997). Also, other animals have been found to be able to discriminate among some human vowels. For example, cats (Dewson, 1964), chinchillas (Miller, 1977), baboons (Hienz and Brady, 1988), and pigeons (Hienz *et al.*, 1981) have been shown to be able to discriminate between some of the human vowels. Among the nonhuman primates, vervet monkeys (*Cercopithecus aethiops*) and Japanese macaques (*Macaca fuscata*) were able to discriminate between some of the common English vowels (Sinnott, 1989). Perhaps there is a limited number of ways in which information might be encoded in vocalizations that have multiple formants. However, we should point out that the duration of English vowels varies between 0.180 and 0.330 s (Lieberman and Blumstein, 1988), while the acoustic structures that we have identified vary between 0.0065 and 0.0148 s in du-

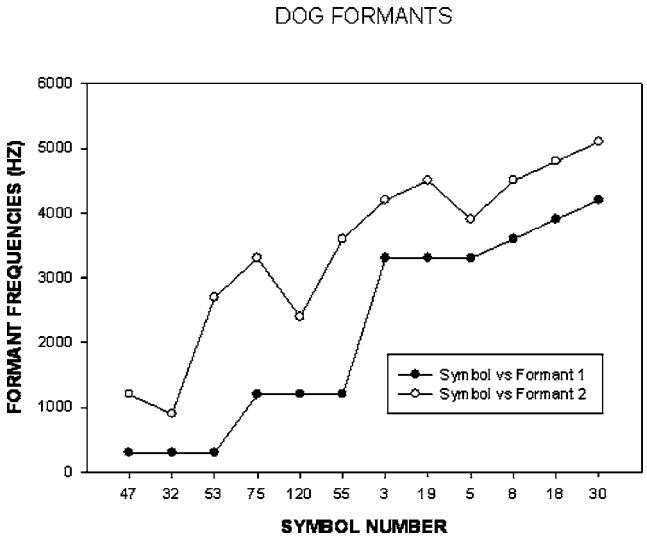


FIG. 10. Formant frequencies for calls elicited by domestic dogs, with the lower frequency ascending.

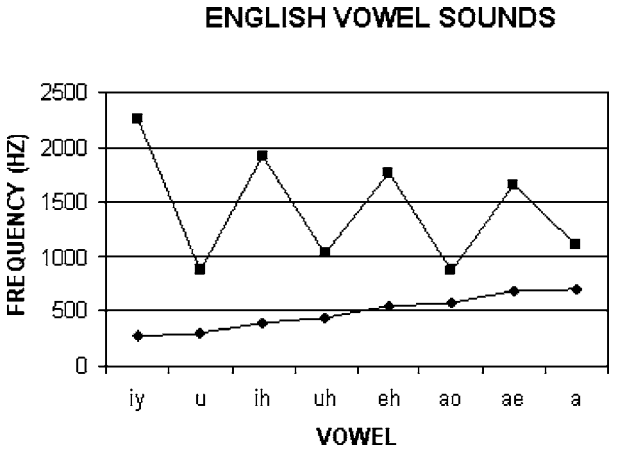


FIG. 11. Frequencies (Hz) for eight American English vowels, showing the lower frequency (F1) and the higher frequency (F2) for each vowel. The vowel symbols stand for the following: iy as in "beet," u as in "put," ih as in "bit," uh as in "Buddhist," eh as in "met," ao as in "cause," ae as in "bad," and a as in "cat." Sound frequencies of vowels from Kim (1991).

ration. Since the entire alarm calls of the prairie dogs vary between 0.084 and 0.193 s in duration and there is a considerable amount of information encoded within that total time interval, it is possible that prairie dogs might process acoustic information more rapidly than humans.

Somewhat more surprising is the acoustic structure of these vowellike sounds. In humans and other primates, vowellike sounds are produced through a combination of the first two formants, that is, the two formants with the lowest frequencies (Owren *et al.*, 1997). In prairie dog alarm vocalizations, the most accurate associations of alarm calls with predator species occurred when the two most dominant formants, the ones carrying the most acoustic energy, were used, rather than the first two. Perhaps this is an adaptation to the long-distance propagation of sound in the alarm call, since prairie dog calls travel over long distances, often through vegetation that can potentially filter out some frequencies (e.g., see Perla and Slobodchikoff, 2002).

ACKNOWLEDGMENTS

We thank Anne Epstein for digitizing calls from an extensive library of tapes of prairie dog vocalizations compiled over a span of 10 years.

Ackers, S. H., and Slobodchikoff, C. N. (**1999**). "Communication of stimulus size and shape in alarm calls of Gunnison's prairie dogs," Ethology **105**, 149–162.

Andrew, R. J. (**1976**). "Use of formants in the grunts of baboons and other nonhuman primates," Ann. N.Y. Acad. Sci. **280**, 673–693.

Beynon, P., and Rasa, O. A. E. (**1989**). "Do dwarf mongooses have a language?: warning vocalisations transmit complex information," S Afr. J. Commun. Disord. **85**, 447–450.

Cheney, D. L., and Seyfarth, R. M. (**1990**). *How Monkeys See the World* (Chicago U. P., Chicago).

Dewson, J. (**1964**). "Speech sound discrimination by cats," Science **144**, 555–556.

Evans, C. S., and Evans, L. (**1999**). "Chicken calls are functionally referential," Anim. Behav. **58**, 307–319.

Fischer, J., Hammerschmidt, K., and Todt, D. (**1995**). "Factors affecting acoustic variation in Barbary-macaque (*Macaca sylvanus*) disturbancy calls," Ethology **101**, 51–66.

Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (**2002**). "Acoustic features of male baboon loud calls: Influences of context, age, and individuality," J. Acoust. Soc. Am. **111**, 1465–1474.

Frederiksen, K. (**2005**). "A comparative analysis of alarm calls across the five species of North American prairie dogs," Ph.D. thesis, Northern Arizona University, Flagstaff.

Greene, E., and Meagher, T. (**1998**). "Red squirrels, *Tamiasciurus hudsonicus*, produce predator-class specific alarm calls," Anim. Behav. **55**, 511–518.

Hauser, M. D., and Fowler, C. A. (**1992**). "Fundamental frequency declination is not unique to human speech: Evidence from nonhuman primates," J. Acoust. Soc. Am. **91**, 363–369.

Hienz, R., and Brady, J. (**1988**). "The acquisition of vowel discriminations by non-human primates," J. Acoust. Soc. Am. **84**, 186–194.

Hienz, R., Sachs, M., and Sinnott, J. (**1981**). "Discrimination of steady state vowels by blackbirds and pigeons," J. Acoust. Soc. Am. **70**, 699–706.

Hockett, C. F. (**1960**). "Logical considerations in the study of animal communication," in *Animal Sounds and Communication*, edited by W. E. Lanyon and W. N. Tavolga (American Institute of Biological Sciences, No. 7, Washington D.C.), pp. 392–430.

Kim, C.-W. (**1991**). "Acoustical phonetics," in *The Linguistics Encyclopedia*, edited by K. Malmkjaer (Routlidge, London), pp. 1–10.

Kiriazis, J. (**1991**). "Communication and sociality in Gunnison's prairie dogs," Ph.D. thesis, Northern Arizona University, Flagstaff.

Kiriazis, J., and Slobodchikoff, C. N. (**2006**). "Perceptual specificity in the alarm calls of Gunnison's prairie dogs," Behav. Processes (in press).

Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge U. P., Cambridge).

Manser, M. B. (**2001**). "The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency," Proc. R. Soc. London, Ser. B **268**, 2315–2324.

Manser, M. B., Bell, M. B., and Fletcher, L. B. (**2001**). "The information that receivers extract from alarm calls in suricates," Proc. R. Soc. London, Ser. B **268**, 2485–2491.

Miller, J. D. (**1977**). "Perception of speech sounds in animals: evidence for speech processing by mammalian auditory mechanisms," in *Recognition of Complex Acoustic Signals*, edited by T. Bullock (Springer, Berlin), pp. 49–58.

Owings, D. H., and Hennessy, D. F. (**1984**). "The importance of variation in sciurid visual and vocal communication," in *The Biology of Ground-dwelling Squirrels*, edited by J. O. Murie and G. R. Michener (U. Nebraska P., Lincoln), pp. 169–200.

Owren, M. J., and Bernacki, R. H. (**1988**). "The acoustic features of vervet monkey alarm calls," J. Acoust. Soc. Am. **83**, 1927–1935.

Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). "The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions," J. Acoust. Soc. Am. **101**, 2951–2963.

Perla, B. S., and Slobodchikoff, C. N. (**2002**). "Habitat structure and alarm call dialects in Gunnison's prairie dog (*Cynomys gunnisoni*)," Behav. Ecol. Sociobiol. **13**, 844–850.

Placer, J., and Slobodchikoff, C. N. (**2000**). "A fuzzy-neural system for identification of species-specific alarm calls of Gunnison's prairie dogs," Behav. Processes **52**, 1–9.

Placer, J., and Slobodchikoff, C. N. (**2001**). "Developing new metrics for the investigation of animal vocalizations," Intell. Autom. Soft Comput. **7**, 1–11.

Placer, J., and Slobodchikoff, C. N. (**2004**). "A method for identifying sounds used in the classification of alarm calls," Behav. Processes **67**, 87–98.

Riede, T., and Zuberbuhler, K. (**2003**). "Pulse register phonation in Diana monkey alarm calls," J. Acoust. Soc. Am. **113**, 2919–2926.

Seyfarth, R. M., and Cheney, D. L. (**1984**). "The acoustic features of vervet monkey grunts," J. Acoust. Soc. Am. **75**, 1623–1628.

Sinnott, J. M. (**1989**). "Detection and discrimination of synthetic English vowels by Old World monkeys (*Cercopithecus, Macaca)* and humans," J. Acoust. Soc. Am. **86**, 557–565.

Slobodchikoff, C. N. (**2002**). "Cognition and communication in prairie dogs," in *The Cognitive Animal*, edited by M. Bekoff, M. C. Allen, and G. Burghardt (MIT, Cambridge, MA), pp. 257–264.

Slobodchikoff, C. N., Kiriazis, J., Fischer, C., and Creef, E. (**1991**). "Semantic information distinguishing individual predators in the alarm calls of Gunnison's prairie dogs," Anim. Behav. **42**, 713–719.

Verdolin, J., and Slobodchikoff, C. N. (**2002**). "Vigilance and predation risk

in Gunnison's prairie dogs," Can. J. Zool. **80**, 1197–1203.
Vosburgh, T. C., and Irby, L. R. (**1998**). "Effects of recreational shooting on prairie dog colonies," J. Wildl Dis. **62**, 363–372.
Zhinkin, N. I. (**1963**). "An application of the theory of algorithms to the study of animal speech," in *Acoustic Behaviour of Animals*, edited by R. G. Busnel (Elsevier, Amsterdam), pp. 132–180.
Zuberbuhler, K. (**2000**). "Referential labeling in Diana monkeys," Anim. Behav. **59**, 917–927.
Zuberbuhler, K. (**2001**). "Predator-specific alarm calls in Campbell's guenons," Behav. Ecol. Sociobiol. **50**, 414–422.

Using an artificial neural network to classify black-capped chickadee (*Poecile atricapillus*) call note types

Michael R. W. Dawson, Isabelle Charrier, and Christopher B. Sturdy[a)]
Department of Psychology, Centre for Neuroscience, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

(Received 9 May 2005; revised 17 October 2005; accepted 2 March 2006)

The "chick-a-dee" call of the black-capped chickadee (*Poecile atricapillus*) contains four note types, A, B, C, and D that have important functional roles. This provides strong motivation for studying the classification of acoustic components of the call into different note types. In this paper, the spectrograms from a sample of A, B, and C notes (370 in total) were each described as a set of 9 summary features. An artificial neural network was trained to identify note type on the basis of these features, and was able to obtain better than 98% accuracy. An internal analysis of this network revealed a distributed code in which different hidden units generated high activities to different subsets of notes. By combining these different sensitivities, the network could discriminate all three types of notes. The performance of this network was compared to a discriminant analysis of the same data. This analysis also achieved a high level of performance (95%). A comparison between the two approaches revealed some striking similarities, but also some intriguing differences. These results are discussed in terms of two related issues: developing a research tool for note classification, and developing a theory of how birds themselves might classify notes. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2189028]

PACS number(s): 43.80.Ka, 43.80.Lb, 43.80.Jz [JAS] Pages: 3161–3172

I. INTRODUCTION

Songbirds are a popular model for the study of vertebrate acoustic communication (e.g., Slater, 2003). One reason for this is that the circumstances under which songbirds learn song from a model can be viewed as being analogous to human language learning (Doupe and Kuhl, 1999; Goldstein *et al.*, 2003). That said, song is not the only class of vocalizations that songbirds learn from a model. For example, songbirds also learn *calls* that are used in the context of specific behavioral interactions. As a result, more general studies of acoustic signaling and communication in songbirds are broadening their scope to include examinations of learned calls and other vocalizations in addition to learned songs (e.g., Hailman and Ficken, 1996; Hughes *et al.*, 1998; Vicario *et al.*, 2002). The purpose of this paper is to report the results of one such examination of the notes of a learned call of a small songbird native to North America, the black-capped chickadee (*Poecile atricapillus*). We show that when the notes of this call are represented as small sets of acoustic features, a nonlinear statistical method (artificial neural network) and a linear statistical method (linear discriminant analysis) are both capable of using these features to correctly classify notes into different note types.

The black-capped chickadee is one songbird that has provided a great deal of information about learned calls. One of the learned calls for this species is the "chick-a-dee" call, for which the species was named (Ficken *et al.*, 1978). Chick-a-dee calls contain four note types, labeled A, B, C, and D. Chick-a-dee calls are interesting because while their component notes are generated in a fixed order (A→B→C→D), in a particular call note types can be either repeated or omitted. As a result, a seemingly infinite variety of different chick-a-dee calls can be produced (e.g., ACCCCD, ABDDD; see Figs. 1 and 2 for examples of calls and the constituent note types, respectively). Indeed, the combinatorial nature of the chick-a-dee call has been used to draw interesting analogies between it and human speech (Hailman, 1985).

The different note types from which the chick-a-dee call is composed are important for a number of reasons. First, chickadees appear to vary the note-type composition of their chick-a-dee calls depending on the particular context (e.g., Ficken *et al.*, 1994; Gaddis, 1985; Smith, 1972), suggesting that different distributions of note types within calls convey different information. Second, chickadees fail to respond in a species-typical fashion to the playback of chick-a-dee calls that have an atypical syntax. Reduced responsiveness can be caused by gross syntactical violations, such as complete reversal of call syntax (DCBA rather than ABCD); (Charrier and Sturdy, 2005), or from more subtle syntactical violations (CACACA instead of AAACCC; Clucas *et al.*, 2004). Third, even more subtle alterations of notes within chickadee calls can cause birds to treat calls as atypical. Charrier and Sturdy (2005) found that chickadee calls with the correct syntax (ABCD) but in which each individual note was time-reversed, thus altering both amplitude and frequency modulation patterns of the notes within the calls, led to significantly reduced responding compared to normal calls. Fourth, black-capped chickadees are themselves able to discriminate and categorize the different note types of the chick-a-dee call, although they sometimes confuse adjacent note types (Sturdy *et al.*, 2000). Finally, different chick-a-dee call note

[a)]Electronic mail: csturdy@ualberta.ca

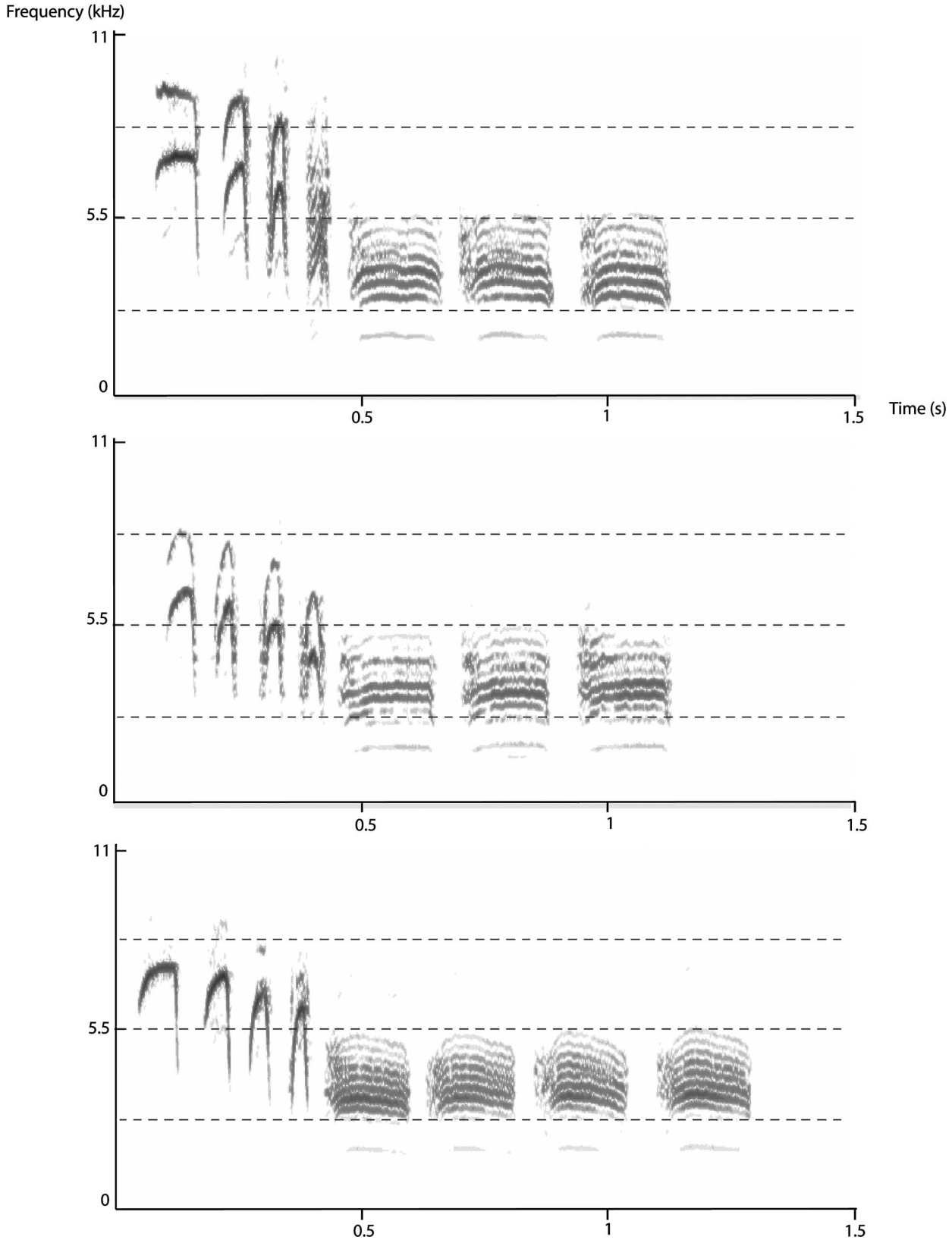


FIG. 1. Sound spectrograms of black-capped chickadee chick-a-dee calls from 3 different birds. The note composition of these calls is AABCDDD (top) AABBDDD (middle), and AAABDDDD (bottom), respectively. (Hamming window size=1024 points, frequency precision=21 Hz).

types have been associated with conveying specific information. For instance, the D note appears to be critical for birds being able to distinguish between chick-a-dee calls from members of their flock from the same kind of calls produced by members of a foreign flock (Mammen and Nowicki, 1981; Nowicki, 1983). The C note appears to be used to indicate the location or availability of food sources (Freeberg and Lucas, 2002). Taken together, it seems clear that the

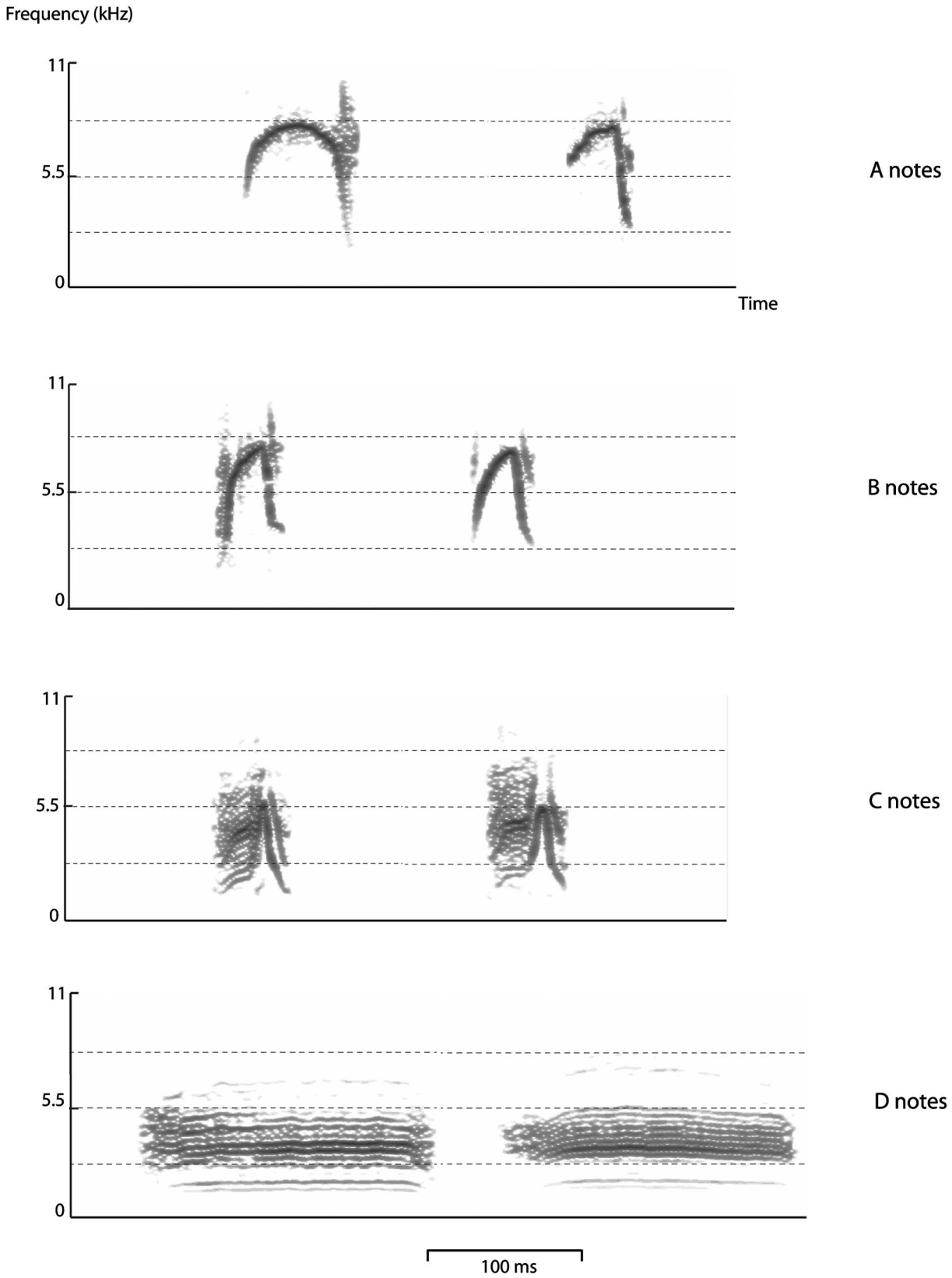


FIG. 2. Sound spectrograms of the four different note types found in black-capped chickadee chick-a-dee calls: A, B, C, and D notes. Exemplars for each note type from 2 different birds. (Hamming window size=1024 points, frequency precision=21 Hz).

accurate perception of individual note types in the chick-a-dee call is a crucial and fundamental building block of chickadee's acoustic communication system.

The fact that the constituent notes of the chick-a-dee call have important functional roles provides strong motivation for studying the classification of acoustic components of the call into one of the four different note types. Nowicki and Nelson (1990) describe such note classification as identifying natural categories in the acoustic signal. They point out that one critical research question is to determine how black-

capped chickadees themselves identify such natural categories. A related and equally crucial question, studied in detail by Nowicki and Nelson, is how researchers themselves can identify these natural categories to aid in studying the properties of the call.

The traditional approach to this classification problem is exemplified by bioacousticians, who examine the visual representation of a vocalization (e.g., a chick-a-dee call) in the form of a sound spectrogram, and use visual features of the call to identify and classify its component note types. This technique has been used by several researchers to classify the notes of vocalizations from several songbird species [e.g., mountain chickadee, *Poecile gambeli*, chick-a-dee call notes, Bloomfield *et al.* (2004); Carolina chickadee, *P. carolinensis*, chick-a-dee call notes, Bloomfield *et al.* (2005); Black-capped chickadee chick-a-dee call notes, Charrier *et al.* (2005); Ficken *et al.* (1978); and Nowicki and Nelson (1990); zebra finch, *Taeniopygia guttata*, song notes, Sturdy *et al.* (1999b)]. Some of these visual classifications have been verified with the species that produces them via operant conditioning discrimination tasks (e.g., Sturdy *et al.*, 1999a, 2000), providing a necessary validation to the visual classification. That said, not all visually-guided classifications of animal vocalizations have been verified empirically, and depending on the species under study, many laboratory-based verifications may be impractical or impossible to conduct.

A less subjective approach to identifying natural categories of notes is to take spectrograms as raw data for traditional statistical analyses. Nowicki and Nelson (1990) provided one example of this approach. They took spectrograms representing 240 different notes (60 notes from each of 4 different birds), and cross-correlated all possible pairs of notes from this dataset. They then took the resulting correlation matrix and analyzed it with multidimensional scaling (MDS). MDS is a technique that provides a graphical representation of data, where data instances are represented as points in a multidimensional space (Kruskal and Wish, 1978). The more similar two instances of data are to one another, the nearer are their corresponding points in the space. Nowicki and Nelson found that a two-dimensional space provided an excellent account of their data. The MDS plot revealed a good segregation of each note type into different clusters with little overlap. This result was particularly encouraging because Nowicki and Nelson included D notes in addition to A, B, and C notes in their analysis. Because D notes are dramatically different than the other three in terms of acoustic properties, their inclusion possibly made differences between A, B, and C notes less apparent.

Another approach, also explored by Nowicki and Nelson (1990), is to take raw spectrograms and summarize them into a more tractable set of summary features. These summary features can then be used as input to a variety of statistical methods as an aid to note classification.

For example, Nowicki and Nelson (1990) described each of their non-D notes, that is A, B, and C notes (a sample of 185 stimuli), as a small set of different acoustic features that were derived from the spectrogram of the note. These features included such measurements as the total note duration, the starting, peak, and ending frequencies of the note, and so on. They created a matrix of correlations between every pair of notes on the basis of these features, and then performed principal components analysis on this matrix. They were able to account for over 68% of the variance in the correlation matrix with three principal components. They then described each of the 185 notes in terms of their loadings on these three factors, and processed this stimulus representation with *k*-means cluster analysis. They found that three clusters provided the most parsimonious account of the data.

A classification matrix based on the *k*-means solution indicated that this statistical approach generated a 77.3% agreement with visual classification of notes. That is, 143 of the notes were given the same classification that was assigned after a visual inspection of the spectrogram. With respect to misclassifications, 8 of the 71 A notes were classified as B notes, and 6 were classified as C notes. Similarly, 1 of the 65 B notes was classified as an A note, and 22 were classified as C notes. Finally, 5 of the 49 C notes were classified as B notes.

On the one hand, Nowicki and Nelson (1990) demonstrated that it is possible to use a variety of statistical analyses to classify notes in such a way that there is a high amount of agreement between statistical classifications and visual classifications. On the other hand, their research raises two questions that deserve additional study. First, is it possible to improve the already good fit between statistical and visual classifications by exploring alternative statistical methods? Second, is it possible to explore some form of statistical classification that might provide additional insight into how the birds themselves might perform note classification?

The purpose of this paper is to explore both of these questions by examining the ability of an artificial neural network to classify notes when these notes are described as a small set of features that have been derived from a spectrogram. The artificial neural network that we investigated is called a multilayer perceptron. In comparison to traditional statistics, this kind of network can provide classification power because it determines an optimal nonlinear combination of features. This kind of technique is also of interest because many researchers have argued that it is more biologically plausible than other techniques (Clark, 1989; McClelland *et al.*, 1986; Schneider, 1987; Smolensky, 1988). As a result, if one can produce an artificial neural network that is capable of classifying note types, this network might suggest general ways in which note classification is performed by the birds themselves.

This paper proceeds as follows. First we describe the methods used to obtain and summarize a set of A, B, and C notes from the chick-a-dee call. Second, we provide a brief overview of the multilayer perceptron, and examine its ability to classify this new set of notes. Third, to compare this approach to more traditional statistical approaches, we explore the ability of discriminant analysis to classify this same set of stimuli. Finally, we use the results of these two sets of analyses to discuss some issues related to the statistical and to the biological classification of chick-a-dee call notes.

II. METHODS

The animals, recording procedure, acoustic measurements, and analyses have been published previously (Charrier *et al.*, 2004) and will be outlined only briefly here.

A. Animals

Six male and four female adult black-capped chickadees were captured during the winters of 2002 and 2003 from two locations in Alberta. Birds were housed in individual cages, had *ad libitum* access to food and were maintained on natural day–night cycle typical for the season in Edmonton and at approximately 20 °C.

B. Recording procedure, signal acquisition, and analysis

Each bird was recorded until we had a sample of at least 20 calls for each bird. Recordings (effective frequency range: 90–12 000 Hz) were digitized at 44 100 Hz, 16 bit samples/s using a 16 bit DartDisk Direct-to-Disk recorder (Engineering Design, Belmont, Massachusetts, USA). Calls were analyzed using SIGNAL version 4.0 (Engineering Design 2003).

From a sample of 100 calls, 370 non-D (i.e., A, B, and C) notes were visually classified and measured (205 A notes, 134 B notes, and 31 C notes) as was described in a previous paper (Charrier *et al.*, 2004). Nine acoustic features, partially based on the methods described in Nowicki and Nelson (1990), were measured including: start frequency (SF in hertz), peak frequency (PF in hertz), end frequency (EF in hertz) all of which were measured on a digital spectrogram [window size=1024 points, frequency precision=43 Hz; Fig. 2(a)] using a cutoff amplitude of −35 dB relative to the peak amplitude in the note. We also measured the loudest frequency (F_{max} in hertz) using a power spectrum [average window size=4096 points, frequency precision=11 Hz; Fig. 2(c)]. The duration measurements included were ascending duration (AD in milliseconds), descending duration (DD in milliseconds), and total duration (TD in milliseconds), and were measured on a digital spectrogram [window size =256 points, temporal precision=5.8 ms; Fig. 2(b)].

Two other measurements of frequency modulation were also made: the slope of the ascending frequency modulation [FM_{asc} in hertz per milliseconds following the formula (PF–SF)/AD] and the slope of the descending frequency modulation [FM_{desc} in hertz per milliseconds following the formula (EF–PF)/ DD].

C. Data preprocessing

There is a tremendous range in the values of the different input features that were described earlier. This raised a concern about how the artificial neural networks would perform on these raw data, because the processing units of the networks must "squash" input values into a narrow range of from zero to one. We decided to preprocess the data in such a way that the range of the input features would be reduced, but the essential characteristics of the raw data would be unchanged. This was accomplished by normalizing the value of each feature. That is, for each feature we took the 370 features values and converted them into z-scores. This preprocessed data served as the input to both of the different classification tools that are described in the following.

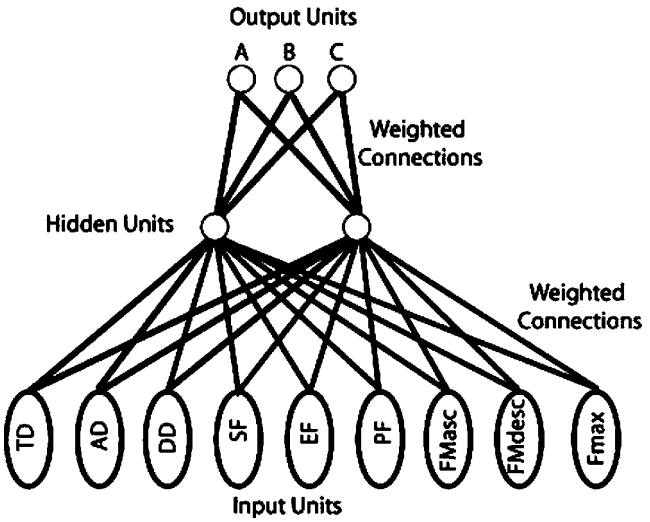


FIG. 3. An example multilayer perceptron. The network trained in the current manuscript is illustrated in which two hidden units processed the values of nine input features in order to assign a note to one of three different output categories.

D. Note classification by artificial neural networks: A brief introduction to artificial neural networks

An artificial neural network is a computer simulation of a "brain-like" system of interconnected processing units (see Fig. 3). In general, such a network can be viewed as a multiple-layer system that generates a desired response to an input stimulus. The stimulus is provided by the environment, and is encoded as a pattern of activity in a set of input units. The response of the system is represented as a pattern of activity in the network's output units. Intervening layers of processors in the system, called hidden units, detect features in the input stimulus that allow the network to make a correct or appropriate response.

The behavior of a single processing unit in this system can be characterized as follows: First, the unit computes the total signal being sent to it by other processors in the network. Second, the unit adopts a particular level of internal activation on the basis of this computed signal. Third, the unit generates its own signal, which is based on its level of internal activity, and sends this signal on to other processors.

The signal sent by one processor to another is transmitted through a weighted connection, which amplifies or attenuates a numerical signal being sent through it. This is accomplished by multiplying the signal's value by the weight associated with the connection. The weight defines the nature and strength of the connection. For example, inhibitory connections are defined with negative weights, and excitatory connections are defined with positive weights. Strong connections have strong weights (i.e., the absolute value of the weight is large), while weak connections have near-zero weights.

The pattern of connectivity in an artificial neural network (i.e., the network's entire set of connection weights)

defines how signals flow between the processors. As a result, a network's connection weights are analogous to a program in a conventional computer (Smolensky, 1988). However, in contrast to a conventional computer, an artificial neural network is not given an explicit program to perform some desired task. Instead, the network is taught to do the task.

Multiple layer networks (i.e., networks that contain one or more layers of hidden units), are frequently trained with a learning procedure called the generalized delta rule (Rumelhart *et al.*, 1986). One starts with a network that has small, randomly assigned connection weights. The network is then taught by presenting it a set of training patterns, each of which is associated with a known correct response. Each pattern is presented to the network's input units, and (on the basis of its existing connection weights) the network generates a response to it. An error term for each output unit is calculated by measuring the difference between the desired response of the unit and its actual response.

Error is used to modify connection weights by sending it backwards through the network. Once the error term for each output unit has been calculated, the weights of the connections directly attached to each output unit are modified. Then the output units send their error as a signal through the modified connections to the next layer of hidden units. Each hidden unit computes its overall error by treating the incoming error signals as net input (i.e., a hidden unit's total error is the sum of the weighted error signals that it is receiving from each output unit). Once a hidden unit has computed its overall error, then the weights of the connections that are directly attached to it can be modified. This process can be repeated, if necessary, to send error signals to the next layer of hidden units, and stops once all of the connections in the network have been modified. By repeating this procedure a large number of times for each pattern in the training set, the network's response errors for each pattern can be reduced to near zero. At the end of this training, the network will have a very specific pattern of connectivity (in comparison to its random start), and will have learned to perform a particular stimulus/response pairing.

E. Network architecture

The network trained in this study was a multilayer perceptron with 9 input units, 2 hidden units, and 3 output units (see Fig. 3). Each input unit was used to represent one of the 9 spectrogram features; the input data for network training is identical to that used as predictors in the discriminant analysis reported later. Each input unit was connected to each hidden unit, and each hidden unit was connected to each output unit. There were no direct connections between input units and output units. The hidden units and the output units used the sigmoid-shaped logistic function to convert their incoming signals into internal activity that ranged between 0 and 1.

The network was trained to activate an appropriate output unit to classify a 9-feature input pattern. If the input pattern was an "A" note, then the network was trained to turn the first output unit on (and to turn the other two output units off). If the input pattern was a "B" note, then the network was trained to turn the second output unit on. If the input pattern was a "C" note, then the network was trained to turn the third output unit on.

TABLE I. Classification of three types of input notes (first column) into three note classes by an artificial neural network.

	Note type according to artificial neural network			
Note type from visual examination	A	B	C	% correct
A	203	3	0	98.5
B	1	132	1	98.5
C	0	0	30	100
Total	195	144	31	98.6

F. Training

The ANN was trained using generalized delta rule. Initially, all connection weights were randomly assigned values between 0.1 and −0.1. The biases of the output unit and the two hidden units (which are analogous to neuronal thresholds) were initialized to a value of 0. The network was trained with a learning rate of 0.1 and zero momentum. Weights and biases were updated after the presentation of each training pattern. Each input pattern was presented once during a training epoch. Prior to each epoch, the order of presentation of the individual patterns was randomized. Pilot studies had indicated that the ANN never converged to a perfect solution to this problem—that is, it never learned to classify all of the notes with 100% accuracy. As a result, the achievement of perfect performance could not be used as a criterion to terminate training. Instead, the network was trained to the point that total network error (i.e., the sum of squared errors, with the sum taken over all of the output units and all of the training patterns) reached an acceptably low level. For the network reported in this paper, this was accomplished after training it for 2000 epochs. At the end of this amount of training, network SSE was 6.02. As is shown in the following, this level of SSE is associated with a very satisfactory degree of classification performance.

III. RESULTS

A. Network performance

The first question of interest at the end of network training concerns the amount of agreement between the visual classification of note types and the classifications performed by the network. In order to make this comparison, we first converted the activation patterns in the network's three output units into a specific note classification (A, B, or C). This was easily accomplished, because for each of the input notes, one of the output units generated activity of 0.9 or higher, while the other output units generated activity of 0.1 or lower. We classified the network's response by assigning it the note name associated with the output unit that was producing high activity. Table I provides the relationship between the visual classifications and the network classifica-

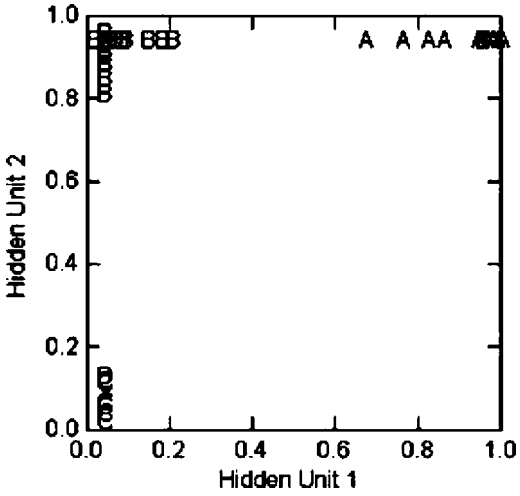


FIG. 4. The distribution of note types as a function of the activity produced in each hidden unit. See the text for details.

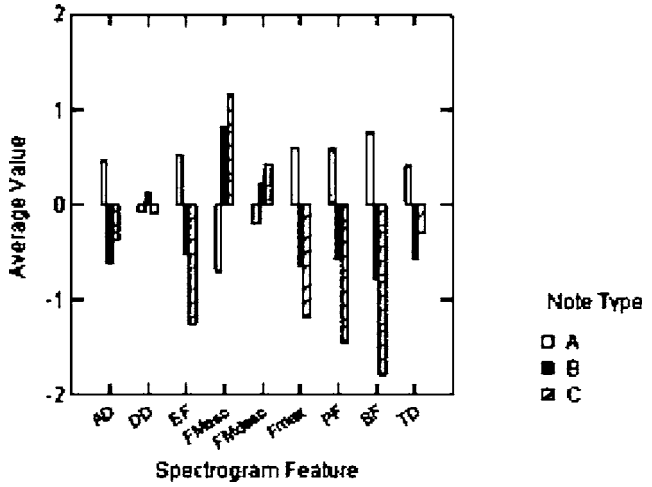


FIG. 5. Average values of the normalized features for each of the three note types. See the text for details.

tions. As can be seen from that table, the network was extremely accurate, misclassifying only 5 of the 370 stimuli for an accuracy level of 98.6%.

How is the network capable of performing at this high level? Some information relevant to answering this question can be obtained by examining the responses of the hidden units to all of the input patterns. In this network, the role of the hidden units is to map each pattern from a nine-dimensional space defined by the input features into a two-dimensional space defined by hidden unit activity. That is, each pattern can be described as a point in a two-dimensional space in which the coordinates of the point are provided by the activities produced by the pattern in the two hidden units.

A graph of this space is provided in Fig. 4. In this graph, each input pattern is represented as a letter that provides the note's type according to visual classification. By examining this graph, it appears that the two hidden units have very well defined functional roles. Hidden unit 1 is an "A-note detector," producing high activity to notes that belong to the A category, and producing little activity to other notes. In contrast, hidden unit 2 is a "not-C note detector," producing high activities to notes that do not belong to the C class, and producing near zero activities to notes that do belong to this class. These two hidden units classify notes as follow, then: if hidden units 1 and 2 are both on, the note is assigned to class A. If hidden units 1 and 2 are both off, then the note is assigned to class C. If hidden unit 1 is off, and hidden unit 2 is on, then the note is assigned to class B. This classification scheme works for 365 of the 370 notes in the training set.

How do these two hidden units respond in this fashion to discriminate between the different note types? One approach to answering this question requires an examination of the average feature profile of each note type in relation to the size of the connection weights that feed into each hidden unit.

Figure 5 provides the first type of information that is required for this interpretation. It is a graph of the average feature values for each of the three different note types in the training set. Two important observations follow from this graph. First, the feature profile for the average A note is markedly opposite to the feature profiles for the other two note types—the bars for each of its features point in an opposite direction than do the corresponding bars for the other two note types. Indeed, when the profile of the A note in Fig. 5 is correlated with the B note profile, the result is −0.99. When the A note profile is correlated with the C note profile, the result is −0.94.

The second important observation from this graph is the similarity of the profiles for B and C notes. For almost all of the features, the bars for these two notes presented in Fig. 5 point in the same direction. The correlation between these two profiles is 0.89. The difference between the two notes, as indicated in Fig. 5, would appear to be the fact that for many of the features the average C note has a more extreme value than does the average B note.

In order for a hidden unit to differentially respond to the various note types, the connection weights that feed into the hidden unit must transform the input features in such a way that the total signal is high for some notes (causing the hidden unit to activate) and is low for other notes (preventing the hidden unit from activating). Consider hidden unit 1, which generates high activity to A notes, and which generates little activity to B or C notes. The connection weights for this unit are presented in Table II. The bias for this unit, which is analogous to the unit's threshold, is also presented in Table II. The total signal coming into this unit from a pattern is equal to the sum of the input values of the features, after these input values have been multiplied by the weights. This signal is then combined with the bias, and passed as an input into the logistic equation $[f(net_i)=1/(1+\exp(-net_i+\theta_j))$, where net_i is the total signal and θ_j is the bias] to convert net input into an activation value that falls in the range from 0 to 1.

The connection weights for hidden unit 1 are arranged in such a way that A note features tend to generate high positive net inputs, while features from the other two note types generate strong negative net inputs. For example, if the average feature values of an A note from Fig. 5 are converted into a net input using the weights in Table II, the result is 31.36. In contrast, the average feature values of a B note produce a net

TABLE II. Connection weights from the 9 input units (features) to the two hidden units of the network. The bias of each hidden unit—analogous to a threshold—is also provided.

Input unit	Hidden unit	
	1	2
Unit bias	6.55	28.33
TD	−2.20	−4.08
AD	−0.05	−0.57
DD	−4.24	10.82
SF	33.88	29.80
PF	8.73	−9.36
EF	−16.20	1.94
FM_{asc}	−7.11	4.41
FM_{desc}	1.96	−6.06
F_{max}	−2.80	0.04

input of −19.08, and the average feature values of a C note produce a net input of −49.18. Clearly, if one views the weights feeding into this hidden unit as a vector, then the orientation of this vector is far more similar to the orientation of the vector an average set of A note features than to the orientation of the vectors representing B and C notes.

This is not to say that all of the input features have equal importance in determining the response of hidden unit 1. The following regression equation includes only those input features that were found to be significant predictors of hidden unit 1's activity, and accounted for 80.1% of the variance in this activity ($F_{5364}=293.83, p<0.0001$):

$$\mathrm{H1}_{\mathrm{activity}} = 0.375\ \mathrm{SF} - 0.114\ \mathrm{EF} - 0.114\ \mathrm{FM}_{\mathrm{asc}} + 0.059\ F_{\max} - 0.039\ \mathrm{TD} + 0.550.$$

From Fig. 5, it can be seen that all of these features correspond to average feature values that demonstrate the largest differences between A notes and the other two note types. The one exception is peak frequency (PF). However, this feature has a very large correlation with F_{max} ($r=0.964$), and because of this redundancy the inclusion of PF in the regression equation would not provide a significant increase in predictive power.

A similar analysis can be provided for hidden unit 2, whose connection weights and bias are also presented in Table II. This hidden unit has a different function than hidden unit 1: it generates low responses to C notes, and high responses to the other two note types. Again, this is because the orientation of its weight vector is more similar to the orientation of the average feature vectors for A and B notes, and much less similar to the orientation of the average feature vector for a C note. When average features for A and B notes are combined with the Table II weights to produce hidden unit 2 net input, the results are 42.19 and 15.84, respectively. In contrast, when the average features of a C note are used to determine net input, the result is −10.54.

Again, not all nine input features are equally important for determining the activity of this hidden unit. The following regression equation includes only those input features that were found to be significant predictors of hidden unit 1's activity, and accounted for 47.0% of the variance in this activity ($F_{6363}=53.60, p<0.0001$):

$$\mathrm{H2}_{\mathrm{activity}} = 0.332\ \mathrm{SF} - 0.214\ F_{\max} + 0.152\ \mathrm{PF} - 0.098\ \mathrm{TD} + 0.087\ \mathrm{FM}_{\mathrm{asc}} + 0.042\ \mathrm{DD} + 0.916.$$

From Fig. 5, it can be seen that many of these features correspond to average feature values that demonstrate the largest differences between A notes and the other two note types. Indeed, there are numerous features shared in both regression equations. Interestingly, this equation includes PF and not EF, these two features have a high correlation with each other ($r=0.672$), and EF also has a high correlation with F_{max} ($r=0.68$). This equation also includes an additional variable not found in the first regression equation, DD.

B. Note classification using discriminant analysis

The previous section demonstrated that a neural network is able to classify chick-a-dee call notes, described in terms of a small set of spectrogram features, with a very high degree of accuracy. This was accomplished by having two different hidden units separate one note type from the other two (in one case, A from B and C; in the other, A and B from C), and by then combining the activities of these hidden units to segregate all three note types. The hidden units operated by weighting input features in such a way to maximize the distinction between note types, and appeared to favor features that seem to distinguish note types (Fig. 5). Therefore it may not be surprising, though it is interesting, to find that both hidden units are sensitive to an overlapping set of note features, though they weight these features differently, and that one hidden unit is also sensitive to a handful of features that the other unit is not sensitive to.

While these results are encouraging, it is important to compare them to those produced by alternative classification methods. For instance, can traditional statistical methods perform this classification task as well as the neural network did? If so, do these methods use the same sets of features, or do they use an alternative method to classify notes? In this section, we investigate these questions by using discriminant analysis to categorize notes using the identical data that were presented to the neural network that was described previously.

C. Discriminant analysis

When researchers are faced with data cases that can be assigned to a set of predefined classes, they often classify the data using discriminant analysis (e.g., Klecka, 1980), although Nowicki and Nelson (1990) did not study this method. Discriminant analysis is similar in many ways to multiple regression. Discriminant analysis determines a set of discriminant functions that can be used to categorize each instance. Each discriminant function has the form of a regression equation, $L=b_1x_1+b_2x_2+\cdots+b_nx_n+c$, where L is the value of the function, each b_i is a discriminant coeffi-

TABLE III. Classification of three types of input notes (first column) into three note classes by discriminant analysis.

Note type from visual examination	Note type according to discriminant analysis A	B	C	% correct
A	193	13	0	94
B	2	129	3	96
C	0	2	28	93
Total	195	144	31	95

cient, each x_i is a predictor variable, and c is a constant. If there are k different classes that cases can be assigned to, then discriminant analysis will compute a set of k different discriminant functions. The coefficients of these functions are selected in such a way that classification errors are minimized. A case is classified by providing its input features as the predictors for each of the discriminant functions. The case is assigned to the class whose discriminant function generates the maximum value.

We used discriminant analysis to classify the 370 notes into the three predefined classes of A, B, and C on the basis of the 9 input features that were described earlier. Table III provides the classification matrix that was produced from this analysis. In general, while discriminant analysis was slightly poorer at classifying the note types than was the artificial neural network, its performance was still excellent: it correctly classified 95% of the 370 notes.

As noted earlier, in order to make the classifications that are summarized in Table III, three different discriminant functions—one for each note type—had to be computed. Each of these functions is similar to the connection weights that were presented in Table II, insofar as each discriminant function is a set of coefficients that are used to weight an input pattern's features to make a classification decision. The constants and discriminant coefficients for the three discriminant functions are provided in Table IV.

As was the case for the two hidden units of the neural network, not all of the input features are equally important contributors to the discriminant functions. Table V provides the F-test of the contribution of each predictor variable to the set of discriminant functions. As can be seen from Table V, six of the nine note features were statistically significant predictors. Table V reveals a striking similarity between the discriminant analysis of the notes and the classifications made by the artificial neural network. All six of the statistically significant contributors in Table V also appear in at least one of the regression equations that were reported earlier; four of the six appear in both of the regression equations. This suggests that the classification power of both techniques may be rooted in processing similar sets of features. Only one feature (DD) was not a significant contributor in the table below, but was a significant contributor in the regression equation that predicted the activity of hidden unit 2.

The similarities between the discriminant analysis and the artificial neural network, in terms of their sensitivity to input features, suggest that a more detailed comparison of the two is in order. Table VI provides such a comparison. It presents the correlations between the two sets of connection weights in Table II (excluding the biases) and the three sets of coefficients in Table IV (excluding the constants).

These correlations are very informative. First, consider the correlations among the three sets of coefficients. The discriminant function for A notes has a strong negative correlation with the other two discriminant functions; however, the discriminant functions for B and C notes have a strong

TABLE IV. Discriminant coefficients for the three discriminant functions used to classify the 370 notes into classes A, B, and C on the basis of the 9 predictors listed in the leftmost column.

Predictor	Discriminant Function A	B	C
Constant	−3.027	−3.033	−14.696
TD	−0.846	0.547	3.350
AD	−0.205	0.144	0.760
DD	−0.092	0.241	−0.462
SF	5.284	−4.468	−16.291
PF	1.737	−0.870	−8.007
EF	−1.206	1.332	2.303
FM_{asc}	−0.965	1.376	0.463
FM_{desc}	0.370	−0.344	−0.984
F_{max}	−1.246	0.109	8.049

TABLE V. F statistics for evaluating the contributions of each note feature to the discriminant functions. Each F value was evaluated with degrees of freedom of 2367.

Feature	F To remove	p value
SF	89.59	0.0001
F_{max}	26.99	0.0001
TD	25.39	0.0001
PF	14.83	0.0001
EF	14.26	0.0001
FM_{asc}	8.82	0.0002
DD	1.66	0.193
FM_{desc}	1.17	0.312
AD	1.04	0.354

TABLE VI. Pearson product moment correlations between the two vectors of connection weights that feed into the neural network's hidden units and the three vectors of coefficients for the discriminant functions.

	H1 weights	H2 weights	A function	B function	C function
Weights for hidden unit 1	1.000				
Weights for hidden unit 2	0.600	1.000			
Discriminant function for A	0.958	0.649	1.000		
Discriminant function for B notes	−0.979	−0.663	−0.966	1.000	
Discriminant function for C notes	−0.860	−0.583	−0.958	0.851	1.000

positive correlation. This suggests that the discriminant functions are taking advantage of the same regularities in note properties that the artificial neural network also exploited. Recall from Fig. 5 that, on average, A notes have an opposite feature profile to both B and C notes, while B and C notes have similar profiles, but that C notes tend to have more extreme feature values. The relationships between notes in Fig. 5 are consistent with the correlations between discriminant function coefficients in Table VI.

Now consider the correlation between the connection weights and the discriminant function coefficients. Recall that in order to be classified as an A note, a note's features must produce higher activity in the A function than in the other two discriminant equations. In this sense, the discriminant function for A notes has exactly the same function as hidden unit 1 in the network, which only generates high activity to A notes. One can see in Table VI that this similarity in function is reflected in the strikingly high correlation between the weights for hidden unit 1 and the discriminant function for A notes. Convergent evidence for this is reflected in the extreme negative correlations between these weights and the other two sets of coefficients; this mirrors the relationship between the A function coefficients and the other two sets of coefficients. In short, the mechanisms feeding into hidden unit 1 are functionally equivalent to the discriminant function for classifying A notes.

In contrast, the relationship between the connection weights for hidden unit 2 and the discriminant functions is less clear. On the one hand, the correlations are all quite strong, and show the same general pattern as the weights for hidden unit 1: a positive relationship with the A function's coefficients, and a negative relationship with the other two sets of coefficients. This is to be expected, perhaps, because of the relationships between average note profiles in Fig. 5, which showed an opposite trend of features between A notes and the other two note types. This is also to be expected because there is a strong positive correlation between the two sets of connection weights, indicating that they should have similar relationships to the three discriminant functions.

On the other hand, because none of the correlations involving the weights of hidden unit 2 are as high as those for hidden unit 1, we are not in a position to declare any functional equivalence. However, this is completely consistent with our previous analysis of that unit (e.g., Fig. 4). Hidden unit 2 generates high activity to A notes or B notes, and does not respond to C notes. Therefore its behavior, unlike that of hidden unit 1, is much less likely to be directly related to any one of the discriminant functions.

IV. DISCUSSION

It was noted earlier that there are two important and related issues concerning note identification in the chick-a-dee call. The first is the development of an objective research tool for identifying note types in the lab. The second is the development of a theory about how the birds themselves process the acoustic signal to identify the different note types.

With respect to the first issue, the above-reported results are significant in several respects. They show that two very different techniques—artificial neural networks and linear discriminant analysis—are capable of classifying notes with a very high degree of accuracy. This is the case even when the large amount of information contained in a spectrogram is reduced to a small, manageable set of descriptive features. For our stimuli, the artificial neural network achieved an accuracy level of over 98%, and discriminant analysis achieved an accuracy level of 95%, for a relatively large set of stimuli (370 notes). The fact that both techniques performed so well would indicate that there is little reason to prefer one technique over another. Either method would perform extremely well as a note classification tool in a research lab.

With respect to the second issue, our analysis of the methods by which both techniques classified notes provides some important information to be considered in the development of a theory of how the birds themselves process acoustic signals to classify notes.

First, both the neural network and the discriminant analysis were able to classify notes by focusing on a particular subset of acoustic features. From a statistical perspective, it would appear that these features are particularly useful for note classification. The empirical issue that this observation raises is whether black-capped chickadees also focus on these features when they process this call.

Recent research from our laboratory suggests that in fact, black-capped chickadees do in fact attend to at least a subset of the acoustic features described here. Charrier *et al.* (2005) trained black-capped chickadees on four different note type discrimination tasks, A+B−, A−B+, B+C− and B−C+, and then conducted transfer tests with two sets of probe stimuli to determine to which features birds were attending in order to solve the task. Test notes were either pitch shifted, with the overall pitch of the note shifted up or down in frequency, or truncated notes, with either the ascending portion or descending portion presented. The results of these tests were quite clear. When notes were shifted in pitch toward the pitch space of another note type (e.g., when A notes had their overall note pitch lowered closer to that typical for B notes) the perception of that note also shifted from one type to the next. Further, birds could discriminate among note types using only the ascending portion of the note, and could generally do so better than using the descending portion of the note. This suggests that pitch cues contained in the notes such as SF, PF, EF and F_{max}, as well as particular "dynamic" temporal/spectral cues such as FM_{asc}, are salient to the birds and control the perception of note types in the chick-a-dee call. Interestingly, these same features also appear to contribute to the functioning of the ANN reported here, based on the interpretation of the activity of hidden units 1 and 2, and, suggesting that perhaps the ANN is tapping into similar features when it sorts notes as when the birds themselves do an analogous task.

A second important piece of information, related to the first, is the fact that both statistical techniques took advantage of the fact that when notes are described as normalized sets of summary features, B and C notes are more related to one another (and therefore harder to discriminate from one

another), while A notes seem quite different (Fig. 5). This was reflected in the patterns of correlations among the three sets of discriminant function coefficients, and in the fact that one of the functional components of the neural network was an A note detector. Again, this raises empirical questions about how birds process this call. Which note types do they appear to have more trouble discriminating? If they behave as if A notes are easier to differentiate than the other two note types, then this would suggest that the birds are processing a similar set of features as those represented in Fig. 5. However, if a different pattern of difficulties is in evidence, then this would suggest that the birds are processing different feature profiles.

Initial research on note type perception (Sturdy *et al.*, 2000) did in fact suggest that birds confused A, B, and C note types. When the pattern of errors was more closely examined, some interesting patterns emerged. Rewarded A note were confused with unrewarded B notes, rewarded B notes were confused with unrewarded A *and* unrewarded C notes, and rewarded C notes were confused with unrewarded B notes. These results, of an apparent perceptual continuum from A to B to C notes, mirrored the observation made earlier by Hailman *et al.* (1985) of a production continuum in these same notes types when the call is produced. Said another way, A notes, although sometimes confused, are only confused with one note type, B notes, whereas B notes are confused with both A and C note types. C notes are confused with B notes more than A notes, but Sturdy *et al.* also found that there was a trend for C notes to be confused with the next note type in the call sequence, D notes, not analyzed in the current paper. These findings provides some support, albeit indirectly, for the notion that both the birds and the ANN have difficulty in similar note type assignments, with A notes being misclassified less often than either B or C notes. This notion would have to be tested directly, perhaps by determining the number of notes of each type that a bird could discriminate, and to what level of accuracy, possibly in a choice task that required the discrimination and classification of all three note-types simultaneously, in a fashion similar to that of the ANN.

A third important point of information is that while the two techniques that we studied demonstrated similarities in features being processed, and also demonstrated some functional similarities, there were important differences between them too. The key similarity between the two techniques was that both included a distinct detector for A notes. By definition, the discriminant analysis produced distinct detectors for B and for C notes as well. Importantly, the artificial neural network did not develop these latter two types of detectors. Instead, its second hidden unit was an A or B note detector, which only turned off to C notes.

This difference between the two techniques demonstrates two different representational formats. In a local representation, each functional component of a system has a particular localized role, and can perform a meaningful task independently of the other functional components. The three discriminant functions created by the discriminant analysis are example of such local functional components. In contrast, artificial neural networks often develop distributed codes (Dawson, 2004; Dawson *et al.*, 2000a; Dawson *et al.*, 2000b; Dawson and Piercey, 2001; Hinton *et al.*, 1986; Leighton and Dawson, 2001). In a distributed code, different functional components of a system may not individually function in a fashion that can be directly related to the system's output. Instead, the meaningfulness of the functional components requires that their separate functions be considered simultaneously. The artificial neural network demonstrated this kind of representation. While the one hidden unit can be considered as being a local component (an A note detector), the behavior of the second hidden unit only makes sense when considered in the context of the behavior of hidden unit 1. That is, the representation is distributed because whether a note is an A, B, or C requires considering the activities of both hidden units at the same time (see the previous discussion of Fig. 4).

This is an important point for developing a theory of how birds process acoustic signals because it demonstrates that even when one focuses on the same set of features, there are different ways in which these features can be combined or processed to mediate note classification. The two techniques that we have described in this paper demonstrate two different approaches to using similar sets of features to make the same kind of classification. As theories of how chickadees process acoustic signals become related to the brain, one issue to address is whether acoustic features are represented in a local or distributed fashion. One advantage of artificial neural networks is that they provide examples of possible neural representations that can be used to guide this aspect of theory development (Dawson, 2004).

ACKNOWLEDGMENTS

This research was supported by an NSERC discovery grant, start-up funding from the University of Alberta and an Alberta Ingenuity New Faculty Grant to C.B.S. and NSERC and SSHRC research grants to M.R.W.D. I.C. was supported by an Isaac Walton Killam Memorial Trust postdoctoral fellowship and an Alberta Ingenuity Fund postdoctoral fellowship. I.C. is now a member of the Bioacoustics Team, NAMC-CNRS UMR 8620, Université Paris Sud, Bat.446 F-91405, Orsay, France. This research was approved by the University of Alberta Biological Sciences Animal Care Committee (protocol No. 351201, originally approved on 14 November 2001) and the University of Calgary Life and Environmental Sciences Animal Care Committee (protocol No. BI2001-028 originally approved on 29 November 2001). Birds were captured under an Environment Canada, Canadian Wildlife Service Scientific permit (permit No. WSA-1-02, originally approved on 25 January 2002) and Alberta Fish and Wildlife Research permits (Nos. 4619 GP, 4621 GP, and 8734 GP) and collection licenses (Nos. 088 CN, 089 CN, and 147 CN) that were originally approved on 21 December 2001. Correspondence can be addressed to either C.B.S. or M.R.W.D.

Bloomfield, L. L., Phillmore, L. S., Weisman, R. G., and Sturdy, C. B. (**2005**). "Note types and coding in Parid vocalizations. III. The chick-a-dee call of the Carolina chickadee (*Poecile carolinensis*)," Can. J. Zool. **83**, 820–833.

Bloomfield, L. L., Charrier, I., and Sturdy, C. B. (**2004**). "Note types and coding in Parid vocalizations. II. The chick-a-dee call of the Mountain chickadee (*Poecile gambeli*)," Can. J. Zool. **82**, 780–793.
Charrier, I., Bloomfield, L. L., and Sturdy, C. B. (**2004**). "Note types and coding in Parid vocalizations. I. The chick-a-dee call of the black-capped chickadee (*Poecile atricapilla*)," Can. J. Zool. **82**, 769–779.
Charrier, I., Lee, T. T-Y., Bloomfield, L. L., and Sturdy, C. B. (**2005**). "Acoustic mechanisms of note-type perception in black-capped chickadee calls," J. Comp. Psychol.. **199(4)**, 371–380.
Charrier, I., and Sturdy, C. B. (**2005**). "Call-based species recognition in black-capped chickadees," Behav. Processes. **70**, 271–281.
Clark, A. (**1989**). *Microcognition* (MIT, Cambridge, MA).
Clucas, B. A., Freeberg, T. M., and Lucas, J. R. (**2004**). "Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*)," Behav. Ecol. Sociobiol. **57**, 187–196.
Dawson, M. R. W. (**2004**). *Minds And Machines: Connectionism And Psychological Modeling* (Blackwell, Malden, MA).
Dawson, M. R. W., Boechler, P. M., and Valsangkar-Smyth, M. (**2000a**). "*Representing space in a PDP network: Coarse allocentric coding can mediate metric and nonmetric spatial judgements*," Spat. Cog. Comp.. **2**, 181–218.
Dawson, M. R. W., Medler, D. A., McCaughan, D. B., Willson, L., and Carbonaro, M. (**2000b**). "Using extra output learning to insert a symbolic theory into a connectionist network," Mind. Mach. **10**, 171–201.
Dawson, M. R. W., and Piercey, C. D. (**2001**). "On the subsymbolic nature of a PDP architecture that uses a nonmonotonic activation function," Mind. Mach. **11**, 197–218.
Doupe, A. J., and Kuhl, P. K. (**1999**). "Birdsong and human speech: Common themes and mechanisms," Annu. Rev. Neurosci. **22**, 567–631.
Ficken, M. S., Ficken, R. W., and Witkin, S. R. (**1978**). "Vocal repertoire of the black-capped chickadee," Auk **95**, 34–48.
Ficken, M. S., Hailman, E. D., and Hailman, J. P. (**1994**). "The chick-a-dee call system of the Mexican chickadee," Condor **96**, 70–82.
Freeberg, T. M., and Lucas, J. R. (**2002**). "Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*," Anim. Behav. **63**, 837–845.
Gaddis, P. K. (**1985**). "Structure and variability in the vocal repertoire of the mountain chickadee," Wilson Bull. **97**, 30–46.
Goldstein, M. H., King, A. P., and West, M. J. (**2003**). "Social interaction shapes babbling: Testing parallels between birdsong and speech," Proc. Natl. Acad. Sci. U.S.A. **100**, 8030–8035.
Hailman, J. P. (**1985**). "Ethology, zoosemiotic and sociobiology," Am. Zool. **25**, 695–705.
Hailman, J. P., and Ficken, M. S. (**1996**). "Comparative analysis of vocal repertoires, with reference to chickadees," in *Ecology and Evolution of Avian Vocal Communication*, edited by D. E. Kroodsma and H. H. Miller (Cornell University Press, Ithaca, NY), pp. 136–159.
Hailman, J. P., Ficken, M. S., and Ficken, R. W. (**1985**). "The 'chick-a-dee' calls of *Parus atricapillus*: A recombinant system of animal communication compared with written English," Semiotica **56**, 191–224.
Hinton, G. E., McClelland, J., and Rumelhart, D. (**1986**). "Distributed representations," in *Parallel Distributed Processing*, edited by D. Rumelhart and J. McClelland (MIT, Cambridge, MA), Vol. **1**, pp. 77–109.
Hughes, M., Nowicki, S., and Lohr, B. (**1998**). "Call learning in black-capped chickadees (Parus atricapillus): The role of experience in the development of 'chick-a-dee' calls," Ethology **104**, 232–249.
Klecka, W. R. (**1980**). *Discriminant Analysis* (Sage, Beverly Hills, CA).
Kruskal, J. B., and Wish, M. (**1978**). *Multidimensional Scaling* (Sage, Beverly Hills, CA).
Leighton, J. P., and Dawson, M. R. W. (**2001**). "A parallel distributed processing model of Wason's selection task," Cog. Sys. Res. **2**, 207–231.
Mammen, D. L., and Nowicki, S. (**1981**). "Individual-differences and within-flock convergence in chickadee calls," Behav. Ecol. Sociobiol. **9**, 179–186.
McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (**1986**). "The appeal of parallel distributed processing," in *Parallel Distributed Processing*, edited by D. Rumelhart and J. McClelland (MIT, Cambridge, MA), Vol. **1**, pp. 3–44.
Nowicki, S. (**1983**). "Flock-specific recognition of chickadee calls," Behav. Ecol. Sociobiol. **12**, 317–320.
Nowicki, S., and Nelson, D. A. (**1990**). "Defining natural categories in acoustic-signals—Comparison of 3 methods applied to chick-a-dee call notes," Ethology **86**, 89–101.
Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (**1986**). "Learning representations by back-propagating errors," Nature (London) **323**, 533–536.
Schneider, W. (**1987**). "Connectionism: Is it a paradigm shift for psychology?," Behav. Res. Methods Instrum. Comput. **19**, 73–83.
Slater, P. J. B. (**2003**). "Fifty years of bird song research: A case study in animal behaviour," Anim. Behav. **65**, 633–639.
Smith, S. T. (**1972**). *Communication and Other Social Behavior in Parus Carolinensis*, Publi. No. 11 (Nuttall Ornithological Club, Cambridge, MA).
Smolensky, P. (**1988**). "On the proper treatment of connectionism," Behav. Brain Sci. **11**, 1–74.
Sturdy, C. B., Phillmore, L. S., Price, J. L., and Weisman, R. G. (**1999a**). "Song-note discriminations in zebra finches (*Taeniopygia guttata*): Categories and pseudocategories," J. Comp. Psychol. **113**, 204–212.
Sturdy, C. B., Phillmore, L. S., and Weisman, R. G. (**1999b**). "Note types, harmonic structure, and note order in the songs of zebra finches (*Taeniopygia guttata*)," J. Comp. Psychol. **113**, 194–203.
Sturdy, C. B., Phillmore, L. S., and Weisman, R. G. (**2000**). "Call-note discriminations in black-capped chickadees (*Poecile atricapillus*)," J. Comp. Psychol. **114**, 357–364.
Vicario, D. S., Raksin, J. N., Naqvi, N. H., Thande, N., and Simpson, H. B. (**2002**). "The relationship between perception and production in songbird vocal imitation: What learned calls can teach us," J. Comp. Physiol., A **188**, 897–908.

Nonconstant quality of auditory filters in the porpoises, *Phocoena phocoena* and *Neophocaena phocaenoides* (Cetacea, Phocoenidae)

Vladimir V. Popov[a] and Alexander Ya. Supin[b]
Institute of Ecology and Evolution, The Russian Academy of Sciences, 33 Leninsky Prosp., 119071 Moscow, Russian Federation

Ding Wang[c] and Kexiong Wang[d]
Institute of Hydrobiology, The Chinese Academy of Sciences, Wuhan, Hubei 430072, People's Republic of China

(Received 10 October 2005; revised 24 January 2006; accepted 14 February 2006)

Simultaneous tone-tone masking in conjunction with the envelope-following response (EFR) recording was used to obtain tuning curves in porpoises *Phocoena phocoena* and *Neophocaena phocaenoides asiaeorientalis*. The EFR was evoked by amplitude-modulated probes with a modulation rate of 1000 Hz and carrier frequencies from 22.5 to 140 kHz. Equivalent rectangular quality Q_{ERB} of the obtained tuning curves varied from 8.3–8.6 at lower (22.5–32 kHz) probe frequencies to 44.8–47.4 at high (128–140 kHz) frequencies. The Q_{ERB} dependence on probe frequency could be approximated by regression lines with a slope of 0.83 to 0.86 in log-log scale, which corresponded to almost frequency-proportional quality and almost constant bandwidth of 3–4 kHz. Thus, the frequency representation in the porpoise auditory system is much closer to a constant-bandwidth rather that to a constant-quality manner. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2184290]

PACS number(s): 43.80.Lb [WWA] Pages: 3173–3180

I. INTRODUCTION

The auditory system of mammals functions as a bank of frequency-tuned bandbass filters which performs frequency analysis of sound signals. In the majority of mammals, passbands of these filters are roughly frequency proportional (except lowest frequencies); in other words, filter quality (the ratio of center frequency to the passband width) is almost constant throughout a major part of the frequency range of hearing. This feature is known as a constant-Q (quality) representation of frequencies in the auditory system. In humans, this regularity is known from the very first measurements of critical ratios (Fletcher, 1940) and critical bands, which are psychophysical equivalents of auditory filters (Zwicker, 1961). Later this regularity was confirmed in a variety of psychophysical investigations in humans with the use of various paradigms of frequency-tuning measurements: frequency-tuning curves [see Zwicker (1974) for a review], narrow-band noise masking [critical bands, see Zwicker (1982) for a review], notch-noise masking (Patterson, 1976; Patterson *et al.*, 1982), not to mention numerous other studies. To summarize the great number of psychophysical investigations, several analytical expressions have been proposed to describe filter bandwidth variation with central frequency (Greenwood, 1961; Zwicker and Terhardt, 1980; Moore and Glasberg, 1983; Glasberg and Moore, 1990). All of them are rather close to one another. For example, a simple equation given by Glasberg and Moore (1990) may be considered. Its original form suggested by the authors is

$$\mathrm{ERB} = 24.7(4.37F + 1), \tag{1}$$

where ERB is the filter equivalent rectangular bandwidth in Hz and F is the central frequency in kHz. With expression of both ERB and frequency in equal units, this equation may be presented as

$$\mathrm{ERB} = 0.108F + 24.7\ \mathrm{Hz}, \tag{2}$$

where F is frequency expressed in the same units as ERB. That is, at all frequencies above a few hundred Hz, where the constant of 24.7 Hz can be neglected, the filter bandwidth is of an almost constant proportion (10.8%) of frequency; respectively, the filter quality Q_{ERB} is constant and equal to $1/0.108 = 9.26$.

The very same regularity has been demonstrated by tuning curves obtained in humans and some mammals using evoked potentials such as cochlear action potentials (AP) (Dallos and Cheatham, 1976; Eggermont, 1977; Harris, 1978; Abbas and Gorga, 1981; Gorga and Abbas, 1981; Harrison *et al.*, 1981) and auditory brain-stem responses (ABRs) (Mitchell and Fowler, 1980; Salvi *et al.*, 1982; Gorga *et al.*, 1983; Brown and Abbas, 1987). The majority of these studies presented the quality of the obtained tuning curves as Q_{10} (the center frequency divided by the bandwidth at a level 10 dB above the tip of the curve) which is about twice lower than Q_{ERB}, but the principal result was the same: the quality was almost constant throughout a wide frequency range.

[a]Electronic mail: popov_vl@sevin.ru
[b]Electronic mail: alex_supin@sevin.ru
[c]Electronic mail: wangd@ihb.ac.cn
[d]Electronic mail: wangk@ihb.ac.cn

These and many other data led to a generally adopted idea that each relative frequency unit (e.g., octave) occupies a constant space on the basilar membrane (about 4 mm per octave in humans), as well as each critical band occuping a constant space (Greenwood, 1961). The result is a frequency-proportional filter (or critical band) width, which is the same as a constant filter quality.

The only well-known significant deviation from this rule is the auditory system of mustache bats. It has expanded representation of a narrow (a few kHz) frequency band around 60 kHz with much higher filter quality than at frequencies both above and below this "acoustic fovea" (e.g., Suga, 1978; Pollak and Casseday, 1989; Pollak, 1992). This exception of a common rule is obviously associated with properties of the biosonar of mustache bats, which employs narrow-band acoustic pulses of around 60-kHz frequency and requires very fine frequency analysis in just this frequency range to detect Doppler shift arising due to movement of the prey.

In the study described herein, we encountered another intriguing deviation from the rule of constant-Q frequency representation in the auditory system. We investigated hearing abilities of porpoises (Phocoenidae, Cetacea). The investigations of hearing of these small cetaceans were motivated both by concern for their protection and conservation and for their unique bioacoustics. Similar to many other odontocetes (toothed whales, dolphins, and porpoises), phocoenids have extremely wide frequency range (more than 150 kHz) and high sensitivity of hearing, obviously associated with their ability to echolocate (rev. Au, 1993; Supin *et al.*, 2001). However, contrary to many other odontocetes, phocoenids use for echolocation narrow-band pips instead of wide-band clicks (Dubrovskii *et al.*, 1971; Møhl and Andersen, 1973; Kamminga *et al.*, 1986, 1996; Kamminga, 1988; Li *et al.*, 2005). Therefore, some specific features of frequency tuning of hearing might be expected. Bearing this in mind, measurements of frequency tuning of porpoise's hearing were carried out. The measurements were done with use of the evoked-potential method in conjunction with the simultaneous tone-tone masking paradigm. The tone-tone masking (both the probe and masker are tones) provides tuning curves (masked thresholds as a function of frequency difference between the probe and masker) which reflects the auditory frequency-tuned filter form. The use of tuning-curve paradigm in conjunction with evoked-potential recording has already demonstrated its effectiveness for measurements of frequency tuning of hearing in a few species of cetaceans (Supin *et al.*, 1993; Popov *et al.*, 1995; Supin *et al.*, 2001). Therefore, the same method was used to investigate hearing of porpoises. These measurements provided unexpected results indicative of frequency representation more close to *constant-B* (bandwidth) rather than to *constant-Q* (quality). These results are presented herein.

II. METHODS

A. Subjects, facilities, and experimental design

Experimental animals were one harbor porpoise *Phocoena phocoena* (an adult female) and two Yangtze finless porpoises *Neophocaena phocaenoides asiaeorientialis*, a male and female.

Investigation of the harbor porpoise was carried out in facilities of the Utrish Marine Station of the Russian Academy of Sciences, Black Sea coast, Russia. The animal was kept in a pool ($9\times4\times1.2$ m^3) filled with sea water. During the experiment, the animal was taken from the home pool and placed on a stretcher in a plastic bath $4\times0.6\times0.6$ m^3 filled with sea water in such a manner that the dorsal surface of the head with the blowhole remained above the water surface. The walls and bottom of the bath and the water surface in front of the animal's head were covered by sound-absorbing material (rubber with cone-shaped closed air cavities) to reduce sound reflections and make the stimulus sound field more uniform.

Investigation of the finless porpoises was carried out in facilities of the Institute of Hydrobiology of the Chinese Academy of Sciences, Wuhan, P. R. China. The animals were kept in a pool ($20\times7\times3.0$ m^3) filled with fresh water. During the experiment, the animal was taken out of the home pool, placed on a stretcher in a wooden bath $2.25\times0.85\times0.6$ m^3 filled with water, and positioned in the same manner as the harbor porpoise. The walls and bottom of the bath were covered by open-cell neoprene to reduce sound reflection.

The care and use of all the animals were performed under the Guidelines of the Russian Ministry of Higher Education on the use of animals in biomedical research adhering to the Ethical Principles of the Acoustical Society of America.

B. Evoked-potential recording

The evoked-potential recording technique was similar for both the harbor porpoise and finless porpoises. For non-invasive evoked-potential recording, suction-cup electrodes were used consisting of a 15-mm stainless steel disk mounted within a 60-mm silicon suction cup. The active electrode was fixed at the vertex head surface, 5 cm behind the blowhole, above the water surface. The reference electrode was fixed at the dorsal or pectoral fin. The electrodes were connected by shielded cables to the input of a custom-made EEG amplifier that provided 88-dB gain within a frequency range of 200 to 5000 Hz, as defined at −3-dB points of 6-dB/oct slopes. The amplified signal was digitized and collected using an E-6040 data acquisition board (National Instruments) and stored in computer memory. To extract signal from noise, the digitized signal was coherently averaged (1000 original records per one averaged record) using triggering from the stimulus onset.

C. Sound signals

Sound signals were digitally synthesized at an update rate of 500 kHz and digital-to-analog converted by the same E-6040 board, amplified, attenuated, and played through a B&K 8104 transducer. The transducer was positioned at a distance of 60 cm in front of the animal's head, near the front wall of the bath. The playback channel was calibrated both before and after the experiments by positioning a calibrated

receiving hydrophone (B&K 8103) at the same location as the animal's head. Both probe and masker levels are specified below as their rms values.

Two kinds of sound signals were used to obtain frequency-tuning curves: probes and maskers. Probes were 20-ms-long sinusoidally amplitude-modulated tone bursts, modulation rate was 1000 Hz, and modulation depth was 100%. These parameters of modulation were chosen since they are optimal to produce in odontocetes (dolphins and porpoises) a robust rhythmic evoked-potential response (Supin *et al.*, 2001). Carrier frequencies of probes varied from 22.5 to 140 kHz, separated by $\frac{1}{4}$- or $\frac{1}{8}$-oct steps. Modulation always started from a zero phase in coherence with evoked-response collection. The probe bursts were presented at a rate of 10 s^{-1}.

Maskers were continuous nonmodulated tones. Their frequencies varied around the probe-carrier frequency with steps of 0.5 to 1.5 kHz. Intensities of probes and maskers were varied by independent attenuators with 5-dB steps; after attenuation, the probe and masker were mixed and emitted through the common transducer.

Even though the walls of the baths were covered with sound-attenuating material, some sound reflections from the bath walls and water surface were inevitable, thus resulting in interference patterns in the small enclosed space. To assess the influence of these interference patterns on the stimulus parameters, sounds were monitored by a B&K 8103 hydrophone near the animal's head. The monitoring showed that despite the sound reflections within the bath, the real modulation depth of the stimuli remained not less than 70%–80%, and local sound levels varied by not more than 5 dB.

D. Masked threshold evaluation and tuning-curve derivation

For tuning-curve derivation, probe levels 40 dB above absolute ABR thresholds were used. Thus, at thresholds from 50 to 80 dB *re*:1 μPa depending on frequency (Popov *et al.*, 1995, 2005), probe levels varied from 90 to 120 dB *re*:1 μPa. For masked threshold evaluation, masker level was varied while keeping the probe level constant. A 16-ms-long part of the rhythmic evoked-potential response to the sinusoidally modulated probe stimulus, from 6 to 22 ms, was Fourier transformed to obtain its frequency spectrum. The magnitude of 1-kHz peak was plotted as a function of masker intensity and an oblique near-threshold part of the plot not longer than 20 dB was approximated by a straight regression line [the criteria for selection of a range for straight-line approximation are motivated in more detail by Supin *et al.* (2001)]. The intersection of this line with the zero-amplitude level was adopted as a masked threshold estimate. This threshold-determination procedure was repeated at 6 to 12 masker frequencies around the probe frequency. The resulting function of masked threshold versus masker frequency was taken as the tuning curve.

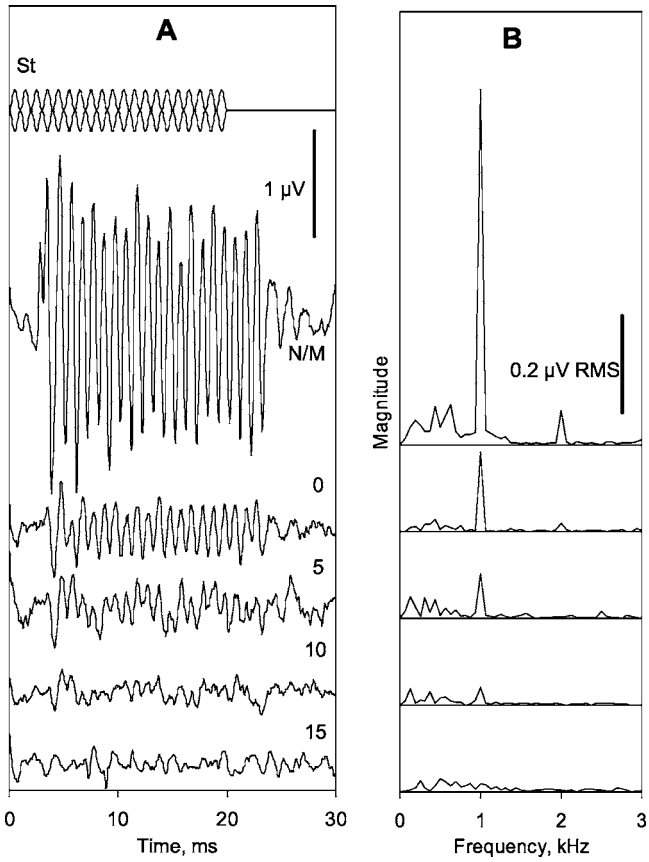


FIG. 1. Examples of EFR records (A) and their frequency spectra (B) at different masker-to-probe ratios. The finless porpoise, male. Probe 128 kHz, 90 dB *re*:1 μPa; masker 128 kHz, level (dB *re* probe) indicated near the records, N/M —no masker (control).

III. RESULTS

A. Evoked response waveform and manifestation of masking

Amplitude-modulated probe sounds evoked robust rhythmic responses, which followed the modulation rate, i.e., the envelope-following response (EFR). As Fig. 1(A) exemplifies, a probe of a moderate intensity (40 dB above the response threshold) without a masker evoked a response of more than 1 μV peak-to-peak amplitude (no-masker control). The burst onset evoked a small transient on-response, which after a few milliseconds transformed to EFR. Both the start and the end of the response appeared with a few ms lag relative to the stimulus. This lag provided a good opportunity to check artifact contamination of records. The response-free initial part of the records showed clearly that the records were not contaminated with electromagnetic artifacts; as well, the response persistence until about 4.5 ms after the stimulus end showed the physiological nature of the response.

The same probe in masker background evoked smaller responses. As the masker level increased (from 0 to 15 dB above the probe level in Fig. 1), EFR diminished until it disappeared in the noise.

Frequency spectra of the records are shown in Fig. 1(B). These spectra were obtained by Fourier transform of a part of the record, from 6 to 22 ms. This 16-ms window contained a

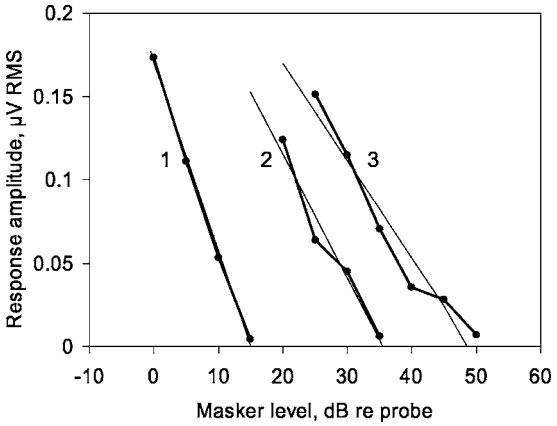


FIG. 2. EFR magnitude dependence on masker-to-probe ratio at different masker-to-probe frequency spacings. The finless porpoise, male. Probe 128 kHz, 90 dB *re*:1 μPa. Masker frequencies: 128 kHz (1), 131.5 kHz (2), and 135 kHz (3). Masker level (abscissa) in dB *re* probe level.

major part of the EFR record but did not contain the latency and the initial transient part of the response. At masker levels below 15 dB, all the spectra featured a peak at the probe envelope frequency of 1 kHz. At the masker level of 10 dB, this peak was comparable with the spectrum noise level but still definite, and at the masker level of 15 dB it completely disappeared.

Figure 2 demonstrates the masked threshold determination procedure. The magnitude of the 1-kHz peak of the response spectrum was taken as an estimate of the response magnitude and plotted against masker level. The near-threshold part of the plot could be satisfactorily approximated by a straight regression line crossing the zero-amplitude level in a point accepted as the threshold estimate.

As it usually is in the tuning-curve paradigm, masked thresholds depended on frequency difference between the probe and masker. The lowest masked threshold (i.e., the most effective masking) appeared at a coincidence of the probe and masker frequencies. With the difference increase, the masked threshold increased. Figure 2 exemplifies response amplitude versus masker intensity functions at three probe-to-masker frequency differences: (1) zero, (2) 3.5 kHz, and (3) 7 kHz. The example demonstrates increasing the masked threshold from 15.0 dB (relative to the probe level) at the zero difference to 35.5 dB at a 3.5-kHz difference and further to 48.6 dB at a 7-kHz difference.

Using the same procedure, masked thresholds were determined at a variety of masker frequencies around the probe frequency to obtain a tuning curve as a masked threshold versus masker frequency function (at a certain probe frequency). The same measurements were repeated at probe frequencies from 22.5 kHz (in the harbor porpoise) or 32 kHz (in the finless porpoises) to 140 kHz, thus obtaining a family of tuning curves. Lower probe frequencies were not tested since evoked-response amplitude was too low at lower frequencies [as described by Popov and Supin (2001)] and did not provide satisfactory precision of measurements. Higher probe frequencies were not tested because of proximity to the upper limit of the frequency range of hearing.

The family of tuning curves obtained from the harbor

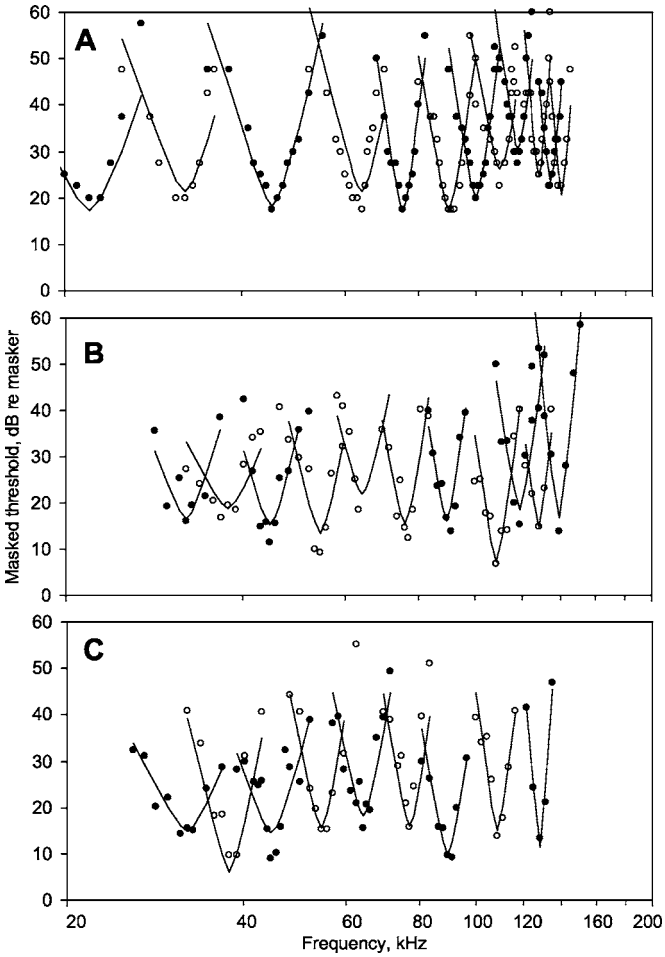


FIG. 3. Tuning curve families obtained in three subjects: the harbor porpoise (A), finless porpoise female (B), and finless porpoise male (C). The curves are masked thresholds as functions of masker frequency. Dots represent experimental data, curves—roex approximations.

porpoise is presented in Fig. 3(A), and similar families obtained in the finless porpoises are presented in Figs. 3(B) and 3(C). All the families looked rather similar. In particular, a noticeable feature of all the tuning curves was that they are not of a constant width when plotted on a log frequency scale, thus indicating nonconstant Q across the frequency scale: high-frequency tuning curves were much more acute than low-frequency ones.

In order to characterize the curves quantitatively, the experimental points were approximated by the rounded-exponential (*roex*) function, which is widely used to approximate auditory filters. Its original form suggested by Patterson *et al.* (1982) is

$$W(g) = (1 + pg)e^{-pg}, \tag{3}$$

where $W(g)$ is the filter form in terms of transferred power, g is the normalized frequency deviation from the center, and the term p determines the filter tuning. ERB of the *roex* function is

$$B = F\int_{-\infty}^{\infty}(1+pg)e^{-pg}\,dg = 4F/p, \tag{4}$$

where B is ERB and F is the center frequency. Respectively, the quality Q is

$$Q = F/B = p/4. \tag{5}$$

Curves like those presented in Fig. 3 may be approximated by an inverted *roex* function with $W(g)$ presented on a dB scale:

$$T(g) = T_{\min} - 10\log[W(g)], \tag{6}$$

where $T(g)$ is the masked threshold and $T_{\min}$ is the minimal masked threshold at the curve tip. To approximate experimental data by *roex* functions, the parameters $T_{\min}$ and p were adjusted to reach the best fit to experimental data according to the least-mean-square criterion. These approximating functions are presented in Fig. 3 along with the experimental points. The obtained value of p was used to calculate ERB and Q of the curves according to the equations presented above. The result was obvious dependence of Q on the probe frequency: from 8.3–8.6 at rather low frequencies (22.5–32 kHz) to 44.8–47.4 at high frequencies (128–140 kHz).

All the obtained values of Q are presented in Fig. 4 as functions of probe frequency. All the functions look rather similar: neglecting some data scatter, there was an almost proportional increase of Q with frequency. Being presented on log-log scales, these functions were satisfactorily approximated by straight lines (r^2 from 0.85 to 0.91, Table I).The slope of these regression lines in the three subjects varied from 0.83 to 0.86, as presented in Table I. In other words, the dependence of Q on frequency might be approximated by a function

$$Q = q_0 F^k, \tag{7}$$

where F is frequency, q_0 is a quality at $F=1$, and k determines the degree of Q dependence on F. Constant (frequency-independent) Q corresponds to $k=0$; frequency-proportional Q corresponds to $k=1$. The found values of k from 0.83 to 0.86 indicated not exact but almost frequency-proportional Q. With such degree of dependence, approximated Q increased from around 10 to around 40 within a frequency range from 30 to 150 kHz.

According to Eqs. (5) and (7), ERB dependence on frequency was approximated as

$$B = F/Q = F^{1-k}/q_0. \tag{8}$$

ERB approximations computed in such a way were little dependent on frequency, e.g., from around 3 kHz at 30-kHz frequency to around 4 kHz at 150-kHz frequency.

IV. DISCUSSION

The tuning curves presented above were obtained in conditions of nonideal acoustic field. Even though the walls of the experimental baths were covered with sound-attenuating material, sound interference resulted in some spatial variation of both the sound level and modulation depth. Nevertheless, we suppose that the obtained tuning curves are credible because of the reasons as follows.

(i) As shown (Supin and Popov, 1995), variation of modulation depth influences the envelope following response amplitude rather than threshold. So, it could not influence the form of tuning curves based on masked thresholds.

(ii) Variation of sound (both the probe and masker) levels by ±5 dB does directly influence the masked threshold estimates. However, this variation should increase the data scatter, but could hardly systematically influence the width of tuning curves.

At least two features of the porpoise's tuning curves described above deserve attention: (*i*) very high quality (up to 40) in the upper part of the hearing frequency range and

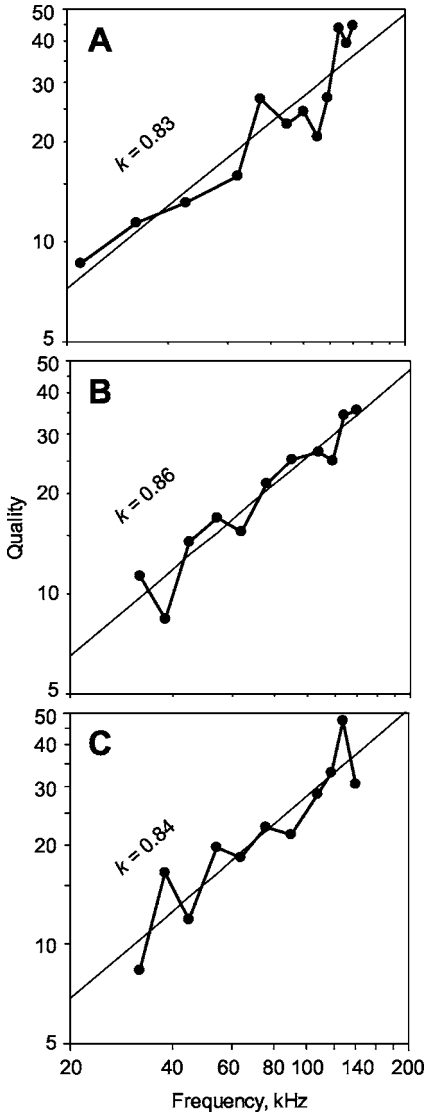


FIG. 4. Frequency-tuning quality (Q) dependence on frequency in three subjects: the harbor porpoise (A), finless porpoise female (B), and finless porpoise male (C). Solid lines—experimental data, thin straight lines—approximating regression lines. k is indicated for each regression line.

TABLE I. Q dependence on frequency in different mammal species.

	r^2	$k \pm$SE	Range (kHz)	References
		Present data		
Harbor porpoise	0.87	0.83±0.10	22–140	
Finless porpoise female	0.91	0.86±0.09	32–139	
Finless porpoise male	0.85	0.84±0.12	32–139	
Mean	0.88	0.84±0.06		
		Summarizing analytical expression		
Human		0.06±0.01	1–16	Glasberg and Moore (1990)
		Psychophysical critical bands (F/B)		
Cat		−0.13±0.05	1–8	Pickles (1975)
Cat		0.12±0.09	4–12	Nienhuys and Clark (1979)
Mouse		−0.03±0.05	10–50	Ehret (1976)
Chinchilla		0.21±0.07	1–4	Seaton and Trahiotis (1975)
Macaque		−0.13±0.07	1–16	Gourevitch (1970)
		Psychophysical tuning curves ($Q_{10\ \mathrm{dB}}$)		
Chinchilla		0.32±0.26	1–12	McGee *et al.* (1976)
Macaque		0.00±0.02	1–16	Serafin *et al.* (1982)
		Evoked-potential tuning curves ($Q_{10\ \mathrm{dB}}$)		
Guinea pig		0.40±0.06	2–8	Mitchell and Fowler (1980)
Bottlenose dolphin		0.18±0.07	11–110	Popov *et al.* (1995)
Beluga whale		0.54±0.12	32–108	Klishin *et al.* (2000)

(*ii*) tuning dependence on frequency which is much closer to a constant-bandwidth rather than to a constant-quality manner.

The frequency tuning of $Q_{\mathrm{ERB}} \approx 40$ is several times more acute than that in many mammals and humans ($Q_{\mathrm{ERB}} \approx 10$, see Sec. I). However, this feature of the porpoise's tuning curves is not very surprising, since auditory filters with quality higher than 30 were already described in a few dolphin species (Supin *et al.*, 1993; Popov *et al.*, 1995, 1997; Klishin *et al.*, 2000).

The cause of very acute frequency tuning in dolphins and porpoises is not known yet, but we may hypothesize that it is associated with the wide frequency range of their hearing. Indeed, both frequency tuning and temporal resolution in the auditory system are limited by passbands of peripheral filters. Acute frequency tuning requires narrow passbands whereas high temporal resolution requires wide passbands. Thus, increasing frequency-tuning acuity above a certain limit is only possible at the cost of decreased temporal resolution and vice versa. However, this contradiction becomes less limiting at high frequencies, because at high frequency F, auditory filters can combine high quality with wide passband. Thus, at high frequencies, acute frequency tuning does not prevent high temporal resolution. Odontocetes (toothed whales, dolphins, and porpoises) have frequency ranges of hearing of more than 100 kHz, in particular, up to 150 kHz in porpoises, which is several times wider than in the majority of terrestrial mammals (except bats). This frequency range allows combining acute frequency tuning with high temporal resolution.

The dependence of tuning on frequency found in porpoises was unexpected. Contrary to the majority of terrestrial mammals and humans, this dependence is much closer to constant-B rather than to constant-Q representation.

In principle, the constant-B frequency representation is not prohibited by any auditory mechanisms. In particular, it is typical of the lowest part of the frequency range of the human's hearing: as follows from Eqs. (1) and (2), at low frequencies, ERB cannot be infinitely narrow and therefore approximates a constant value of around 25 Hz. But it is only typical of the lowest part of the frequency range. Within the upper part, a constant-Q (or approaching constant-Q) representation is more typical of both humans and a variety of mammals. We analyzed, retrospectively, data obtained in a few mammal species which are commonly used in hearing investigations, in order to calculate Q dependence on frequency within an upper part of the frequency range. For critical-band data (Gourevitch, 1970; Pickles, 1975; Seaton and Trahiotis, 1975; Ehret, 1976; Nienhuys and Clark, 1979), Q was calculated as the frequency-to-band ratio; for tuning-curve data (McGee *et al.*, 1976; Mitchell and Fowler, 1980; Serafin *et al.*, 1982; Popov *et al.*, 1995; Klishin *et al.*, 2000), $Q_{10\ \mathrm{dB}}$ estimates were directly taken. The Q dependence on frequency was approximated by a log-log regression line, exactly in the same manner as we did for the experimental data presented above. The values obtained for the factor k [see Eq. (5)] are presented in Table I along with our experimental data. Many of them were not far different from zero, thus indicating the constant-Q representation; in some cases, k reached values as high as 0.4–0.5, thus indicating a representation type intermediate between constant-Q and constant-B. But only in porpoises did values as high as 0.85 indicate a representation type much closer to constant-B rather than to constant-Q.

It is remarkable that frequency-tuning measurements in a few dolphin species carried out with the use of the very same technique (tuning curves obtained with sinusoidally amplitude-modulated tones as a probe, continuous tones for simultaneous masking, and EFR as an indicative response) resulted in a representation type much closer to constant-Q (Popov *et al.*, 1995; Klishin *et al.*, 2000). So, the unusual manner of frequency-tuning dependence on frequency in porpoises cannot be explained by some peculiarities of the methods used for hearing investigation in cetaceans. It may be considered as a real feature of the porpoise's auditory system.

It should be noted that the results presented above do not cover the entire frequency range of hearing but only the upper part of this region. We failed to perform measurements precise enough at lower frequencies, so the type of frequency representation in the low-frequency range in porpoises remains uncertain. But the investigated high-frequency region is just one where constant-Q representation appears normally in other mammals. So the representation manner in porpoises really differs from that in other mammals and humans.

Again, we know nothing about the cause of such type of frequency representation. One of the advantages of such an organization is that a constant bandwidth of peripheral auditory filters provides a constant limit for temporal resolution throughout all the frequency range. A filter of a bandwidth B is capable of transferring modulation rates up to $B/2$; thus, the porpoise's auditory filters of a bandwidth of 3–4 kHz are capable of transferring amplitude modulations of a signal as fast as 1.5–2 kHz. This value is very close to estimates of temporal resolution of the dolphin's and porpoise's hearing obtained by both psychophysical and evoked-potential methods [see reviews Au (1993) and Supin *et al.* (2001)].

ACKNOWLEDGMENTS

Grants provided by National Foundation of Natural Sciences of China (30170142), the Chinese Academy of Sciences (KSCX2-SW-118), the Institute of Hydrobiology, the Chinese Academy of Sciences (220103), Russian Foundation for Basic Research (02-04-39017, 03-04-48117), and Russian Ministry of Science and Education (NSh-2152.2003.4) are greatly appreciated.

Abbas, P. J., and Gorga, M. P. (**1981**). "AP responses in forward-masking paradigms and their relationship to responses of auditory-nerve fibers," J. Acoust. Soc. Am. **69**, 492–499.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Brown, C. J., and Abbas, P. J. (**1987**). "A comparison of AP and ABR tuning curves in guinea pig," Hear. Res. **25**, 193–204.

Dallos, P., and Cheatham, M. A. (**1976**). "Compound action potential (AP) tuning curves," J. Acoust. Soc. Am. **59**, 591–597.

Dubrovskii, N. A., Krasnov, P. S., and Titov, A. A. (**1971**). "On the emission of echolocation signals by the Azov sea harbor porpoise," Sov. Phys. Acoust. **16**, 444–448.

Eggermont, J. J. (**1977**). "Compound action potential tuning curves in normal and pathological human ears," J. Acoust. Soc. Am. **62**, 1247–1251.

Ehret, G. (**1976**). "Critical bands and filter characteristics of the ear of the housemouse (*Mus musculus*)," Biol. Cybern. **24**, 35–42.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gorga, M. P., and Abbas, P. J. (**1981**). "Forward-masked AP tuning curves in normal and acoustically traumatized ears," J. Acoust. Soc. Am. **70**, 1322–1330.

Gorga, M. P., McGee, J., Walsh, E. J., Javel, E., and Farley, G. R. (**1983**). "ABR measurement in the cat using a forward-masking paradigm," J. Acoust. Soc. Am. **73**, 256–261.

Gourevitch, G. (**1970**). "*Detectability of tones in quiet and in noise by rats and monkeys*," in Animal Psychophysics: The Design and Conduct of Sensory Experiments, edited by W. C. Stebbins (Appleton, New York), pp. 67–97.

Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

Harris, D. M. (**1978**). "Action potential suppression, tuning curves and thresholds: Comparison with single fiber data," Hear. Res. **1**, 133–154.

Harrison, R. V., Aran, J.-M., and Erre, J.-P. (**1981**). "AP tuning curves from normal and pathological human and guinea pig cochleas," J. Acoust. Soc. Am. **69**, 1374–1385.

Kamminga, C. (**1988**). "Echolocation signal types of odontocetes," in *Animal Sonar, Processes and Performance*, edited by P. E. Nachtigall, and P. W. B. Moore (Plenum, New York), pp. 9–22.

Kamminga, C., Cohen, S. A. B., and Silber, G. K. (**1996**). "Investigation of cetacean sonar XI: Intrinsic comparison of the wave shapes of some members of the Phocoenidae family," Aquat. Mamm. **22**, 45–55.

Kamminga, C., Kataoka, T., and Engelsma, F. J. (**1986**). "Investigation of cetacean sonar VII: Underwater sounds of Neophocaena phocaenoides of the Japanese coastal population," Aquat. Mamm. **12**, 52–60.

Klishin, V. O., Popov, V. V., and Supin, A. Ya. (**2000**). "Hearing capabilities of a beluga whale, *Delphinapterus leucas*," Aquat. Mamm. **26**(3), 212–228.

Li, S., Wang, K., Wang, D., and Akamatsu, T. (**2005**). "Echolocation signals of the free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*)," J. Acoust. Soc. Am. **117**, 3288–3296.

McGee, T., Ryan, A., and Dallos, P. (**1976**). "Psychophysical tuning curves of chinchillas," J. Acoust. Soc. Am. **60**, 1146–1150.

Mitchell, C., and Fowler, C. (**1980**). "Tuning curves of cochlear and brainstem responses in the guinea pig," J. Acoust. Soc. Am. **68**, 896–900.

Møhl, B., and Andersen, S. (**1973**). "Echolocation: High frequency component in the click of the harbour porpoise (*Phocoena phocoena* L.)," J. Acoust. Soc. Am. **54**, 1368–1373.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Nienhuys, Y. W., and Clark, G. M. (**1979**). "Critical bands following the selective destruction of cochlear inner and outer hair cells," Acta Oto-Laryngol. **88**, 350–358.

Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milory, R. (**1982**). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.

Pickles, J. O. (**1975**). "Normal critical bands in the cat," Acta Oto-Laryngol. **80**, 245–254.

Pollak, G. D. (**1992**). "Adaptation of basic structures and mechanisms in the cochlea and central auditory pathway of the mustache bat," in *The Evolutionary Biology of Hearing*, edited by D. B. Webster, R. R. Fay, and A. N. Popper (Springer, New York), pp. 751–778.

Pollak, G. D., and Casseday, J. H. (**1989**). *The Neural Basis of Echolocation in Bats* (Springer, Heidelberg).

Popov, V. V., and Supin, A. Ya. (**2001**). "Contribution of various frequency bands to ABR in dolphins," Hear. Res. **151**, 250–260.

Popov, V. V., Supin, A. Ya., and Klishin, V. O. (**1995**). "Frequency tuning curves of the dolphin's hearing: Envelope-following response study," J. Comp. Physiol., A **178**, 571–578.

Popov, V. V., Supin, A. Ya., and Klishin, V. O. (**1997**). "Frequency tuning of the dolphin's hearing as revealed by auditory brain-stem response with notch-noise masking," J. Acoust. Soc. Am. **102**, 3759–3801.

Popov, V. V., Supin, A. Ya., Wang, D., Wang, K., Xiao, J., and Li, S. (**2005**). "Evoked-potential audiogram of the Yangtze finless porpoise *Neophocaena phocaenoides asiaeorientalis* (L)," J. Acoust. Soc. Am. **117**, 2728–2731.

Salvi, R. J., Ahroon, W. A., Perry, J. W., Gunnarson, A. D., and Henderson, D. (**1982**). "Comparison of psychophysical and evoked potential tuning curves in the chinchilla," Am. J. Otolaryngol. **3**, 408–416.

Seaton, W. H., and Trahiotis, C. (**1975**). "Comparison of critical ratios and

critical bands in the monaural chinchilla," J. Acoust. Soc. Am. **57**, 193–199.
Serafin, S. V., Moody, D. B., and Stebbins, W. C. (**1982**). "Frequency selectivity of the monkey's auditory system: Psychophysical tuning curves," J. Acoust. Soc. Am. **71**, 1513–1518.
Suga, N. (**1978**). "Specialization of the auditory system for reception and processing of species-specific sounds," Fed. Proc. **37**, 2342–2354.
Supin, A. Ya., and Popov, V. V. (**1995**). "Envelope-following response and modulation transfer function in the dolphin's auditory system," Hear. Res. **92**, 38–46.
Supin, A. Ya., Popov, V. V., and Klishin, V. O. (**1993**). "ABR frequency tuning curves in dolphins," J. Comp. Physiol., A **173**, 649–656.
Supin, A. Y., Popov, V. V., and Mass, A. M. (**2001**). *The Sensory Physiology of Aquatic Mammals* (Kluwer, New York).
Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.
Zwicker, E. (**1974**). "On a psychoacoustical equivalent of tuning curves," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, Berlin), pp. 132–141.
Zwicker, E. (**1982**). *Psychoacustic* (Springer, Berlin).
Zwicker, E., and Terhardt, E. (**1980**). "Analytical expression for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.

Comparison of in-air evoked potential and underwater behavioral hearing thresholds in four bottlenose dolphins (*Tursiops truncatus*)

James J. Finneran
U.S. Navy Marine Mammal Program, Space and Naval Warfare Systems Center, San Diego, Code 2351, 53560 Hull Street, San Diego, California 92152

Dorian S. Houser
Biomimetica, 7951 Shantung Drive, Santee, California 92071

(Received 14 September 2005; revised 23 January 2006; accepted 3 February 2006)

Traditional behavioral techniques for hearing assessment in marine mammals are limited by the time and access required to train subjects. Electrophysiological methods, where passive electrodes are used to measure auditory evoked potentials (AEPs), are attractive alternatives to behavioral techniques; however, there have been few attempts to compare AEP and behavioral results for the same subject. In this study, behavioral and AEP hearing thresholds were compared in four bottlenose dolphins. AEP thresholds were measured in-air using a piezoelectric sound projector embedded in a suction cup to deliver amplitude modulated tones to the dolphin through the lower jaw. Evoked potentials were recorded noninvasively using surface electrodes. Adaptive procedures allowed AEP hearing thresholds to be estimated from 10 to 150 kHz in a single ear in about 45 min. Behavioral thresholds were measured in a quiet pool and in San Diego Bay. AEP and behavioral threshold estimates agreed closely as to the upper cutoff frequency beyond which thresholds increased sharply. AEP thresholds were strongly correlated with pool behavioral thresholds across the range of hearing; differences between AEP and pool behavioral thresholds increased with threshold magnitude and ranged from 0 to +18 dB.
[DOI: 10.1121/1.2180208]

PACS number(s): 43.80.Lb, 43.80.Ev [WWA] Pages: 3181–3192

I. INTRODUCTION

Marine mammal hearing capabilities have traditionally been assessed using behavioral response paradigms (e.g., Johnson, 1967; Schusterman and Moore, 1978; Moore and Schusterman, 1987; Awbrey *et al.*, 1988; Kastak and Schusterman, 1998; Finneran *et al.*, 2002; Kastelein *et al.*, 2002). In these methods, operant conditioning techniques are used to train animals to perform specific actions when they hear sounds. Features of the sound (e.g., amplitude, frequency) are varied and the subject's responses recorded and used to estimate thresholds or other desired features of the auditory system. These techniques provide the most direct assessment of the performance of the auditory system. For some smaller cetaceans, behavioral techniques have been used to define auditory system features such as temporal integration (Johnson, 1968), auditory filter widths (Au and Moore, 1990; Finneran *et al.*, 2002), auditory fatigue (Schlundt *et al.*, 2000; Nachtigall *et al.*, 2004), and receiving directivity (Au and Moore, 1984; Kastelein *et al.*, 2005).

Behavioral techniques require considerable access to subjects for training before reliable data can be obtained, thus preventing these methods from being used on larger marine mammals not routinely kept under human care. Behavioral approaches are also time consuming, both for initial training and data collection. The limited availability of subjects and the required time and costs have limited the number of individuals and species of marine mammals tested with behavioral techniques. At present, behavioral hearing thresholds have been obtained in only about 21 of the more than 100 extant marine mammal species. The National Research Council (National Research Council (NRC), 2000) stated that "population-level audiograms of many individuals are necessary for establishing baseline hearing capabilities and normal hearing loss in marine mammals" and has repeatedly recommended that studies be conducted to obtain such data (NRC, 1994; 2000). The NRC Ocean Studies Board further singled out the need to individually train behavioral hearing test subjects as the "major barrier to large-scale testing of the hearing of many individuals of the same species" (NRC, 2000). The need to specifically train individuals for hearing tests has prevented widespread testing, even at facilities with large numbers of trained marine mammals.

Electrophysiological hearing tests use passive electrodes to detect small voltages generated by the brain in response to acoustic stimuli. The voltages, called auditory evoked potentials (AEPs), are routinely measured in infants, children, and patients who are unwilling or unable to respond behaviorally. AEP methods are attractive alternatives to behavioral methods because they are fast and require minimal cooperation from the subject. AEPs have been measured in several marine mammal species, including bottlenose dolphins (Ridgway *et al.*, 1981; Popov and Supin, 1990b), harbor porpoises (*Phocoena phocoena*; Popov *et al.*, 1986), belugas (*Delphinapterus leucas*; Popov and Supin, 1987), killer whales (*Orca orcinus*; Szymanski *et al.*, 1999), tucuxi dolphins (*So-*

talia fluviatilis; Popov and Supin, 1990a), Amazon river dolphins (*Inia geoffrensis*; Popov and Supin, 1990c), the false killer whale (*Pseudorca crassidens*; Supin *et al.*, 2003; Yuen *et al.*, 2005), a striped dolphin (*Stenella coeruleoalba*; Andre *et al.*, 2003), Risso's dolphin (Grampus griseus; Dolphin, 1997) the finless porpoise (*Neophocaena phocaenoides asiaeorientalis*; Popov *et al.*, 2005), sea lions (*Zalophus californianus*; Bullock *et al.*, 1971) and harbor seals (*Phoca vitulina*; Wolski *et al.*, 2003).

The broader application of AEPs to the study of marine mammal hearing would permit more rapid collection of information on marine mammal auditory systems and address issues acknowledged by the NRC as limiting our understanding of marine mammal hearing at the population level. Unfortunately, most AEP measurements have occurred in laboratory settings using a single individual and there are few direct comparisons between AEP and behavioral hearing thresholds in the same individual.

This paper describes an AEP system developed to address a need for periodic hearing assessment of dolphins at the U.S. Navy Marine Mammal Program at the Space and Naval Warfare Systems Center, San Diego (SSC San Diego). Navy dolphins are trained to find and mark the location of underwater objects and waterborne intruders. To accomplish these tasks, they rely upon their biological sonar (echolocation) system. Since echolocation requires both generation of outgoing pulses and reception (hearing) of returning echoes, loss of hearing sensitivity would eventually decrease a dolphin's ability to detect and identify targets. Periodic hearing assessments of Navy dolphins are therefore needed to identify those animals with degraded hearing. This paper presents the operating principles and performance characteristics of the AEP system. AEP hearing thresholds obtained in air are compared to underwater behavioral thresholds in the same subjects to determine their suitability as proxies for underwater behavioral data. System effectiveness and the time required for hearing estimates in delphinids are discussed relative to the goals of routine screening of Navy dolphins and expanding our knowledge of population level audiometry within marine mammal species.

II. METHODS

A. Subjects

The subjects were four bottlenose dolphins: BLU (female, 39-y, 200 kg), BUG (male, 23-y, 190 kg), BUS (male, 24-y, 180 kg), and WEN (male, 21-y, 210 kg). All subjects had previously participated in cooperative psychophysical tasks, including auditory detection tasks. Subjects were housed in floating netted enclosures, 9×9 to 12×24 m, located in San Diego Bay (SD Bay), California. The study followed a protocol approved by the Institutional Animal Care and Use Committee of the Biosciences Division, SSC San Diego, and followed all applicable U.S. Department of Defense guidelines for the care of laboratory animals.

B. AEP measurements

AEP measurements were conducted in air. In-air measurements were preferable to underwater measurements because AEP amplitudes are generally larger in air compared to those collected underwater (seawater provides an alternative low resistance path for the electrical potentials), and the AEP measurements could be coupled to clinical examinations of dolphins, which are performed with the dolphin removed from water.

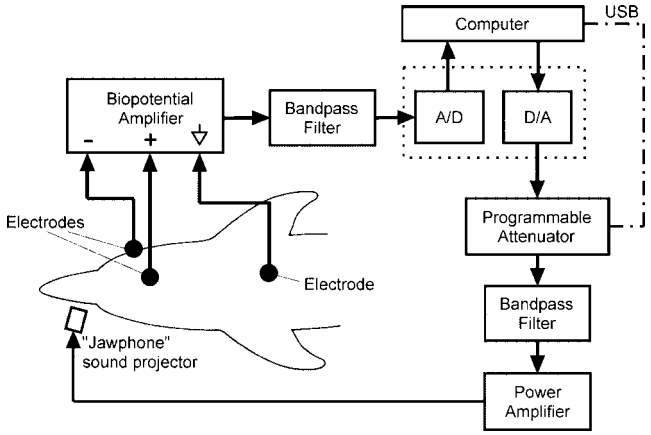


FIG. 1. AEP measurement system configuration.

1. Hardware

Figure 1 is a block diagram of the AEP measurement system. A personal computer (PC) with a multifunction data acquisition card (National Instruments PCI-MIO-16E-1) was used to generate sound stimuli and digitize the evoked responses. Sound stimuli were attenuated (Tucker-Davis Technologies PA-5), bandpass filtered (Krohn-Hite 3C module, 1–150 kHz), and amplified (Hafler P1000) before being presented to the subject. The attenuator featured a USB interface that allowed the PC to programmatically change the attenuation.

Sounds were presented to the subjects using piezoelectric sound projectors embedded in silicon rubber (Rhodia V-1065) suction cups. The suction cup-embedded projectors have been labeled "jawphones" in prior literature (Moore *et al.*, 1995; Brill *et al.*, 2001). The jawphone was placed on the subject's lower jaw, in the pan region, a site which has been previously demonstrated to be an important pathway for high frequency sound reception in dolphins (Brill *et al.*, 1988; Møhl *et al.*, 1999a). Jawphones were calibrated by measuring underwater rms sound pressure levels (SPLs) at a distance of 15 cm from the jawphone [Fig. 2(a)]. The 15-cm distance between jawphone and calibration hydrophone was based on the estimated distance between the jawphone attachment point and the auditory bullae. This distance was measured from a computed tomography (CT) scan of WEN (Houser *et al.*, 2004). SPLs were measured using a calibrated hydrophone (B&K 8105) and charge amplifier (B&K 2692 or 2635). Three individual jawphones, designated JP1, JP2, and JP4, were used. Table I details the size and operating range of the various jawphones. JP1 was used above 100 kHz; JP2 and JP4 were used at and below 100 kHz. Figure 2(b) compares the transmitting voltage response (the SPL at 15 cm in response to a 1-V jawphone input) mea-

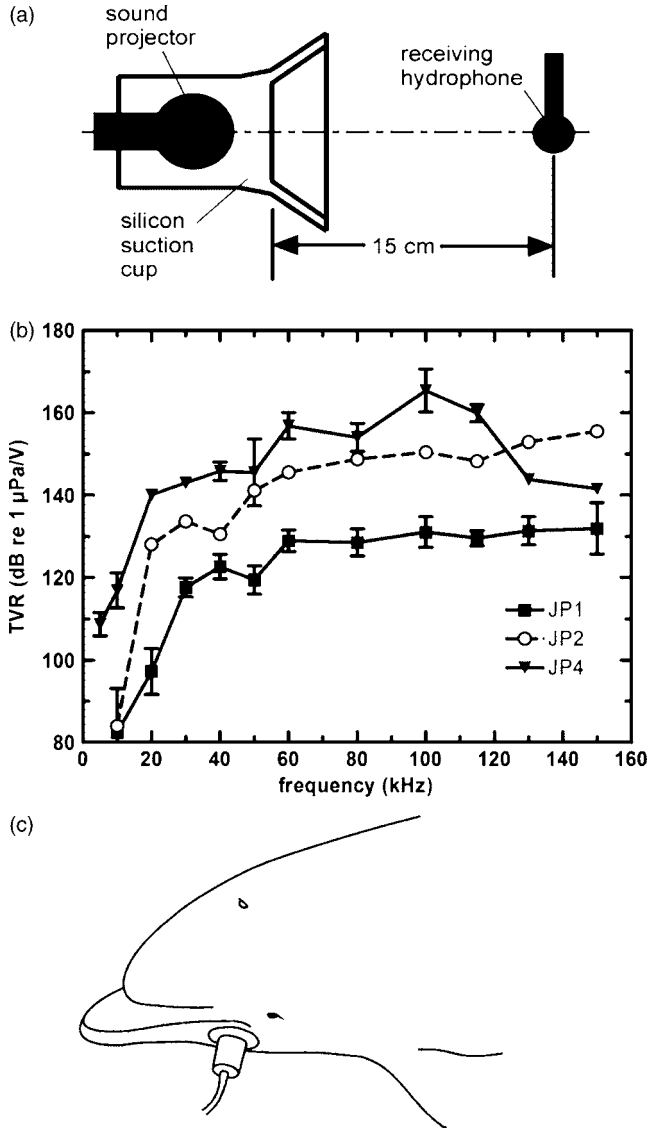


FIG. 2. (a) Jawphone construction and calibration arrangement. (b) Transmitting voltage response (TVR) from jawphones measured at 15 cm. (c) Example of jawphone position on subject.

sured from the various jawphones. Figure 2(c) shows the attachment site of the jawphone for testing the left ear.

Two types of surface electrodes were used to detect evoked potentials: disposable electrodes consisting of a flexible, conductive, self-adhesive Ag/Ag–Cl pad (Ambu Neuroline 710 series) and reusable 6 and 10 mm diameter gold cup electrodes (e.g., Grass FH-E6G series) embedded in 25 mm diameter silicon suction cups. Electrode placement sites were dried using gauze pads and/or alcohol swabs prior to placement. Disposable electrodes were subsequently covered with waterproof bandages (Nexcare Absolute Waterproof) to prevent contact with water periodically sprayed over the dolphin. Reusable electrodes were protected from contact with water by the suction cup itself. A three-electrode configuration was used: the active (+) electrode was placed near the vertex, approximately 10 cm posterior of the blowhole and offset approximately 2 cm contralateral of the ear being tested; the reference (−) electrode was placed contralateral of the ear being tested, just posterior to the external auditory meatus; and the ground (common) electrode was placed on the subject's back near the dorsal fin. When using reusable suction cup electrodes, the exact position of the active electrode varied depending on the shape of the subject's head, i.e., in some cases, curvature of the head prevented attachment of the suction cup electrodes at the desired site behind the blowhole. The electrode signals were amplified and filtered using a biopotential amplifier (Grass IP-511). The biopotential amplifier output was proportional to the voltage difference between the active and reference electrodes. The biopotential amplifier gain was fixed at 100 000. High and low-pass filters varied from 100 to 300 Hz and 3 to 10 kHz, respectively, depending on the stimulus modulation frequency and measurement sampling rate. Additional filtering (Krohn-Hite 3C, bandpass from 0.1 to 5 kHz) was sometimes used, depending on signal quality. The resulting signal was digitized by the PCI-MIO-16E-1 at either 15 or 20 kHz.

TABLE I. Jawphone properties.

Jawphone	Projector	Suction cup base diameter (cm)	Suction cup length (cm)	Frequency range (kHz)
JP1	EDO 1066	2.5	4.0	115–150
JP2	EDO 1066	4.0	6.0	10–100
JP4	ITC 1042	4.0	6.0	10–100

2. Stimuli

Sound stimuli consisted of sinusoidal amplitude modulated (SAM) tones, resulting in steady-state, periodic AEPs with a fundamental at the modulation frequency. An evoked response such as this that is phase-locked to the amplitude modulation rate is sometimes called the envelope following response (EFR) or auditory steady-state response (ASSR) (Dolphin and Mountain, 1993; Rance *et al.*, 1993; Dolphin *et al.*, 1995). Eleven carrier frequencies, from 10 to 150 kHz were tested. These frequencies were chosen to cover the effective range of bottlenose dolphin echolocation (Au, 1993). The modulation rate was 1 kHz. Modulation rate transfer function (MRTF) measurements have shown this rate to produce a strong evoked response in *T. truncatus* (Dolphin *et al.*, 1995; Supin and Popov, 2000; Finneran and Houser, 2004). This was verified by measuring MRTFs on BLU and WEN prior to threshold measurements. The amplitude modulation function was chosen to produce a stimulus with a three-component frequency spectrum (carrier plus sidebands).

BLU and WEN were tested using both intermittent and continuous SAM tones while BUG and BUS were tested only with continuous SAM tones. For intermittent presentations, tone durations varied from 12 to 15 ms (the majority were 13 ms), with a 1 ms cosine envelope rise and fall. Intermittent stimuli were presented at a rate of approximately 50/s. Continuous stimuli were presented for approximately 10 s.

Stimulus generation and evoked response recording were controlled using custom software. Artifact rejection was implemented in software; signals greater than 20 μV

were normally rejected. Frequency analysis was performed on 19 ms epochs for continuous stimuli and 10–13 ms epochs for intermittent stimuli. For continuous stimuli, the first epoch containing the response latency and initial transients was not included in the frequency analysis; for intermittent stimuli, the portions of the evoked response corresponding to the rise/fall of the stimulus were not included. Time durations for frequency analysis were constrained to integral multiples of 1 ms (the harmonic period of the modulation) to avoid spectral leakage in the frequency domain. Coherent averaging (typically 500–1000 averages) was performed in the frequency domain; this allowed the mean amplitude and phase at the modulation frequency, as well as confidence intervals, to be obtained.

Figure 3(a) shows examples of ASSR waveforms measured from BLU in response to an intermittent 30 kHz SAM tone. At high SPLs, responses are easily detectable from the waveforms; however, at lower SPLs the responses become indistinguishable from noise in the time domain. Figure 3(b) shows the frequency spectra of the waveforms from Fig. 3(a). The primary feature of interest for detection of the evoked response is the spectral peak at the 1 kHz modulation frequency.

3. Response detection

The magnitude-squared coherence (MSC) was calculated during each AEP measurement and used to objectively determine if the measured AEP component at the modulation frequency was statistically different from noise (Dobie and Wilson, 1989; Dobie, 1993; Dobie and Wilson, 1996). MSC is a ratio of the power (at a single frequency) contained in the "grand" coherent average to the average of the powers within individual "segments" or "subaverages" of the total data stream. The MSC provides a ratio of the signal power to signal-plus-noise power and varies from zero (all noise) to one (all signal). The MSC calculation used 20 subaverages. Critical values for MSC, using $\alpha=0.01$, were obtained from Amos and Koopmans (1963) and Brillinger (1978). If the calculated MSC was greater than the critical value, the AEP at the modulation frequency was considered to be detected. This process of objective response detection provided a yes/no result for each AEP measurement and permitted adaptive procedures for adjusting stimulus levels (e.g., modified staircase technique). MSC calculations were done on-line during test procedures to permit the use of adaptive search techniques during data collection.

4. Data collection

Data collection began with a stimulus SPL of 80 dB re 1 μPa. An automated staircase technique was used to adjust stimulus levels to quickly reach threshold. Stimulus levels were decreased after each detected response (a "hit") and increased after each measurement without a detected response (a "miss"). The amount the stimulus was raised or lowered (the step size) started at 30 dB and was reduced after each "reversal"—a transition from a hit to a miss or a miss to a hit. The step size was multiplied by 0.4 after each miss-hit reversal and 0.6 after a hit-miss reversal. Unequal multipliers were used to avoid repeated testing at the same levels, provide a more continuous distribution of stimulus

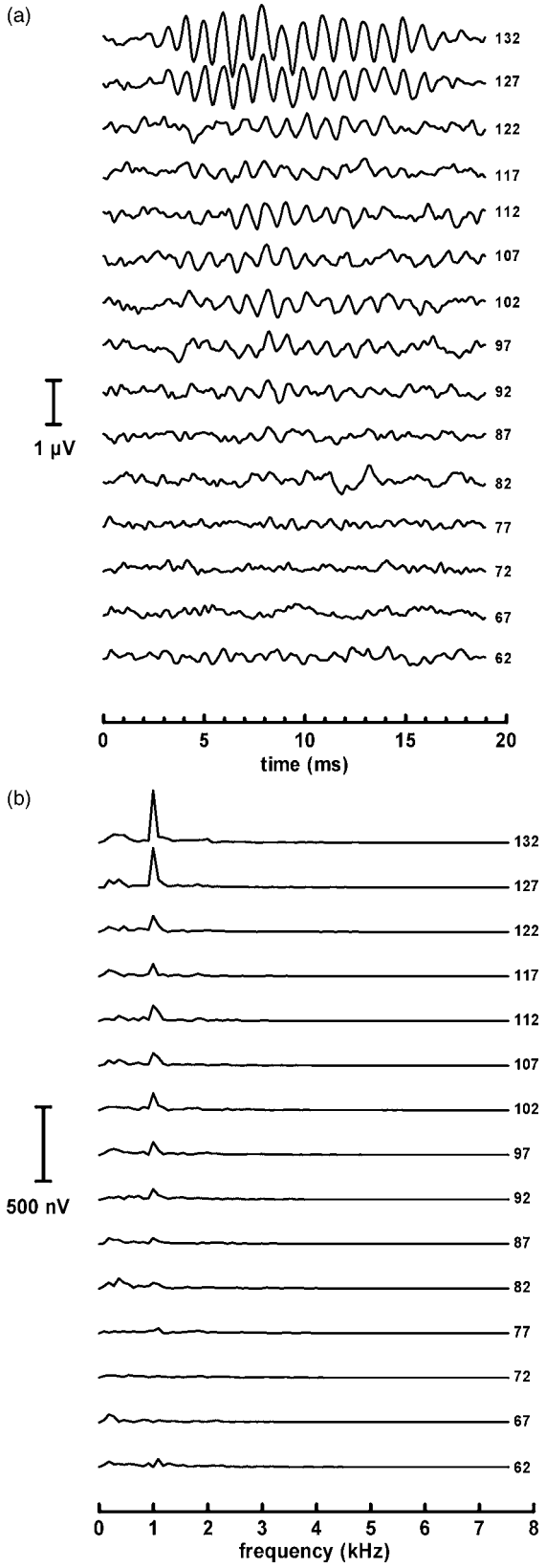


FIG. 3. (a) Steady-state AEP (a) waveforms and (b) frequency spectra measured at different stimulus SPLs for subject BLU, 30 kHz carrier frequency, 1 kHz modulation frequency. The frequency spectra were calculated from a 10 ms analysis window. The numbers next to each series indicate the stimulus SPL (dB re 1 μPa).

levels tested, and to increase the chance of obtaining a hit on the next trial. Multiple samples were not taken at the individual stimulus levels to avoid an uneven sampling of stimuli and ensure that time was available for sampling at all of the frequencies of interest. The staircase was terminated when the step size was reduced below 4 dB.

Considering only responses that were detected, a linear regression was then performed on the AEP amplitude versus stimulus SPL data. Detected responses with amplitudes greater than 400 nV were excluded from the regression analysis. Testing was concluded if the regression r^2 value from a minimum of four detected responses reached 0.9. If the minimum r^2 value was not obtained, additional measurements were conducted at SPLs selected to fill the gaps within the AEP amplitude versus stimulus SPL data. With each additional detected response, another regression was performed on the data series. This process was repeated until the criterion r^2 was met or until a maximum of eight detections at different SPLs was made. When necessary, additional stimulus levels were tested to ensure sufficient data for post-hoc threshold estimates. (This was most often required when a plateau in the AEP amplitude versus SPL data confounded the automated logic.) Using these techniques, a single ear could be tested (at eleven frequencies) in approximately 45 min.

5. Threshold estimation

A post-hoc linear regression technique was used to estimate hearing thresholds for each of the frequencies tested. Linear regression was first performed on the four detected responses with the lowest amplitudes. If the regression r^2 was less than 0.9, additional points obtained at consecutively higher stimulus levels were added to the regression analysis until the r^2 criterion of 0.9 was met or an obvious plateau was reached in the AEP amplitude data. The AEP threshold was defined as the stimulus SPL at which the extrapolated regression line reached 0 V. Almost all regressions were based on more than four points; however, in a few cases, only three points could be included in the regression. If a three point regression could not be performed using detected responses, the data at that frequency were excluded.

Figure 4 shows the amplitude (upper panel) and phase angle (lower panel) at the modulation frequency for the data shown in Fig. 3 and illustrates the threshold estimation procedure. The filled symbols represent detected responses according to the MSC calculation; open symbols indicate measurements that were not statistically different than noise. The solid line shows the linear regression applied according to rules outlined above. In this case, the four lowest amplitude responses resulted in an $r^2=0.92$, so no additional points were included. The dashed line shows the extrapolation of the regression to estimate the threshold (72 dB re 1 μPa).

Each ear of WEN and BLU was tested at least three times and the mean of the threshold estimates are reported here. BUG and BUS were only tested once (except BUG's right ear which was tested twice on the same day at 40 kHz).

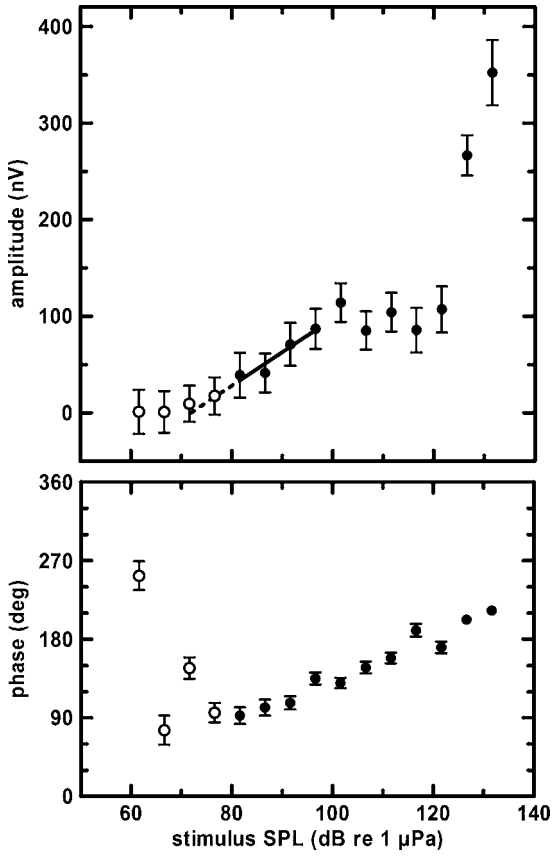


FIG. 4. Steady-state AEP (rms) amplitude and phase as functions of stimulus SPL for the data shown in Fig. 3. Symbols represent mean values; error bars show the 95% confidence intervals. The filled symbols represent detected responses (MSC values were greater than the critical value, MSC_{crit}); the open symbols represent measurements without statistically detected responses ($MSC < MSC_{crit}$). The solid line shows the linear regression performed on the lowest four detected responses ($r^2=0.92$). The dashed line is an extrapolation of the regression line to 0 V. The stimulus SPL at which the dashed line intersects the x-axis is the estimated threshold (72 dB re 1 μPa).

C. Behavioral measurements

Behavioral thresholds were measured underwater in SD Bay (all subjects) and an above-ground pool (BLU only). Ambient noise spectral density levels in the pool ranged from approximately 50 dB re 1 $\mu Pa^2/Hz$ (at 10 kHz) to 40 dB re 1 $\mu Pa^2/Hz$ (50 kHz); noise levels in San Diego Bay were approximately 20 dB higher (Finneran *et al.*, 2005).

Behavioral thresholds for BLU and WEN were obtained before or concurrently with AEP measurements (data were collected over a period of several months). BUG and BUS were tested under "blind" conditions; AEP measurements were conducted first, then behavioral measurements were performed by independent investigators. AEP thresholds were not disclosed until the behavioral testing was complete.

Table II compares test parameters for each subject. All subjects were behaviorally tested using 500 ms duration test tones with linear rise/fall times of 20 ms or larger. BLU, BUS, and WEN were trained to whistle in response to hearing test tones and to remain quiet otherwise. BUG was trained to press a response paddle in the presence of a tone. Fifty-percent of the trials were no-tone ("catch") trials, except for a few sessions with BUS where the tone/no-tone

TABLE II. Behavioral hearing test parameters.

	Subject			
	BLU	BUG	BUS	WEN
Location	Pool, SD Bay	SD Bay	SD Bay	SD Bay
Method	(Descending) Staircase	(Descending) Staircase	(Descending) Staircase	(Descending) Staircase
Trial presentation	MFR	Single interval	Single interval	Single interval
Response	Whistle	Paddle press	Whistle	Whistle
Tone duration	500 ms	500 ms	500 ms	500 ms
Step size	2 dB	2 dB	2 dB	2 dB
Threshold estimate	10 rev. (Pool) 6 rev. (SD Bay)	10 rev.	10 rev.	6 rev.
Reinforcement	Variable	Fixed	Fixed	Fixed

ratio was adjusted to 65/35 to maintain a false alarm rate of around 10%–15%. Stimulus SPLs were adjusted using a staircase procedure (Cornsweet, 1962) with a 2-dB step size for threshold estimates. Threshold estimates were based on 6 (WEN and BLU in SD Bay) or 10 (BUG, BUS, and BLU in the pool) consecutive hit-miss or miss-hit reversals. BUG, BUS, and WEN were tested in the context of a single interval experiment where the subject was rewarded (correct response) or recalled (incorrect response) after each trial. The behavioral threshold procedure for BLU was based on the Method of Free Response (MFR) (Egan *et al.*, 1961) and described in detail by (Finneran *et al.*, 2005). For all subjects, thresholds from independent sessions were averaged to yield a mean threshold estimate based on minimum of two (BUG, BUS) or three (BLU, WEN) independent measurements. At most frequencies, WEN and BLU were tested 4–6 times.

D. Statistical analysis

Left and right ear AEP thresholds for BLU obtained with intermittent and continuous stimuli were compared using a one-way ANOVA (GraphPad Software, 2003). A Tukey-Kramer post-test was used to compare all four data sets: Left ear, continuous; left ear, intermittent; right ear, continuous; right ear, intermittent.

III. RESULTS

Figure 5 presents the behavioral thresholds measured for BLU, BUG, BUS, and WEN. Three of the four subjects showed varying degrees of high-frequency hearing loss, with upper cutoff frequencies ranging from approximately 80 kHz (BUG) to 30 kHz (BUS). WEN possessed full range hearing with an upper frequency cutoff of approximately 140 kHz. Low-frequency thresholds in SD Bay were elevated, presumably from masking caused by the relatively high ambient noise levels.

Figure 6 compares AEP thresholds for BLU measured using both continuous and intermittent stimuli. For this subject, responses could not be detected above 100 kHz at the highest SPLs the jawphones could generate. No significant differences between threshold estimates using intermittent or continuous tones were observed for BLU ($p > 0.05$); for the remaining analyses, intermittent and continuous data were pooled to produce a single mean threshold at each frequency for each ear.

Figure 7 compares the left and right ear AEP thresholds estimates for BLU to behavioral thresholds measured in SD Bay and the pool. All threshold estimates reveal poor high frequency hearing, with AEP and pool behavioral threshold estimates rising at about 47 dB/octave above 30 kHz. Below 40 kHz, a clear separation exists between the behavioral thresholds measured in the pool and SD Bay, most likely from the ambient noise differences between the two environments.

Figure 8 compares BLU's AEP thresholds (ordinate) to the behavioral thresholds measured in the pool (abscissa). The dashed line represents the ideal case in which behavioral and AEP thresholds are identical. AEP threshold estimates tended to be higher than the behavioral estimates. Differences between thresholds (AEP minus behavioral) increased with threshold magnitude and ranged from 0 to +18 dB, with a mean of 8 dB. AEP thresholds were strongly corre-

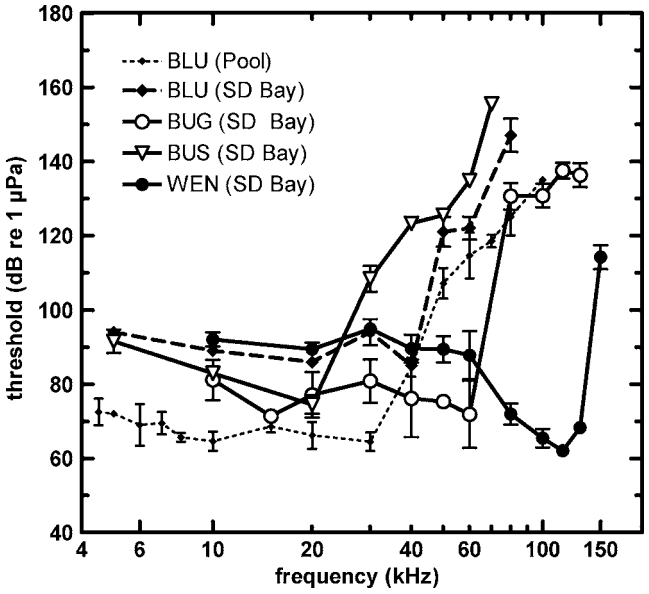


FIG. 5. Behavioral thresholds for BLU, BUG, BUS, and WEN. Symbols represent mean values; error bars show the mean ± one standard deviation.

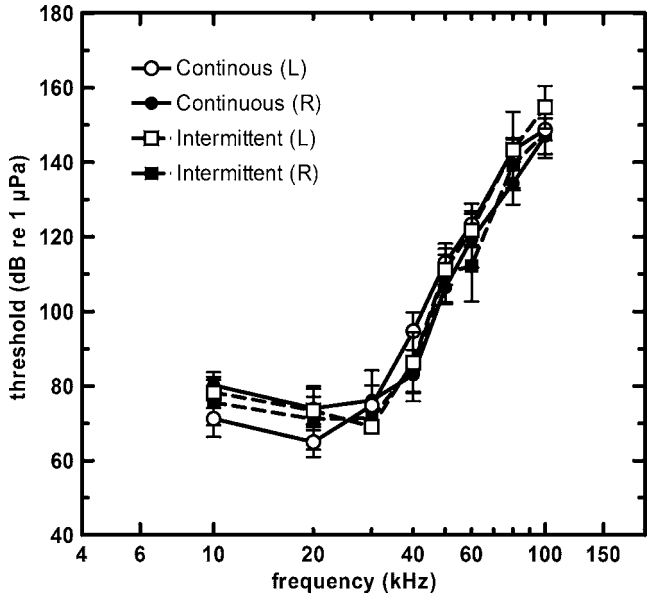


FIG. 6. Comparison of in-air steady-state AEP hearing thresholds for subject BLU obtained from left (L) and right (R) ears using continuous and intermittent stimuli. Symbols represent mean values; error bars show the mean ± one standard deviation. There were no statistically significant differences between the thresholds at a particular frequency.

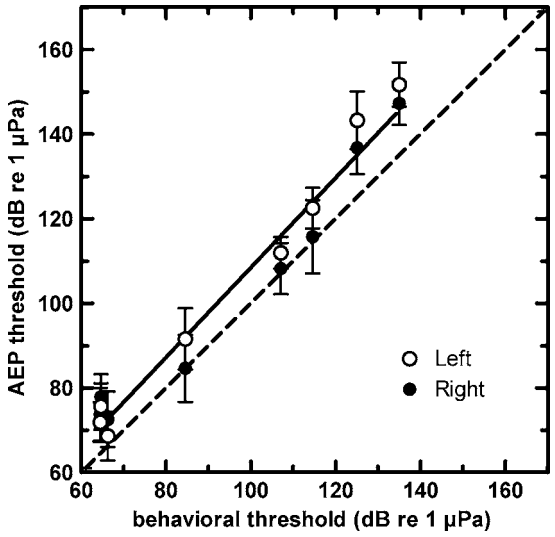


FIG. 8. In-air AEP threshold estimates vs behavioral thresholds measured in the pool for subject BLU. Symbols are mean values. Error bars represent ± one standard deviation. The solid line is a linear regression performed on the experimental data.

lated with behavioral thresholds across the range of hearing (r=0.98) with a regression equation slope of 1.1. The mean AEP thresholds from both ears ranged from 3 to 15 dB above the pool behavioral thresholds.

Figures 9–11 compare AEP and behavioral thresholds for WEN, BUG, and BUS, respectively. Both BUG and BUS exhibited poor high-frequency hearing, with thresholds increasing above 60 and 20 kHz, respectively. In contrast, WEN's thresholds remained relatively low across the expected functional range for dolphins and did not increase until the frequency exceeded 130 kHz. For all three subjects, AEP threshold estimates and behavioral estimates agreed closely as to the upper cutoff frequency beyond which thresholds increased sharply.

Figure 12 compares the AEP thresholds to the SD Bay behavioral thresholds for all four subjects. AEP threshold estimates below 40 kHz were generally lower than behavioral estimates, probably because the behavioral thresholds were masked by ambient noise at these frequencies. At high frequencies, AEP estimates tended to be higher than behavioral estimates, especially in WEN, the only subject with relatively low thresholds at high frequencies. Mean AEP thresholds (average of both ears) were within −26 to +20 dB of the SD Bay behavioral thresholds for all four subjects. AEP threshold estimates for individual ears ranged

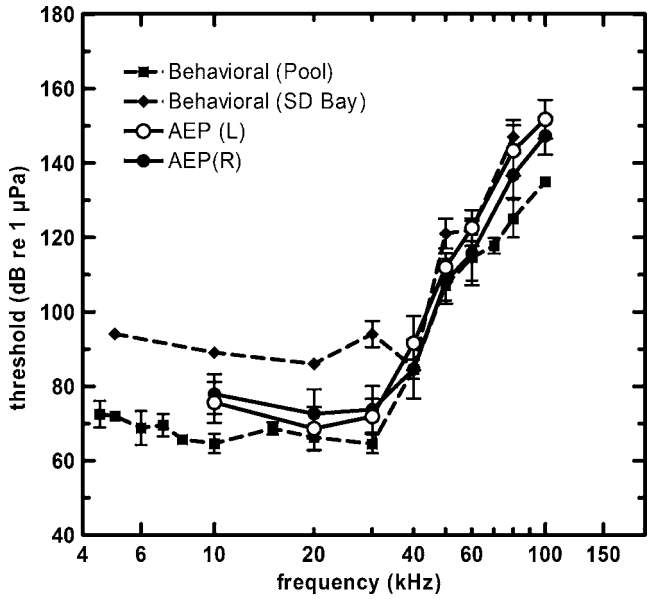


FIG. 7. In-air AEP and underwater behavioral hearing thresholds measured for BLU. Symbols are mean values. Error bars represent ± one standard deviation.

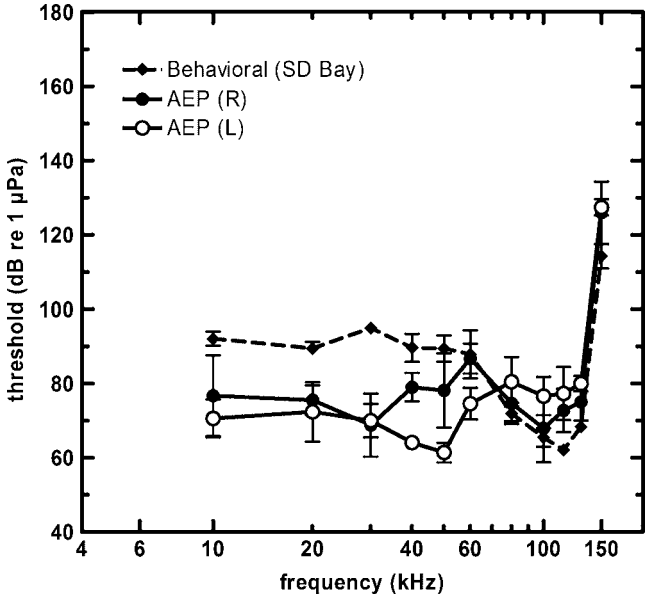


FIG. 9. In-air AEP and underwater behavioral hearing thresholds measured for WEN. Symbols are mean values. Error bars represent ± one standard deviation.

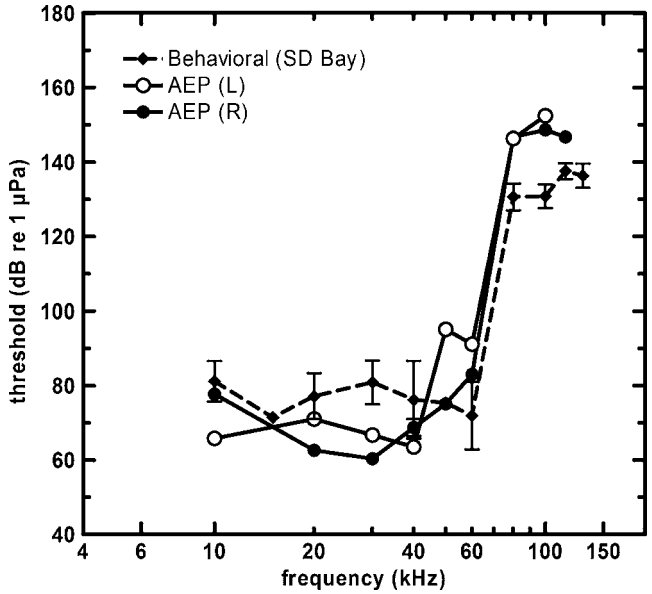


FIG. 10. In-air AEP and underwater behavioral hearing thresholds measured for BUG. Symbols are mean values. Error bars represent ± one standard deviation.

from −28 to +22 dB. Excluding comparisons at 40 kHz and below, where masking likely affected threshold measurements, mean AEP thresholds (average of both ears) were within −20 to +20 dB of the SD Bay behavioral thresholds for all four subjects, with a mean difference of +6 dB.

IV. DISCUSSION

A. Behavioral thresholds

Comparisons of behavioral thresholds between the BLU and WEN for SD Bay and the pool suggest that thresholds measured in SD Bay were masked by ambient noise at frequencies up to at least 40 kHz, and possibly higher. BUS and

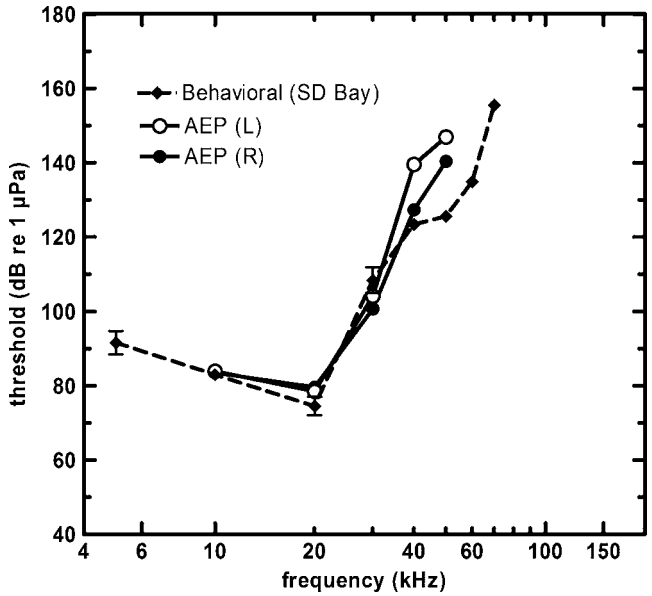


FIG. 11. In-air AEP and underwater behavioral hearing thresholds measured for BUS. Symbols are mean values. Error bars represent ± one standard deviation.

BUG had similar thresholds at low frequencies; however, their thresholds were consistently lower than thresholds for BLU and WEN. The similar shapes, but roughly 10 dB shift in amplitude, of the threshold curves for thresholds for WEN and BUG suggest that motivational, methodological, or calibration differences may have affected threshold estimates. An implicit assumption is often made that behavioral thresholds are correct and that any differences between AEP and behavior threshold estimates are attributed to errors in the AEP estimate. It must be remembered that the classical behavioral threshold is a statistical concept depending on experimental and motivational factors as well as the subject's sensitivity. Changes in a subject's response bias or differences in threshold estimation methods will affect the numeric thresholds obtained. Until standard methods for marine mammal behavioral threshold estimates are developed, comparisons between AEP and behaviorally measured data should consider the details of how the data were collected.

B. AEP thresholds

1. AEP methodology

The relationship between the AEP amplitude and the stimulus SPL, called the input/output function or AEP growth curve, is nonlinear (Fig. 4). The plateau and subsequent increase in the input/output function have been proposed to result from auditory adaptation during stimulation, followed by increased activation from spectral spreading at higher SPLs (Supin *et al.*, 2001). The shape may also reflect basilar membrane displacement, which exhibits a similar nonlinear response believed to arise from a combination of active and passive processes (Pickles, 1988). Identification of the linear regions and plateaus within the nonlinear input/output function requires sufficient measurements and under sampling the curve can result in unsatisfactory threshold estimates (Lins *et al.*, 1995; Vander Werff and Brown, 2005). The existence of a plateau can confound automatic threshold testing logic and, if occurring sufficiently close to the noise floor, prevent the application of regression techniques.

As the stimulus SPL is lowered, the AEP amplitude will, at some point, become indistinguishable from noise. Discriminating AEP responses from noise based only on visual observation cannot avoid some degree of subjectivity and a dependence on observer skill. Furthermore, AEP threshold estimates based on a regression technique can be markedly skewed by inclusion/rejection of individual points within the regression (Vander Werff and Brown, 2005).

This study employed an objective technique (the MSC) to distinguish responses from noise and avoid problems associated with observer subjectivity. This allowed objective decisions to be made regarding the presence or absence of a response. Data collection and stimulus adjustment could thus be automated to a great extent, since the presence or absence of a response could be used as a basis for adjusting the stimulus SPL for the next trial. Figure 4 illustrates the utility of statistical approaches; without objective response detection, one may be tempted to include the three highest amplitude undetected responses in the regression, producing a potentially erroneous threshold estimate.

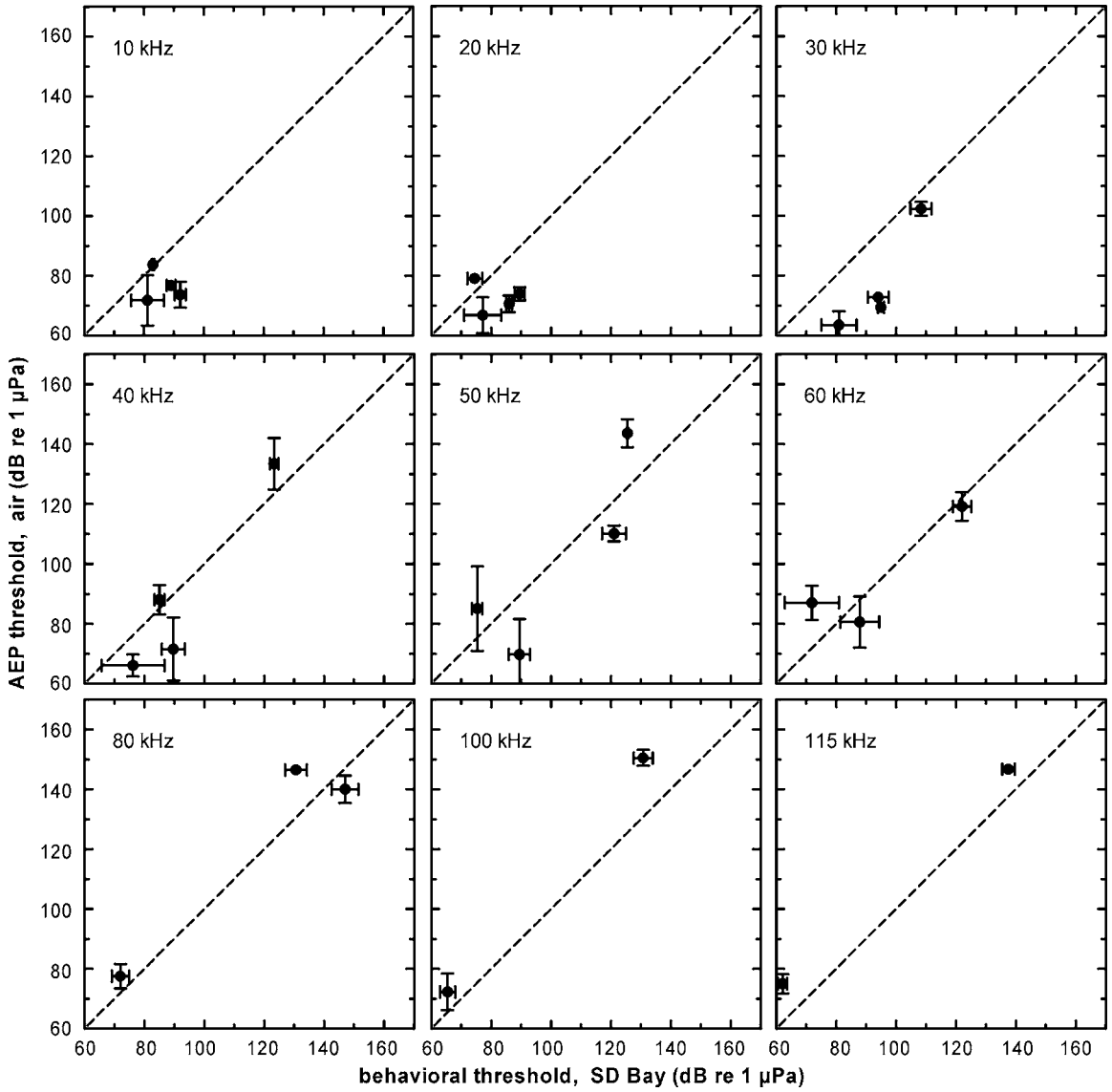


FIG. 12. In-air AEP thresholds plotted versus the San Diego Bay behavioral thresholds. Each panel shows data from all four subjects (if possible) for a single frequency. Error bars represent ± one standard deviation. The dotted line indicates where perfect agreement (equivalence) between the threshold predictions would occur. Note that frequencies below 40 kHz are likely to be masked, augmenting differences between behavioral and AEP estimates of hearing sensitivity at these frequencies.

2. Relationships between AEP and behavioral data

Because they were not affected by masking noise in SD Bay, behavioral thresholds in the pool provide more appropriate comparisons to the AEP thresholds than behavioral data obtained in SD Bay. The largest differences between the two threshold estimates occurred at the lowest (10 kHz) and highest frequencies (80 and 100 kHz). The magnitude and direction of the differences between AEP and behavioral thresholds agree with previous comparisons in humans with normal hearing and sensorineural hearing loss (Lins *et al.*, 1995; Rance *et al.*, 1995; Aoyagi *et al.*, 1999; Dimitrijevic *et al.*, 2002; Vander Werff and Brown, 2005) and show similarly strong correlations in predicted thresholds (Rance *et al.*, 1995; Vander Werff and Brown, 2005). The results reveal that in-air AEP thresholds measured using the techniques described here are useful proxies for underwater behavioral data, especially with respect to the audiogram shape and upper cutoff frequencies.

3. Left/right ear differences

Differences between BLU's left and right thresholds were small (relative to the standard deviation), ranging from 2 to 7 dB, whereas WEN's left ear had 12–17 dB lower thresholds between 40 and 60 kHz, but 6–10 dB higher thresholds between 80 and 130 kHz. BUG and BUS had similar thresholds for their left and right ears at most frequencies (most frequencies were only tested once, so there are no variance measures for these data). The ability to differentiate the relative hearing sensitivity between two ears would be beneficial in clinical diagnoses of auditory system dysfunction as well as the characterization of auditory system health in animals that might be used in psychoacoustic research. At present it is not clear how best to compare behavioral thresholds obtained binaurally with monaural AEP data. The mean of the left and right ear thresholds seems a reasonable approach, especially when the thresholds are

similar; however, one could imagine an animal with large sensitivity differences between ears relying on the more sensitive ear during binaural tests.

4. Continuous and intermittent stimuli

No difference between threshold estimates for BLU was found when using either continuous or intermittent stimuli. Continuous stimuli allow data to be acquired more rapidly because the same amount of time domain response data will take less time to acquire using continuous stimuli than with intermittent stimuli. The duration of the time segment available for spectral analysis dictates the resulting frequency resolution (frequency "bin" width). Assuming that the noise spectral density is relatively flat, smaller frequency bins will result in lower noise levels, making detection of responses near threshold easier. So, for identical data collection time periods, continuous stimuli would allow longer time segments for analysis and lower noise levels. The primary advantage of intermittent stimuli is that the response latency may be observed. This allows one to clearly separate physiological response from stimulus artifacts, which may have zero latency. In many cases, this advantage can offset the additional time that is required for data collection.

5. Approach limitations

The AEP threshold estimation in delphinids using the combined techniques described here is promising given that estimates demonstrate a similar degree of variation between behavioral and physiological estimates that are observed with other species. However, electrophysiological techniques have limitations when applied as surrogates for behavioral tests. Some limitations may be controlled from a methodological standpoint (e.g., the two methods typically do not use equivalent stimuli), while others are inherent to physiological measures (e.g., animals are "noisy" because of spontaneous brain activity and myogenic potentials). Differences between AEP and behavioral thresholds quantified for the methods used in this study should not be broadly applied because they are likely to vary as a function of methodology. For example, underwater, direct field stimulation of a subject may result in sound transmission to the ear via acoustic pathways that are not used when the sound projector is coupled to a point on the lower jaw of an animal that is out of the water.

The delivery of acoustic stimuli through a jawphone has been previously applied in studies of dolphin audition (Moore *et al.*, 1995; Mohl *et al.*, 1999; Brill *et al.*, 2001), but no standard approach exists for jawphone calibration. In this study, the jawphone was calibrated underwater as would be a free transducer with the receiving hydrophone 15 cm away. This approach does not account for potential distortion of the stimulus signal resulting from coupling the jawphone to the animal. The impact of the coupling is unknown. Nor can it presently be verified whether the direct path distance from the coupling point to the auditory bulla is representative of the true acoustic path to the ear, thus validating that distance for use in future calibrations. However, jawphones constructed for this project have effectively been calibrated against the animals on which they were used, i.e., by comparison of AEP and behavioral threshold estimates. The agreement between threshold estimates, within the variation expected between AEP and behavioral methods, suggests that the calibration used for this project is suitable for the types of animals and circumstances under which the jawphone is used.

6. Approach benefits

The AEP approach used here to estimate hearing thresholds in dolphins has characteristics that make it attractive for use on captive, stranded or wild caught marine mammals. The statistical approach to signal detection and defined rules for inclusion of data points used in the regression analysis diminish the potential bias introduced by the subjective selection of data for threshold estimation. Increasing the application of such objective approaches to AEP threshold estimates can help reduce interstudy variability and facilitate comparisons of results from different studies. Because it employs an automated search algorithm, automated statistical detection of signals and threshold estimation, and automatic transition between test frequencies, the approach permits rapid testing across a broad range of frequencies. The testing of 11 frequencies from 10 to 150 kHz was achieved in approximately 45 min, and the sensitivity of both ears was obtained in 1.5 h. It is important to note that the total of 1.5 h is a single test session in which the full range of hearing of both ears was tested. This relatively short testing time substantially exceeds the rate at which behavioral hearing tests can be performed (e.g., weeks to months) and permits multiple animals to be rapidly tested in sequence. Due to the potential impact of handling stress, the ability to rapidly test animals is a major consideration when dealing with stranded and rehabilitated animals and efforts to collect AEP threshold estimates from these animals will benefit from the implementation of rapid data acquisition systems.

V. CONCLUSIONS

The automated methodology described here offers advantages to assessing hearing sensitivity in delphinids by increasing the rapidity of data collection and the objectivity of the data analysis. Threshold estimates derived from this approach in four bottlenose dolphins demonstrated agreement with behavioral thresholds: (1) differences were consistent with the magnitude and variation of those observed in similar clinical studies; (2) the upper cutoff frequency of hearing sensitivity determined by both methods were close. AEP thresholds measured using the techniques described here are useful proxies for underwater behavioral data.

ACKNOWLEDGMENTS

We thank R. Dear and C. Schlundt for invaluable help in data collection and analysis and L. Green, C. Quintos, L. Lewis, E. Linberg for animal training. Behavioral audiograms for BUS and BUG were provided by D. Katsak and L. Dankiewicz. S. Ridgway, D. Carder, and P. Moore provided much helpful advice and technical assistance. Financial support was provided by the U.S. Office of Naval Research,

Marine Mammal S&T Program and the SSC San Diego In-House Laboratory, Independent Research (ILIR) program.

Amos, D. E., and Koopmans, L. H. (**1963**). *Tables of the Distribution of the Coefficient of Coherence for Stationary Bivariate Gaussian Processes* (Sandia Corporation, Livermore, California), p. 328.

Andre, M., Supin, A., Delory, E., Kamminga, C., Degollada, E., and Alonso, J. M. (**2003**). "Evidence of deafness in a striped dolphin, Stenella coeruleoalba," Aquat. Mamm. **29**, 3–8.

Aoyagi, M., Suzuki, K., Yokota, M., Furuse, H., Watanabe, T., and Ito, T. (**1999**). "Reliability of 80-Hz amplitude-modulation-following response detected by phase coherence," Audiol. Neuro-Otol. **4**, 28–37.

Au, W. W. L., and Moore, P. W. B. (1984). "Receiving beam patterns and directivity indices of the Atlantic bottlenosed dolphin (Tursiops truncatus)," J. Acoust. Soc. Am. **75**, 255–262.

Au, W. W. L., and Moore, P. W. B. (**1990**). "Critical ratio and critical bandwidth for the Atlantic bottlenosed dolphin," J. Acoust. Soc. Am. **88**, 1635–1638.

Au, W. W. L. (**1993**). "*The sonar of dolphins*," (Springer Verlag, NY).

Awbrey, F. T., Thomas, J. A., and Kastelein, R. A. (**1988**). "Low-frequency underwater hearing sensitivity in belugas, Delphinapterus leucas," J. Acoust. Soc. Am. **84**, 2273–2275.

Brill, R. L., 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.

Brill, R. L., Sevenich, M. L., Sullivan, T. J., Sustman, J. D., and Witt, R. E. (**1988**). "Behavioral evidence for hearing through the lower jaw by an echolocating dolphin (Tursiops truncatus)," Mar. Mammal Sci. **4**, 223–230.

Brillinger, D. R. (**1978**). "A note on the estimation of evoked response," Biol. Cybern. **31**, 141–144.

Bullock, T. H., Ridgway, S. H., and Nobuo, S. (**1971**). "Acoustically evoked potentials in midbrain auditory structures in sea lions (pinnipedia)," Z. vergl. Physiologie **74**, 372–387.

Cornsweet, T. N. (**1962**). "The staircase method in psychophysics," Am. J. Psychol. **75**, 485–491.

Dimitrijevic, A., John, M. S., Van Roon, P., Purcell, D. W., Adamonis, J., Ostroff, J., Nedzelski, J. M., and Picton, T. W. (**2002**). "Estimating the audiogram using multiple auditory steady-state responses," J. Am. Acad. Audiol **13**, 205–224.

Dobie, R. A. (**1993**). "Objective response detection," Ear Hear. **14**, 31–35.

Dobie, R. A., and Wilson, M. J. (**1989**). "Analysis of auditory evoked potentials by magnitude-squared coherence," Ear Hear. **10**, 2–13.

Dobie, R. A., and Wilson, M. J. (**1996**). "A comparison of t test, F test, and coherence methods of detecting steady-state auditory-evoked potentials, distortion-product otoacoustic emissions, or other sinusoids," J. Acoust. Soc. Am. **100**, 2236–2246.

Dolphin, W. F. (**1997**). "Electrophysiological measures of auditory processing in odontocetes," Bioacoustics 8, 79–101.

Dolphin, W. F., Au, W. W., Nachtigall, P. E., and Pawloski, J. (**1995**). "Modulation Rate Transfer Functions to Low-Frequency Carriers in Three Species of Cetaceans," J. Comp. Physiol., A **177**, 235–245.

Dolphin, W. F., and Mountain, D. C. (**1993**). "The envelope following responce (EFR) in the Mongolian gerbil to sinusoidally amplitude-modulated signals in the presence of simultaneously gated pure tones," J. Acoust. Soc. Am., **94**, 3215–3226.

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., Carder, D. A., Schlundt, C. E., and Ridgway, S. H. (**2005**). "Temporary threshold shift (TTS) in bottlenose dolphins (Tursiops truncatus) exposed to mid-frequency tones," J. Acoust. Soc. Am. **118**, 2696–2705.

Finneran, J. J., and Houser, D. S. (**2004**). "A portable system for marine mammal auditory-evoked potential measurements," J. Acoust. Soc. Am. **115**, 2517(A).

Finneran, J. J., Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (**2002**). "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.

GraphPad Software (**2003**). "GraphPad Prism" (GraphPad Software, San Diego, CA).

Houser, D. S., Finneran, J. J., Carder, D. A., Van Bonn, W., Smith, C. R., Hoh, C., Mattrey, R., and Ridgway, S. H. (**2004**). "Structural and functional imaging of Bottlenose Dolphin (Tursiops truncatus) cranial anatomy," J. Exp. Biol. **207**, 3657–3665.

Johnson, C. S. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bioacoustics*, edited by W. N. Tavolga (Pergamon, Oxford), pp. 247–260.

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.

Kastelein, R. A., Bunskoek, P., and Hagedoorn, M. (**2002**). "Audiogram of a harbor porpoise (Phocoena phocoena) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 334–344.

Kastelein, R. A., Janssen, M., Verboom, W. C., and de Haan, D. (**2005**). "Receiving beam patterns in the horizontal plane of a harbor porpoise (Phocoena phocoena)," J. Acoust. Soc. Am. **118**, 1172–1179.

Lins, O. G., Picton, P. E., Picton, T. W., Champagne, S. C., and Durieux-Smith, A. (**1995**). "Auditory steady-state responses to tones amplitude-modulated at 80–110 Hz," J. Acoust. Soc. Am. **97**, 3051–3063.

Mohl, B., Au, W. W. L., Pawloski, J., and Nachtigall, P. E. (**1999**). "Dolphin hearing: Relative sensitivity as a function of point of application of a contact sound source in the jaw and head region," J. Acoust. Soc. Am. **105**, 3421–3424.

Moore, P. W., and Schusterman, R. J. (**1987**). "Audiometric assessment of northern fur seals, Callorhinus ursinus," Marine Mammal Sci. **3**, 31–53.

Moore, P. W. B., Pawloski, D. A., and Dankiewicz, L. (**1995**). "Interaural time and intensity difference thresholds in the bottlenose dolphin (Tursiops truncatus)," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil Publishers, Woerden, The Netherlands).

Nachtigall, P. E., Supin, A., Pawloski, J., and Au, W. W. L. (**2004**). "Temporary threshold shifts after noise exposure in the bottlenose dolphin (Tursiops truncatus) measured using evoked auditory potentials," Marine Mammal Sci. **20**, 673–687.

National Research Council (NRC) (**1994**). *Low-Frequency Sound and Marine Mammals: Current Knowledge and Research Needs* (National Academy Press, Washington, D.C.).

National Research Council (NRC) (**2000**). *Marine Mammals and Low-Frequency Sound: Progress Since 1994* (National Academy Press, Washington, D.C.).

Pickles, J. O. (**1988**). *An Introduction to the Physiology of Hearing* (Academic, San Diego).

Popov, V., and Supin, A. (**1990a**). "Electrophysiological Studies of Hearing in Some Cetaceans and a Manatee," in *Sensory Abilities in Cetaceans*, edited by J. A. a. R. A. K. Thomas (Plenum, New York), pp. 405–415.

Popov, V. V., Ladygina, T. F., and Supin, A. Y. (**1986**). "Evoked potentials of the auditory cortex of the porpoise, Phocoena phocoena," J. Comp. Physiol., A **158**, 705–711.

Popov, V. V., and Supin, A. Y. (**1987**). "Characteristics of hearing in the beluga Delphinapterus leucas," Dokl. Biol. Sci. **294**(5), 1255–1258.

Popov, V. V., and Supin, A. Y. (**1990b**). "Auditory brainstem responses in characterization of dolphin hearing," J. Comp. Physiol., A **166**, 385–393.

Popov, V. V., and Supin, A. Y. (**1990c**). "Electrophysiological investigation of hearing in fresh water dolphin Inia Geoffrensis," Dokl. Biol. Sci. **313**(1), 238–241.

Popov, V. V., Supin, A. Y., Wang, D., Wank, K., Xiao, J., and Li, S. (**2005**). "Evoked-potential audiogram of the Yangtze finless porpoise Neophocaena phocaenoides asiaeorientalis (L)," J. Acoust. Soc. Am. **117**, 2728–2731.

Rance, G., Rickards, F. W., Cohen, L. T., Burton, M. J., and Clark, G. M. (**1993**). "Steady state evoked potentials: a new tool for the accurate assessment of hearing in cochlear implant candidates," Adv. Oto-Rhino-Laryngol. **48**, 44–48.

Rance, G., Rickards, F. W., Cohen, L. T., De Vidi, S., and Clark, G. M. (**1995**). "The automated prediction of hearing thresholds in sleeping subjects using auditory steady-state evoked potentials," Ear Hear. **16**, 499–507.

Ridgway, S. H., Bullock, T. H., Carder, D. A., Seeley, R. L., Woods, D., and Galambos, R. (**1981**). "Auditory brainstem response in dolphins," Neurobiology **78**, 1943–1947.

Schlundt, C. E., Finneran, J. J., Carder, D. A., and Ridgway, S. H. (**2000**). "Temporary shift in masked hearing thresholds of bottlenose dolphins, Tursiops truncatus, and white whales, Delphinapterus leucas, after expo-

sure to intense tones," J. Acoust. Soc. Am. **107**, 3496–3508.
Schusterman, R. J., and Moore, P. W. B. (**1978**). "Underwater audiogram of the northern fur seal Callorhinus ursinus," J. Acoust. Soc. Am. **64**, S87.
Supin, A. Y., 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. Y., Popov, V. V., and Mass, A.M, (**2001**). "*The sensory physiology of aquatic mammals*," (Kluwer Academic Publishers, Norwell, MA).
Supin, A. Y., and Popov, V. V. (**2000**). "Frequency-modulation sensitivity in bottlenose dolphins, Tursiops truncatus: evoked-potential study," Aquat. Mamm. **26**, 83–94.
Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (**1999**). "Killer whale (Orcinus orca) hearing: Auditory brainstem response and behavioral audiograms," J. Acoust. Soc. Am. **106**, 1134–1141.
Vander Werff, K. R., and Brown, C. J. (**2005**). "Effect of audiometric configuration on threshold and suprathreshold auditory steady-state responses," Ear Hear. **26**, 310–326.
Wolski, L. F., Anderson, R. C., Bowles, A. E., and Yochem, P. K. (**2003**). "Measuring hearing in the harbor seal (Phoca vitulina): Comparison of behavioral and auditory brainstem response techniques," J. Acoust. Soc. Am. **113**, 629–637.
Yuen, M. M. L., Nachtigall, P. E., Breese, M., and Supin, A. Y. (**2005**). "Behavioral and auditory evoked potential audiograms of a false killer whale (Pseudorca crassidens)," J. Acoust. Soc. Am. **118**, 2688–2695.

PROGRAM OF

The 151st Meeting of the Acoustical Society of America • Providence, Rhode Island • 5–9 June 2006
Rhode Island Convention Center

NOTE: All Journal articles and Letters to the Editor are peer reviewed before publication. Program abstracts, however, are not reviewed before publication, since we are prohibited by time and schedule.

MONDAY MORNING, 5 JUNE 2006 ROOM 555AB, 8:30 TO 11:30 A.M.

Session 1aAA

Architectural Acoustics: Current Topics in Architectural Acoustics

Lily M. Wang Chair
Univ. of Nebraska—Lincoln, Architectural Engineering, Omaha, NE 68182

Contributed Papers

8:30

1aAA1. An online ear training program designed for students, acousticians, and architects. Nozomi Kamiya, Ning Xiang (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180, nkamiya@eilarassociates.com), and Mendel Kleiner (Chalmers Univ. of Technol., SE-41296 Gothenburg, Sweden)

As part of a continuing research initiative at the Program in Architectural Acoustics at Rensselaer Polytechnic Institute, an online ear training program has been designed and implemented for scholars active in architectural acoustics, as well as professional acousticians and architects. The project aims to provide users the opportunity to learn to evaluate selected acoustical parameters without the employment of measurement equipment. The author believes that this training program will be indispensable for those who seek careers in fields related to acoustics for such skills are currently considered to be obtainable through professional experience in the field alone. The program was created using Macromedia Flash animations to develop an interactive training environment. Features such as user identification and an animated user interface create a personalized and entertaining experience while a score tracking system and high score table encourage friendly competition amongst users. As a result, users are compelled to return for successive training sessions which promote long-term growth and learning. Basic requirements for participation include an installed version of the Flash plug-in, an internet-enabled computer, and a pair of headphones.

8:45

1aAA2. Speech intelligibility of lecture halls. Lauren Ronsse (Univ. of Kansas, 1312 Louisiana St., Lawrence, KS 66044, lronsse@ku.edu)

For this project, different large lecture halls were analyzed using various speech intelligibility measurement techniques. The goal was to determine how well these measurements correlate with each other and convey speech intelligibility. Objective speech intelligibility and other related acoustical measurements were made throughout each hall at varying distances from the speaker. These measurements were repeated in each hall in a standardized manner, allowing the results from the different halls to be compared effectively. These objective measurements were related to the subjective analysis of the hall to determine which measurements indicate an appropriate speech intelligibility level for that space. Background mechanical system noise was introduced into some of the halls to determine the effect on speech intelligibility. Some of the halls tested were modeled in an architectural acoustics computer program to determine how well the actual measurements correlate with the predicted measurements. Measurement techniques and results of the study will be presented.

9:00

1aAA3. A shape variable acoustic panel system for multipurpose rooms. Ken Andria (Acoust. Dimensions, 145 Hugenot St., Ste. 406, New Rochelle, NY 10801, andriak@gmail.com) and Mendel Kleiner (Chalmers Univ. of Technol., SE-41296 Gothenburg, Sweden)

This work proposes an acoustically effective alternative to current passive acoustic solutions. Typical systems involve moving panels to one of either reflective, absorptive, or diffusive surface states. By varying the

shape of an array of ceiling reflectors, acousticians can achieve more precise control of the room response to better serve a wider range of acoustic scenarios. The results of scale model testing show how a single system can direct, absorb, or diffuse sound energy so that the room will be appropriate to four distinct programmatic events: lectures, gathering, simultaneous sources, and flat effect. By including microphones and a digital signal processing system the panel system may be designed to adaptively change its behavior to suit that of the current room use.

9:15

1aAA4. Acoustical design of St. Margaret's Chapel. Akira Ono and Minoru Nagata (Nagata Acoust. Inc., Hongo-Segawa Bldg.3F, 2-35-Hongo Bunko-Ku Tokyo113-0033)

St. Margaret's School is an all-girls high school located in the quiet, upscale Suginami Ward of Tokyo. The school's St. Margaret's Chapel was built some 70 years ago and included a pipe organ. The school decided to replace it with a new pipe organ and commissioned Taylor & Boodie, a U.S. organ builder. They requested adjustments to the chapel's room acoustics so that the new organ would sound its best. Before the project began, the chapel's reverberation time measured 1.4 s at 63 Hz and 1.8 s at 500 Hz and 1 kHz. That is, the reverberation time was shorter for low register sounds. The immediate challenge we faced on this project was the school's requirement that the chapel not be altered at all visually. St. Margaret's Chapel has historical value that the school wished to preserve. The framework of the chapel building and its floor are entirely made of wood. Moreover, the area around the altar is wood paneling. Upon investigation, we confirmed that the vibrations of these wood elements were absorbing the low register sounds. We determined that strengthening the walls and floor of the altar area would be an effective strategy to improve the chapel's acoustics.

9:30

1aAA5. Sound absorption in concert halls: Rectangular versus other shapes; Sabine versus Eyring equations. Leo L. Beranek (975 Memorial Dr., Ste. 804, Cambridge, MA 02138, beranekleo@ieee.org)

The accurate calculation of reverberation time in a concert hall in advance of construction is vital because it relates directly to the volume of the hall, a critical architectural parameter. The calculation can be by either the Sabine or Eyring equations which, broadly speaking, requires two quantities, the absorption coefficients of the seated audience plus orchestra and those of the nonaudience areas, i.e., side walls, ceiling, and the like. The two formulas require different sizes of absorption coefficients. The Sabine and Eyring equations are analyzed here, including their relative accuracies. A means for deriving the Eyring absorption coefficients when the Sabine coefficients are known is presented. Based on published acoustical and architectural data for 20 selected concert halls, which are of two shapes, the audience and chair absorption coefficients are calculated. It is concluded that for the same audience area and a given midfrequency reverberation time the volume of a hall is directly proportional to the midfrequency audience absorption coefficient, which, in turn depends on the type of chair and the shape of the hall. New audience absorption coefficients for medium and heavily upholstered chairs and for the two types of hall are presented.

9:45

1aAA6. An efficient algorithm for Bayesian energy-decay analysis in architectural acoustics. Ning Xiang and Paul Calamia (Grad. Prog. in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

Sound energy-decay analysis, including experimental determination of reverberation times in single-space enclosures and decay-time estimation in acoustically coupled spaces, has relied primarily on Schroeder backward integration of room impulse responses. Recent work has demonstrated that model-based Bayesian approaches [Xiang *et al.*, J. Acoust. Soc. Am. **110**, 1415–1424 (2001); **113**, 2685–2697 (2003); **117**, 3705–3715 (2005)] can be very useful for such analysis in architectural acoustics. In order to remove any concern of a high computational load within the Bayesian framework, this paper describes a search algorithm for the model-based Bayesian approaches which leads to an efficient analysis method, not only for multiple decay rates in acoustically coupled spaces but also for classical reverberation time estimation. Implementation of the algorithm within the Bayesian framework is discussed, and its performance with respect to other existing algorithms is analyzed using experimentally measured room impulse responses.

10:00–10:15 Break

10:15

1aAA7. Experimental investigations of acoustically coupled spaces. Zuhre Su, Ning Xiang (Grad. Prog. in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, New York 12180), and Jason E. Summers (Acoust. Div. Naval Res. Lab., Washington, DC 20024)

Recently, architectural acousticians have been increasingly interested in halls that incorporate coupled-volume systems because of their potential for creating desirable effects in meeting conflicting requirements between clarity and reverberance. Previous works on coupled spaces include studies on theoretical formulation of energy exchange between coupled volumes and applications of Bayesian inference for estimating decay parameters in such spaces. Frequency dependence of sound-energy decay is one of the essential aspects of acoustically coupled spaces that has not yet been extensively investigated. In order to understand the behavior of sound-energy decay over different frequency ranges, experimental studies are carried out. Room impulse responses are collected through real-size and scale-model measurements. Schroeder integrations followed by Bayesian decay analysis reveal location sensitivities over different frequency bands in coupled spaces. This paper discusses extensive analysis results from Bayesian analysis of experimentally measured room impulse responses.

10:30

1aAA8. Subjective response to absorption and aperture size variations in complex computer-modeled coupled volume systems. David T. Bradley and Lily M. Wang (Architectural Eng., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681, dbradley@mail.unomaha.edu)

Computer models of an imagined but realistic coupled volume concert hall have been made to obtain room impulse responses for several different configurations of the space. A coupled volume concert hall is typified by a large, absorbent main volume linked to a smaller, reverberant auxiliary space through a series of sonically transparent openings. Absorption and aperture size have been modified in this study to create each configuration of the virtual hall. Auralizations of the configurations have been used in paired-comparison tests to determine listener preference for the various sound fields, including some that exponentially decay and others that show a double-sloped decay. The results from this subjective testing were statistically analyzed using Thurstone Case V and multidimensional scaling methods. These results will be presented and contrasted with objective parameters extracted from the impulse responses generated by the computer modeled-configurations of the hall.

10:45

1aAA9. Measurements and verification of two mosques in Saudi Arabia. Wasim A. Orfali and Wolfgang Ahnert (Inst. of Techn. Acoust., Tech. Univ. Berlin, Einsteinufer 25, 10587 Berlin, Germany, orfaliwasim@hotmail.com)

Two prestigious mosques, Holy Haram in Makkah and Prophets Mosque in Medinah, have been selected to investigate their acoustical properties. With portable measurement setup basic parameters like reverberation time, intelligibility numbers, sound level distribution, and others have been measured. This was done with and without a sound system installed at the site. In postprocessing the results have been compared with recommendations developed by us for different mosque sizes. Also, proposals are given for further improvements in the existing sound quality. Finally, suggestions for future newly constructed mosques are derived.

11:00

1aAA10. A sound encounter or deaf ignorance: Coherence of new buildings to old mosques. Quinsan Ciao (Dept. of Architecture, Kuwait Univ., P.O. Box 5969, Safat 13060, Kuwait)

This is to raise the issue of a common phenomenon occuring within the booming development of many Islamic cities where a building site is nested within the common cluster of old mosques. The location and exisitance of mosques are fixed as God-given constraints and so is the voice of calling for prayers. The sound of worshiping infiltrated in all urban interior/exterior spaces. A brief study shows the split views in an Arab city where freedom of expression is progressing. Should the new buildings embrace the spiritual deeds and be the powerful bearer of the Holy Tune? Should we try to prevent the noise from interfering with the modern day function? Are there any design guidelines? A quick overview shows that the majority of new buildings recently built in Kuwait City neglected the subject, particularly the large projects commissioned to western design firms whose consultants may have less awareness of the issues related to the small-scale religious entities. This creates not only challenges to their proposal of an ideal form, but also the need for a unique answer to the acoustics regarding the sound of worship. A project involving students' works was carried out to explore the possibilities of tackling the problem.

11:15

1aAA11. Perceptual acoustics in the Mapparium. William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., 4230 BPS Bldg., E. Lansing, MI, 48824, and Dept. of Biomed. Engr., Boston Univ. Boston, MA 02215), H. Steven Colburn, and Gerald Kidd, Jr. (Boston Univ., Boston, MA, 02215)

The Mapparium is a hollow, stained-glass globe, 30 ft. in diameter, at the Mary Baker Eddy Library in Boston. It allows visitors to see the continents of the world from the perspective of an observer inside the Earth. The ears of a visitor walking across the glass bridge traverse a diameter in the equatorial plane. The acoustical properties of this space are of secondary importance, but they are so remarkable that they amaze everyone who experiences them. Extensive listening and recording experiments within the Mapparium (prior to the renovations of 2002) revealed some of the acoustical effects in detail, with source and receiver points near the bridge diameter. Among the effects are the following: anomalous broadband point-to-point amplitude transfer functions with maxima (whispering gallery) and spatially compact minima; localization illusions wherein a source on the listener's left is heard on the right, or high- and low-frequency parts of a sound are localized differently, or sound images change in elevation as the listener walks; spatially compact regions of flutter echo; and spatially unstable reflections of one's own voice. [Work facilitated by the Mary Baker Eddy Library for the Betterment of Humanity.]

MONDAY MORNING, 5 JUNE 2006 ROOM 550AB, 8:55 TO 11:35 A.M.

Session 1aAB

Animal Bioacoustics: Acoustic Interactions in Animal Groups I

Andrea M. Simmons, Chair
Brown Univ., Dept. of Psychology and Neuroscience, Providence, RI 02912

Chair's Introduction—8:55

Invited Papers

9:00

1aAB1. Interference risk and dynamic shifts in calling in the gray treefrog. Joshua Schwartz (Dept. of Biological and Health Sci., Pace Univ., 861 Bedford Rd., Pleasantville, NY 10570, jschwartz2@pace.edu)

Male gray treefrogs call to attract females under challenging acoustic conditions. With increases in chorus density, background noise levels and the chance that male calls will be overlapped increase. Call overlap may jeopardize the probability of mating because females prefer advertisement calls in which the internal pulse structure of calls is not obscured to those that are overlapped by other calls. With increases in the level of acoustic stimulation, males lengthen the duration of their calls but simultaneously reduce call rate such that pulse effort changes little. The interference risk hypothesis was tested, which proposes that males change their calling in a fashion that insures that, on average, there will be a sufficient number of pulses and interpulse intervals clear of acoustic interference per call to attract a female. However, a series of female choice experiments employing call alternatives of different durations, rates, intensity levels and degrees of call overlap refuted this hypothesis. Results leave open the possibilities that the dynamic shift in the two components of pulse effort are related to the problem of call detection in a noisy environment or are responses by males to the perceived threat of competition for females. [Work supported by NSF.]

9:20

1aAB2. Signal timing by the gray treefrog, *Hyla chrysoscelis*: Responses to conspecific and heterospecific advertisement calls. Vincent Marshall (Beckman Inst., Univ. of Illinois, 405 North Mathews Ave., Urbana, IL 61801)

Breeding choruses often consist of multiple species sharing the same communication channel. Whether and how a signaler adjusts the timing of its signals, an often important predictor of receiver response, may therefore depend on the species identity of its neighbor. How males of the gray treefrog, *Hyla chrysoscelis*, adjusted advertisement call timing towards the calls of conspecifics and those of the closely related, syntopic species, *H. versicolor*, was examined. During interactions between calling males, calls were timed such that there was little signal overlap regardless of whether the neighbor was conspecific or heterospecific. In response to playbacks of conspecific and heterospecific advertisement calls, males generally avoided producing calls during either stimulus type,

and instead produced them during the silent gaps between stimuli. There were, however, significant differences in the latencies of response to conspecific and heterospecific signals. Males produced calls with shorter latencies after the heterospecific calls, which resulted in their often rapid placement within the silent intervals between stimuli. In contrast, calls produced after conspecific stimuli almost always had a distinct gap after the stimulus ended. These results are discussed in light of hypotheses for the function of adjustments in the timing of acoustic sexual signals.

9:40

1aAB3. Temporal organization of bullfrog choruses. Andrea M. Simmons, James A. Simmons (Dept. Psych. & Dept. Neurosci., Brown Univ., Providence, RI 02912, Andrea_Simmons@brown.edu), and S. D. Deligeorges (Boston Univ., Boston, MA 02215)

Choruses of advertisement calls of male bullfrogs (*Rana catesbeiana*) were recorded at two different ponds with an array consisting of two 4-microphone sensors to locate and identify calling individuals. The array determined the bearing of frogs from two positions located 10 m apart along the edge of the pond. Frogs were segregated and classified as individuals from the cluster of bearing intersections provided by the two sensor units for each note making up the advertisement calls. At one pond, 6 frogs were identified by bearing intersections, and 2 additional frogs, picked up by only one sensor unit, were separately identified from the pitch of their calls. At the other pond, 9 frogs were localized and 3 additional frogs were identified. Some frogs called independently, with coincident calling between pairs occurring at or near chance levels. These frogs usually were located remote from one another in the pond. At both ponds, a group of 2–3 frogs located in relative proximity called in partially overlapping temporal sequences that recurred throughout recording sessions. Members of the group may have to recognize each other acoustically in order to maintain the integrity of the timing sequence across multiple repetitions. [Work supported by NIH and DARPA.]

10:00–10:20 Break

10:20

1aAB4. Cues used by female barking treefrogs (*Hyla gratiosa*) to assess distances to potential mates. Christopher G. Murphy (Dept. of Biol., James Madison Univ., MSC 7801, Harrisonburg, VA 22807)

In chorusing insects and amphibians, females have the potential to choose from among a number of males located at various distances from the choosing female. In barking treefrogs, the distance to potential mates influences female choice, with females choosing a more attractive call over an unattractive call until the distance to the source of the more attractive call, relative to the source of the unattractive call, becomes too great. Two- and eight-speaker playback experiments were used to determine which cues females use to assess the distances to potential mates. These playback experiments ruled out three possible cues: degradation of signal quality, amplitude at the location of the female, and the sound gradient (how rapidly call amplitude increases as a sound source is approached). Additional playbacks showed that females consistently preferred signals with the highest source amplitude, even when the amplitude of that signal at the location of the female was lower than that of another signal during a substantial proportion of the approach by the female. These results indicate the source amplitude is an important cue in mate choice and suggest that females use a complex mechanism (e.g., triangulation) to judge distance to potential mates.

10:40

1aAB5. The causes and evolutionary consequences of a conflict between species and mate-quality recognition. Karin S. Pfennig (Dept. of Biol., CB#3280, Univ. of North Carolina, Chapel Hill, NC 27599-3280, kpfennig@email.unc.edu)

Mate choice typically involves the identification of conspecifics (species recognition) and the identification of mates that provide females with enhanced fitness (mate-quality recognition). When high-quality conspecifics possess advertisement signals that resemble those of heterospecifics, however, females may be unable to engage in both species and mate-quality recognition simultaneously. This conflict between species and mate-quality recognition may require that females trade off the benefits of mating with high-quality males with the risk of heterospecific matings. In spadefoot toads, females face a conflict between species and mate-quality recognition because the advertisement calls of high-quality conspecifics resemble those of heterospecifics. Because of this conflict, females compromise on male quality to ensure conspecific matings. Moreover, this trade-off has seemingly caused spadefoot toads to evolve smaller body size, reduced fecundity, and different patterns of sexual dimorphism in populations where heterospecifics are present (sympatry) versus populations where heterospecifics are absent (allopatry). Thus, a conflict between species and mate-quality recognition can have profound impacts on how females use male advertisement signals to choose mates, sexual selection, and morphological evolution in different populations, and even the fitness of individuals in sympatry relative to those in allopatry.

11:00

1aAB6. The *cocktail party problem*: What is it? How can it be solved? And why should students of bioacoustics study it? Mark A. Bee (Dept. of Ecology, Evolution, and Behavior, Univ. of Minnesota, 1987 Upper Buford Circle, Rm. 100, St. Paul, MN 55108)

Auditory scene analysis refers to mechanisms that parse a composite acoustic waveform into perceptually distinct auditory *objects* or *streams* that correspond to distinct sound sources in the environment. A familiar and well-studied example of a difficult auditory scene analysis task is the *cocktail party problem*, which refers to the difficulty we have perceiving speech in noisy social environments comprised of multiple people talking simultaneously. Students of human hearing and computer engineering study auditory scene analysis, in part, because people with hearing impairments and computer algorithms for speech recognition do a poor job of solving the cocktail party problem compared to people with normal hearing. Of course, the cocktail party problem is not unique to humans. The animal behavior literature provides numerous examples of cocktail-party-like social environments for animal acoustic commu-

nication (e.g., choruses, roosts, colonies, creches). Compared to previous work on humans, however, students of animal bioacoustics have devoted little attention to how animals that signal in groups or networks solve their own cocktail party problems. This talk will briefly review some issues in auditory scene analysis relevant to the study of animal acoustic communication and present the case for why these issues deserve further study by students of bioacoustics.

Contributed Paper

11:20

1aAB7. Observation and modeling of a spatiotemporal pattern in fish chorusing. Gerald L. D'Spain, Heidi H. Batchelor, and Catherine L. Berchok (Marine Physical Lab., Scripps Inst. of Oceanogr., 291 Rosecrans St., San Diego, CA 92106)

A numerical model was developed to predict the unusual spatiotemporal characteristics of sciaenid chorusing recorded offshore of the Southern California coast. The model, based on the physics of excitable media, is composed of a random distribution of units (fish) on a 2D grid that inhabits one of three states: resting, excitable, and active. Temporal evolution of a unit's state is determined by transition laws. The temporal character of chorusing (increases in ocean sound levels by factors of 2 to 5 for 10–20 s followed by 15–20 s periods of lower levels) can be predicted using a Poisson waiting time distribution, once the proper parameters are selected. For some parameters, the predicted chorusing levels evolve into a continuous din rather than displaying a cycling nature. In one experiment, the spatial evolution of chorusing also has a periodic nature; each chorus begins off the Mexican coast and the chorusing region propagates 25 km upcoast until the fish just south of San Diego harbor reach a chorus peak 16–20 s later. Predicting this observation for arbitrary initial conditions requires coupling between units, achieved by incorporating the received sound level into the probability of transitioning from excitable to active states. [Work supported by ONR.]

MONDAY MORNING, 5 JUNE 2006 BALLROOM A, 8:00 A.M. TO 12:00 NOON

Session 1aBB

Biomedical Ultrasound/Bioresponse to Vibration: Brown Tadpoles and Red Herrings I: Boiling Inertial Cavitation and Nonlinearity in High Intensity Focused Ultrasound Lesion Formation

R. Glynn Holt, Chair

Boston Univ., Dept. of Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Invited Papers

8:00

1aBB1. A chair's introduction to brown tadpoles: What's all the fuss about? R. Glynn Holt (Dept of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, rgholt@bu.edu)

A brief background of the work in understanding irregular HIFU thermal lesion formation will be presented to motivate the current session. Several basic arguments for asymmetric lesion formation will be framed, highlighted by examples of *in vivo* occurrence. Several fundamental questions will be posed, with the hope that such questions will be addressed, clarified, and answered by the speakers and panel members (as well as the audience) by the end of the session. Along with conceptual discussions of relevant mechanisms, recommendations for usage of terms to describe and distinguish these mechanisms will be encouraged.

8:15

1aBB2. Microbubble cavitation, boiling, and nonlinear acoustic propagation in high-intensity focused ultrasound thermal therapy. P. J. Kaczkowski, M. R. Bailey, L. A. Crum (Ctr. for Industrial and Medical Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, peter@apl.washington.edu), V. A. Khokhlova (Moscow State Univ., Moscow 119992, Russia), and A. Anand (Philips Research USA, Briarcliff Manor, NY 10510)

The investigation of high-intensity focused ultrasound (HIFU) as a tool for noninvasive thermally ablative therapy has required deeper understanding of the relative roles of nonlinear mechanisms involved in heat deposition. Attempts at quantifying the dose response to particular exposure conditions *in vitro* are complicated by the interplay of several mechanisms. These include microbubble cavitation, nonlinear acoustic propagation and attenuation, dependence of tissue parameters on temperature and temperature history, and formation and evolution of vapor bubbles due to boiling. One immediately evident consequence of such effects is distortion of coagulative lesion shape and size, colloquially evolving from cigars to tadpoles. Developing a quantitative understanding of the relative roles of relevant nonlinear mechanisms is not straightforward, yet is desirable for design of algorithms for therapy planning and real-time monitoring using ultrasound. A historical perspective of research toward this end will be presented along with a recommendation for suitable terminology for the various physical acoustic regimes encountered in HIFU therapy. [Work supported by Army MRMC, NIH DK43881, NSBRI SMS00402, and RFBR.]

8:45

1aBB3. The relevance of bubbles in clinical high-intensity focused ultrasound treatments. Gail ter Haar (Phys. Dept., Inst. of Cancer Res., Royal Marsden Hospital, Sutton, Surrey, SM2 5PT, UK)

Until recently there has been a dichotomy of opinion as to the utility of bubbles in clinical HIFU treatments. On the one hand uncontrolled acoustic cavitation can lead to lesions of unpredictable shape, size, and position, while on the other, lesions containing bubbles are visible on conventional B-mode ultrasound images, and clinically are generally regarded as being indicative of tissue destruction. In some HIFU treatments of cancer, therefore, the ultrasound exposure levels are increased until bright ups are seen on the real time ultrasound scan of the target volume. It is now known that, in the presence of acoustic cavitation, the rate of heating is considerably enhanced, and investigations are under way to establish whether controlled cavitation induction can be used to speed up clinical HIFU treatments. When temperatures in excess of 100 °C are achieved in tissue, tissue water boiling will occur. This too will result in enhanced scatter in the ultrasound image. Boiling will also result in tissue destruction. The clinical implications of bubble activity will be reviewed in this paper.

9:15

1aBB4. Bubble nucleation, growth and dissolution in soft tissue. Charles C. Church (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Vapor bubbles of sufficient size to serve as cavitation nuclei may form spontaneously in tissue in regions of very low interfacial tension. In the absence of an acoustic wave or other mechanical stress, such nuclei will quickly dissolve and disappear from the medium. Under the influence of an acoustic wave, however, these microbubbles may grow to many times their initial size and then collapse violently, a process known as inertial cavitation. The *in vivo* energetics of the nucleation process are treated by assuming tissue to be a homogeneous fluid, while the initial dynamics of nucleated bubbles are modeled by treating tissue as an isotropic, viscoelastic solid. It is shown that the rate of nucleation-cavitation events is highly sensitive to the instantaneous tissue temperature and moderately sensitive to the rarefactional pressure. The effect of tissue elasticity is to increase the threshold pressure for inertial cavitation and decrease the rate of rectified diffusion relative to bubbles in water. Finally, the increased rigidity of denatured soft tissue enhances the persistence of gas/vapor bubbles following the acoustic pulse.

Contributed Papers

9:45

1aBB5. A model for irregular lesion formation including cavitation bubbles and boiling. Tianming Wu, Caleb H. Farny, Ronald A. Roy, and R Glynn Holt (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Nonlinear propagation and scattering of incident sound due to the presence of gas/vapor bodies and temperature-dependent sound speed and acoustic absorption have been found to contribute to irregular lesion formation during high-intensity focused ultrasound therapy. The appearance and importance of these mechanisms are dependent on the intensity of insonating sound and the medium properties. In the regime where the focal pressure is slightly above the cavitation threshold and the medium possesses sufficient nuclei, photographic observations and model simulations for tissue phantoms reveal that tadpole lesions start to develop prior to large-scale boiling; however, enhanced heating from inertial cavitation greatly accelerates the progression to boiling. The relative time scales for cavitation-triggered boiling and bubble dissolution are studied in order to yield insight on how to control pulse-mode HIFU insonation in order to either avoid bubble behavior outright or utilize cavitation and boiling to promote therapy. [Work supported by the Dept. of the Army (Award No. DAMD17-02-2-0014) and the Center for Subsurface Sensing and Imaging Systems (NSF ERC Award No. EEC-9986821).]

10:00–10:15 Break

10:15

1aBB6. The dependence of temperature rise and cavitation dose on acoustic pressure during high-intensity focused ultrasound (HIFU) application. Caleb H. Farny, Tianming Wu, R. Glynn Holt, and Ronald A. Roy (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, cfarny@bu.edu)

Given sufficient nuclei, inertial cavitation can greatly enhance heating rates during HIFU application. In the absence of bubble shielding effects, one might assume that increasing the driving pressure amplitude will serve to enhance both the temperature elevation and the amount of cavitation activity. Here we report on measurements in agar-graphite phantoms that show that during HIFU application, cavitation dose and temperature rise do increase with increasing pressure but only to a point where there is a measurable decrease in the 2.5–3-MPa range followed by further increase with higher pressures. This behavior can be predicted using single-bubble dynamics with allowed equilibrium bubble sizes that are consistent with waterlike viscosity values for an effective Newtonian fluid host [see Yang *et al.*, J. Acoust. Soc. Am. **116**, 3423–3431 (2004) for details]. These results can serve to help guide an optimal pressure as well as develop a further understanding of the bubble parameters and energy dissipation mechanisms involved in HIFU. [Work supported by the Dept. of the Army (Award No. DAMD17-02-2-0014) and the Center for Subsurface Sensing and Imaging Systems (NSF ERC Award No. EEC-9986821).]

10:30

1aBB7. Interpreting passive cavitation detection signals during high-intensity focused ultrasound (HIFU). Jamie R. T. Collin and Constantin C. Coussios (Dept. of Eng. Sci., Univ. of Oxford, 17 Parks Rd., Oxford OX1 3PJ, UK)

Inertial cavitation is known to be associated with significantly enhanced rates of heating during cancer treatment by high-intensity focused ultrasound (HIFU). However, during prolonged exposure, the cavitation region gradually shifts towards the HIFU transducer, causing shielding of the original focus. It is hoped that controlling the HIFU field to constrain cavitation to the focal region will lead to sustained enhanced focal heating, thus reducing treatment times and improving treatment resolution. Monitoring of cavitation can be achieved during insonation by using a focussed passive cavitation detector (PCD) to listen for broadband noise emissions resulting from microbubble collapse. The ability to effectively control the cavitation field relies on accurate interpretation of the PCD data. During HIFU exposure of a tissue-mimicking material, it was noted that the frequency content of the PCD signal shifted downwards with increasing phantom propagation path length to the PCD, as a result of depth-dependent attenuation. This has significant implications for the effective sensing volume of the PCD. In addition, it was noted that separate analysis of the peak and rms time-domain PCD values provides an incomplete description of inertial cavitation activity, but that combining these two parameters enables a better understanding of bubble cloud dynamics.

10:45

1aBB8. An amplitude-modulated harmonic motion imaging method for mapping tissue ablation. Caroline Maleke, Mathieu Pernot, and Elisa Konofagou (Dept. of Biomed. Eng., Columbia Univ., 1210 Amsterdam Ave., New York, NY 10027)

Previously, a single-element focused transducer method inducing amplitude-modulated (AM) radiation force deep within the tissue was introduced. This technique offers real-time capability for monitoring the change in the tissue mechanical properties during focused ultrasound (FUS) therapy. In this paper, we extend the application of this technique for 2D mapping of the FUS lesion. The harmonic radiation force was generated by a 4.68-MHz focused transducer modulated at 50–100 Hz. Radio frequency (rf) data were acquired at 7.5 MHz with a PRF of 6.5 kHz, and displacements were estimated using a 1D cross-correlation algorithm on successive rf signals. A bandpass analog filter with $fc1 = 5$ MHz and $fc2 = 8$ MHz was employed to remove the spectrum of the focused beam. The real-time monitoring method was adjusted to induce 2-s sonication time using a continuous wave, followed by 100-ms AM signal to image tissue displacements. The experiments were performed on *ex vivo* porcine liver. The tissue was raster-scanned before and after lesion formation. Results indicated that displacement amplitudes decreased after lesion formation due to the associated elasticity change, and thus 1-cm-diameter lesions were accurately mapped. Future investigations of this technique will include temperature and coagulation mapping *in vivo*.

11:00

1aBB9. Development of a cell-embedded tissue-mimicking ultrasound phantom. Yunbo Liu and Pei Zhong (Dept. of Mech. Eng., Duke Univ., Durham, NC, 27708, pzhong@duke.edu)

An optically transparent cell-embedded tissue-mimicking phantom was developed to study high-intensity focused ultrasound (HIUF)-induced gene activation *in vitro*. The phantom was constructed by using agarose hydrogel embedded with dispersed rat mammary carcinoma (Mac) cells that were stably transfected with GFP plasmid under the control of hsp70B promoter. Thermally sensitive bovine serum albumin (BSA) proteins were also incorporated in the phantom so that both lesion formation and gene activation could be quantified simultaneously following HIFU exposure. Both acoustic (sound speed, attenuation, acoustic impedance, B/A ratio) and thermal properties (thermal conductivity and thermal diffusivity) of the phantom were characterized in a temperature-dependent manner. At a peak temperature between 70 °C and 90 °C and a 10-s exposure using a 3.3-MHz HIFU transducer, the inducible GFP gene expression patterns in the phantom were analyzed and compared with the corresponding 3-D temperature distribution quantified by thermocouple measurements. Overall, the cell-embedded tissue-mimicking phantom was found to be a versatile tool for investigating HIFU-induced gene activation and it may also be used for general ultrasound bio-effect studies. [Work supported by NIH.]

11:15

1aBB10. Effects of short-pulsed focused ultrasound: Histological study on the rabbit brain *in vivo*. Natalia Vykhodtseva, Nathan McDannold, and Kullervo Hynynen (Dept. of Radiol., Brigham and Women's Hospital, and Harvard Med. School, 75 Francis St., Boston, MA 02115)

Histological effects of short-pulsed focused ultrasound at pressure amplitudes beyond the cavitation threshold were studied. One hundred thirty-one locations in the brains of 27 rabbits were sonicated using a 16-sector focused transducer (1.63 MHz; $D/F = 10/8$ cm; burst lengths $= 1–100$ ms; PRF $= 1–100$ Hz, total exposure $= 10$ s). A range of effects was observed: (1) Affected blood vessels without any damage to the brain parenchyma. (2) Local lesions with homogenized core, which consisted of a pudding-like mass arranged in a helically coiled, layered fashion in some lesions. In the large lesions, core was completely liquefied. (3) Hemorrhagic and torn regions with fragmented and liquefied tissues, characteristic of inertial cavitation effects. (4) Circular-shaped lesions with a concentric annular pattern: a small central core was surrounded by three or more morphologically different circular layers. (5) Thermal lesions, which had the features of thermal lesions produced by lower-power sonications. These findings appeared to be associated with the nonlinear ultrasound absorption and cavitation-related effects, which could be exploited for new therapies. [Work sponsored by: NIH Grant Nos. R01CA076550 and R25CA089017.]

11:30

1aBB11. Focused ultrasound-induced opening of the blood-brain barrier in mice. James J. Choi, Mathieu Pernot, Scott A. Small, and Elisa E. Konofagou (Columbia Univ., 351 Eng. Terrace, Mail Code 8904, 1210 Amsterdam Ave., New York, NY 10027, ek2191@columbia.edu)

The feasibility of blood-brain barrier (BBB) opening in the hippocampus of mice using transcranial focused ultrasound (FUS) was investigated on simulations and *in vivo* experiments. A micro-CT scan of an excised mouse skull was performed (resolution: 0.01 mm) and mappings of porosity, speed of sound, density, and attenuation were calculated. A 1.5-MHz wave propagating through the skull mappings was simulated, revealing a well-formed beam, minimal attenuation, and minute focus displacement. These results were verified with needle hydrophone measurements through five excised mouse skulls. In experiments *in vivo*, three mice were intravenously injected with a bolus of 0.01-mL ultrasound contrast agents (Optison) and the brains were sonicated transcranially. Ultrasound sonications (frequency: 1.5 MHz; burst length: 20 ms; duty cycle: 20%; acoustic pressure range: 2.0 to 2.7 MPa) were applied five times for 30 s per shot with a 30-s delay between shots. High-resolution (9.4 tesla), contrast-enhanced MR imaging revealed BBB opening and slow permeation of gadolinium from the posterior cerebral artery (10-min postinjection) through the surrounding tissue (35-min postinjection) and to the entire left hippocampus (95-min postinjection). These results demonstrate the feasibility of localized, noninvasive BBB opening in mice using transcranial FUS for potential precise drug delivery.

11:45

1aBB12. Focused ultrasound-induced transgene activation in skeletal muscles. Xiaochun Lu, Georgii Sankin (Dept. of Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708), John Madden, and Pei Zhong (Duke Univ., Durham, NC 27708)

Focused ultrasound (FUS)-induced gene activation in the skeletal muscles of a rat model was investigated. Plasmids that encode the luciferase gene under the control of hsp70B promoter were transfected into the skeletal muscles of the hind limbs of a rat model by electroporation. One day following the gene transfection, the left hind limb was treated with 3.3-MHz FUS by raster scan of a 9×9-mm^3 area while the right hind limb was used as control. The output of the FUS transducer was adjusted to produce a rapid *in vivo* temperature rise (< 4 s) to three different temperature plateaus in the range of 40–50 °C, 50–60 °C, and 60–70 °C for a treatment duration up to 30 s. One day following FUS treatment, gene expression *in vivo* was analyzed and tissue injury assessed by H&E staining. It was found that *in vivo* gene activation and tissue damage were highly correlated with thermal dosage. The highest gene activation was produced by the FUS at 50–60 °C with an exposure time of 20 s, which was found to be statistically equivalent to the gene activation produced by water-bath hyperthermia (42 °C for 35 min). At this thermal dose, FUS-induced tissue injury was sublethal and could recover in a week.

Session 1aSPa

Signal Processing in Acoustics: Signal Processing: Diverse Problems—Similar Solutions

Sean K. Lehman, Chair
Lawrence Livermore National Lab., 700 East Ave., Livermore, CA 94550

Chair's Introduction—7:45

Invited Papers

7:50

1aSPa1. Introduction to deconvolution. Peter A. Jansson (Jansson Associates, 19 Arthur Dr., Hockessin, DE 19707)

Deconvolution tasks will always lie at the frontier of knowledge in many fields, almost by definition, because measurement resolution often defines the state of our knowledge. Rising in the latter 20th century from near-disreputability, first to usefulness, then necessity in some disciplines, the varied techniques of deconvolution have assumed an increasingly important role in the scientist's tool kit. This talk will trace deconvolution's development with examples, including many firsts, drawn from spectroscopy, radio astronomy, cell biology, color science, and diverse other fields. Following a tutorial introduction, detail will be provided on modern super-resolving methods, and lesser-known topics such as selected-ordinate image (SORI) processing.

8:10

1aSPa2. Room-acoustic energy decay analysis and acoustic landmine detection within Bayesian framework. Ning Xiang and Tomislav Jasa (Grad. Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

Model-based signal processing has been applied in sound energy decay analysis in enclosures for estimation of reverberation times [N. Xiang, J. Acoust. Soc. Am. **98**, 2112–2121 (1995)] and estimation of diverse decay times to quantify sound energy decay in coupled spaces. Recent works using Bayesian approaches have demonstrated success in this application [Xiang *et al.*, J. Acoust. Soc. Am. **110**, 1415–1424 (2001); **113**, 2685–2697 (2003); **117**, 3705–3715 (2005)]. Similar approaches have also been applied in acoustic landmine detection [Xiang & Sabatier, J. Acoust. Soc. Am. **113**, 1333–1341 (2003)] for target detection and target localization. Respective models have been developed for each application in order to apply model-based Bayesian approaches in two different acoustical regimes. Using experimentally measured field data, this paper demonstrates an example of data analysis solutions developed for the room-acoustics application which can also be applied to another physical regime: acoustic landmine detection.

8:30

1aSPa3. Geometric methods for real-world imaging problems. Eric Miller, Gregory Boverman, Emre Guven, Mohamed Khamed ben hadj Miled (Dept. ECE, Northeastern Univ., Boston, MA 02115, elmiller@ece.neu.edu), Robin Cleveland (Boston Univ., Boston, MA 02215), and Basak Ulker Karbeyaz (Analogic Corp., Peabody, MA 19610)

The extraction of information concerning the internal structure of a medium given sparsely sampled, externally collected data is a ubiquitous problem arising in application areas ranging from the large-scale (geophysical prospecting and environmental remediation) to the small (biomedical imaging and nondestructive evaluation). The sensors used to collect data are equally diverse, encompassing the full spectrum of electromagnetic and mechanical modalities. In spite of the broad variety seen in this class of problems, many share underlying common challenges making the application of core processing methods of potential utility across a range of specific problems. In this talk, we discuss the application of geometric parametrizations to problems we have encountered in geophysical and biomedical imaging problems using optical, ultrasonic, and electromagnetic sensors. Estimating the geometry of a tumor, plume of toxic waste, or lesion created in the process of hyperthermia cancer treatment often reduces the ill-posed nature of the imaging problem and allows for the extraction of parameters directly linked to the application-specific information we seek to recover from the data. We shall discuss three approaches we have used to parametrize shape as well as our experience in applying these methods to real-world data.

8:50

1aSPa4. Applications of acoustic-wave-field downward continuation to geophysical and medical imaging. Lian-jie Huang, Michael Fehler (Los Alamos Natl. Lab., MS D443, Los Alamos, NM 87545, ljh@lanl.gov), Nebojsa Duric, and Peter Littrup (Karmanos Cancer Inst., Detroit, MI 48201)

We introduce acoustic imaging methods using wave-field downward continuation and describe their application to geophysical imaging and ultrasonic breast imaging. We have developed a number of novel migration imaging methods based on the acoustic-wave equation in heterogeneous media. They have been successfully applied for imaging complex structures beneath buried salt bodies in which the velocity is often more than twice that of the surrounding sediment within the Earth. The length scale in seismic imaging problems is on the order of 100 times of the seismic wavelength. Even though the frequency used in ultrasonic breast imaging is at a few megahertz, which is a million times higher than seismic frequency, its length scale is also on the order of 100 times of the ultrasonic wavelength in the breast. On the other hand, there are many differences between seismic imaging and ultrasonic breast

imaging. For example, the sizes of imaging objects are 10–100 wavelengths for seismic imaging but on the order of the wavelength or less for ultrasonic breast imaging. We have successfully applied the wave-equation-based migration methods to ultrasonic breast imaging. We will present imaging examples both in seismic imaging and ultrasonic breast imaging.

9:10

1aSPa5. High-resolution acoustic and optical imaging underwater. Hanumant Singh (Deep Submergence Lab., Appl. Ocean Phys. and Eng., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, hsingh@whoi.edu)

High-resolution sensing from underwater platforms continues to be limited by the mismatch between sensor resolution and available navigation information. In this talk we explore a similar set of solutions that works across the diverse applications and sensing modalities associated with multibeam sonar and optical camera systems. By utilizing results from the simultaneous localization and mapping (SLAM), computer vision and photogrammetric communities we examine the role of sensor consistency for overlapping sensor swaths to yield a far more accurate representation of the environment. We present our work in the context of real data acquired from a variety of autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs) in a variety of locations including shallow-water coral reef surveys with the Seabed AUV, mapping the RMS TITANIC at 4100 m from the Hercules ROV and multibeam mapping of the TAG hydrothermal mound on the Mid-Atlantic Ridge at 2800 m.

9:30

1aSPa6. Development of equivalent circuit analogies that can predict the loading effects of a multilayered structure on an electro-acoustic transducer. Karl Fisher (Lawrence Livermore Natl. Lab., 7000 E Ave., Livermore, CA 94551)

A technique is being developed based on a direct correlation between the electrical impedance of a standard electro-acoustic transducer and the mechanical loading it experiences when placed in contact with a layered elastic structure. Quantitative structural information can be determined empirically provided that there is an accurate calibration standard to compare the measured electrical impedance to. Normally this would be accomplished using a mechanical structure with known dimensions and construction. However, in some instances this may not be practical. The goal is to investigate simple and accurate calibration standards based on electrical circuit analogies to quantify resonant measurements on multilayered elastic structures. [This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.]

Contributed Paper

9:50

1aSPa7. Applying geophysical imaging algorithms to nondestructive evaluation problems. Sean K. Lehman, Karl A. Fisher (L-154, Lawrence Livermore Natl. Lab., 7000 East Ave., Livermore, CA 94550), Robert A. Clapp, and Brad Artman (Stanford Univ., Palo Alto, CA)

The Center for Subsurface Sensing & Imaging Systems (CenSSIS, www.censsis.neu.edu) has been advancing the concept of "Diverse Problems-Similar Solutions," whereby imaging, inversion, and tomography (IIT) algorithms developed for one field or physical regime are applied to another. Following this philosophy, algorithms developed for geophysical imaging have been successfully applied to a nondestructive evaluation (NDE) problem in a planar multilayer. The depths and acoustic frequencies of geophysical problems are on the orders of kilometers and sub-hertz, while those of NDE are millimeters and megahertz. Just the same, the propagation phenomenon is entirely wave-based in both cases and thus, algorithms developed for one can be applied to the other. [This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48. UCRL-ABS-216281.]

MONDAY MORNING, 5 JUNE 2006 ROOM 556AB, 10:10 A.M. TO 12:05 P.M.

Session 1aSPb

Signal Processing in Acoustics, Underwater Acoustics, and Acoustical Oceanography: Adjoint Modeling in Acoustics I

Jean-Pierre Hermand, Chair

Univ. Libre de Bruxelles, Dept. of Optics and Acoustics, av. F.D. Roosevelt 50, B-1050 Brussels, Belgium

Chair's Introduction—10:10

Invited Papers

10:15

1aSPb1. The method of adjoint field in inverse scattering of plane waves. D. N. Ghosh Roy (Ctr. For Adv. Imaging Res. (UCAIR), Univ. of Utah, 729 Arapeen Dr., Salt Lake City, UT 84108), S. J. Norton (Geophex, Ltd., Raleigh, NC 27603), and M. Schable (Univ. of Utah, Salt Lake City, UT 84108)

The notion of the adjoint operator/field is introduced, and its general importance in physical sciences briefly discussed. The applications of the adjoint in the solutions of inverse scattering problems for plane-wave excitation are overviewed for both obstacles and inhomogeneities. The method is then illustrated by applying it to a specific problem, namely the volume reconstruction of an

acoustic object by inverse scattering. Calculations of noise propagation from the data to the reconstructed image in inverse scattering using the adjoint field algorithm are presented. These calculations are then specialized to the case of Born approximation. Finally, the similarities and differences between inverse scattering and X-ray tomographic covariance are pointed out.

10:35

1aSPb2. Gradient-based shape optimization of an acoustic horn. Daniel Noreland (Laboratoire de Mecanique et d'Acoustique, 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20, France, Erik Bangtsson, and Martin Berggren (Uppsala Univ., 751 05 Uppsala, Sweden)

Numerical shape optimization is employed in order to improve the transmission properties of an acoustic horn. Wave propagation in the horn is modeled by the Helmholtz equation, which is solved using a finite element method. The shape is modified by changing the position of the mesh points along the contour of the horn. The optimization uses a quasi-Newton algorithm, with the gradient of the cost function computed by solving the adjoint equation for the discretized forward problem. Key points are the mesh movement scheme and different procedures in order to enforce smooth design updates for a large number of degrees of freedom. Convergence is fast, and the resulting horns have good transmission properties over a wide frequency band.

10:55

1aSPb3. Adjoint-based sensitivity analysis of fluid-structure interaction vibrations. Haw-Ling Liew and Peter M. Pinsky (Dept. of Mech. Eng., Stanford Univ., Stanford, CA 94305-4040, pinsky@stanford.edu)

Adjoint-based sensitivity analysis for the cavity fluid-structure interaction vibrations problem, with the inclusion of prestress and internal pressurization, is considered. The formulation employs a continuum approach in which the adjoint model is defined by employing a new duality relation between the sensitivities and the adjoint variables. For complex problems, such as fluid-structure interaction, the new duality relation provides an unambiguous approach for the construction of the adjoint variational problem in contrast to the standard approach. It also provides useful insight into the interpretation of the adjoint method. The resulting adjoint formulation for sensitivities is based on evaluation of inner products of the adjoint and state variables and is efficient because it requires the calculation of the adjoint field only once, regardless of the number of decision variables. In order to assess the accuracy of the adjoint approach for sensitivities in fluid-structure interaction problems, a model problem is considered and exact and analytical adjoint sensitivities are compared. The method is then applied to a finite element human eye model to compute the sensitivities with respect to elasticity constants and intraocular pressure for application to inverse estimation of those properties. The results are presented using a particularly effective multidimensional visualization technique.

11:15

1aSPb4. An efficient reconstruction algorithm for inverse scattering based on sensitivity analysis and adjoint equations. Gonzalo R. Feijoo (Woods Hole Oceanograph. Inst., MS 11, Woods Hole, MA 02543)

A common approach when developing imaging algorithms is to construct a linear model that relates the measured data to the properties of interest in the medium. But often a linear model cannot be constructed due to the relevant physics of the problem under study. In these cases, approaches based on optimal control theory offer several advantages: (1) a unified theory can be constructed to deal with different underlying physics and (2) a generic set of algorithms can be developed to solve the resulting system of nonlinear equations. Nonetheless, these methods tend to be too costly for solving problems but those involving few parameters. In this talk a new efficient algorithm for the inverse obstacle problem is presented without making the usual assumptions of a weak scatterer. The algorithm is based on optimal control theory but, different from other methods, it does not rely on iterative techniques to obtain a solution. Rather, it computes a scalar field which is a first approximation to the location and shape of the objects. This scalar field is easily computed in terms of the incident radiation used to probe the medium and an adjoint field constructed from the measured scattering pattern.

Contributed Papers

11:35

1aSPb5. Analytic derivation of adjoint nonlocal boundary conditions for a stratified ocean bottom in parabolic approximation. John S. Papadakis, Evangelia T. Flouri (Inst. of Appl. & Computational Math., F.O.R.T.H., P.O. Box 1527, 71110 Heraklion, Greece), Matthias Meyer (Royal Netherlands Naval College, The Netherlands), and Jean-Pierre Hermand (Univ. libre de Bruxelles, B-1050 Bruxelles, Belgium)

In underwater acoustics various types of nonlocal boundary conditions have been developed to handle the semi-infinite bottom in parabolic approximations and to efficiently reduce the computational domain. This paper proposes new exact nonlocal boundary conditions suitable for a layered ocean bottom and presents an analytic derivation of the corresponding adjoint equations. The new boundary condition has the form of a Neumann-to-Dirichlet map (NtD) that explicitly contains the geoacoustic parameters of the stratified bottom, i.e., thickness, density, sound speed, and attenuation of each layer. By means of the analytic adjoint, exact gradient information can be obtained which in turn allows a direct inversion of these parameters using a gradient-based optimization scheme. [Work supported by Royal Netherlands Navy.]

11:50

1aSPb6. Inferring marine animal characteristics using multi-angle scatter. Jules S. Jaffe and Paul L. D. Roberts (Marine Physical Lab, Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

An important goal in using active sonar systems to assess underwater animals *in situ* is the inference of animal dimension. Although 180-deg backscatter provides an estimate of target reflectivity, in many situations the reflected energy is not easily related to animal size. An additional degree of freedom, rarely exploited, is to measure animal scatter at a multiplicity of angles. Although this is more challenging from a technical point of view, the benefit of this additional information can make it worthwhile. The interpretation of multi-angle scatter from marine organisms can

be most easily understood in the context of diffraction tomography. Depending upon the details of the scattering process, in fortuitous circumstances a complete set of angular data can be inverted to compute the three-dimensional scattering potential which can ultimately lead to a complete inversion for animal morphology. In cases where it is impractical to collect a complete set of angular data it is often possible to infer characteristic dimensions of the animal from an analysis of the obtainable multi-angle data. The interpretation of multi-angle scatter data in the context of diffraction tomgraphy therefore presents opportunities to infer ecologically significant information about animal populations.

MONDAY MORNING, 5 JUNE 2006 ROOM 552A, 9:30 A.M. TO 12:00 NOON

Session 1aUW

Underwater Acoustics and Acoustical Oceanography: High-Frequency Ambient Noise

Juan I. Arvelo, Chair
Johns Hopkins Univ., Applied Physics Lab., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099

Chair's Introduction—9:30

Invited Papers

9:35

1aUW1. Recent developments in ambient noise processing. Martin Siderius, Michael B. Porter (HLS Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130), and Chris H. Harrison (NATO UnderSea Res. Ctr., La Spezia, Italy)

For many sonar applications noise is an annoyance and great effort is expended to minimize its impact. It may be surprising that ocean acoustic noise can itself be used in ways that rival or surpass complex sonar systems with controlled sound sources. Although ocean noise has been studied for many years, recent developments have led to exciting new ways to exploit the noise field for practical applications. In particular, by cross correlating the noise measured on two sensors one can recover a reasonable estimate of the impulse response between the sensors. This coherent processing was first demonstrated for sensors horizontally separated, however, it was also recently applied to vertically separated sensors. The vertical configuration leads to a type of passive fathometer that not only determines water depth but also reveals the seabed layering. A second development uses beamforming on the incoherent noise field to estimate bottom reflection loss. Combining these coherent and incoherent techniques leads to powerful ways to determine the properties of the seabed. This presentation will provide a summary of these recent developments and will show how these are combined for profiling the geoacoustics of the seabed using noise. Examples will be shown using both measured and simulated data.

9:55

1aUW2. High frequency noise modeling: Dependence on application. Chris H. Harrison (NATO Undersea Res. Ctr., Viale S. Bartolomeo 400, 19138 La Spezia, Italy, harrison@nurc.nato.int)

It is tempting to think that one could build a practical, general purpose noise model suitable for high or low frequencies and any application. For many signal processing and system applications this may be true, and one could incorporate directionality and spatial coherence. Recent developments in the use of wind-and sea-surface noise to determine geoacoustic properties by cross correlating hydrophone signals or beams demonstrate the need for more detail. The correlation peaks that correspond to absolute distances or ranges clearly require (frequency domain) coherence in any simulation. In contrast, layer information derived from angular reflection loss measurements by spectral factorization does not require coherence since it originates from an angular power spectrum. In both these cases the assumption of a uniform sheet of incoherent sources is adequate. However, to investigate the temporal statistics in either case, for instance to estimate averaging times, requires a fairly detailed model of the surface wave spatial field. Thus, the noise model is very much application dependent. These mechanisms and effects will be discussed in the context of experimental data.

10:15

1aUW3. Air-deployable sonobuoy array concepts and their performance against high-frequency noise anisotropy. Juan I. Arvelo, Jr., Patrick Ferat, and Ronald Mitnick (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099)

High-frequency sonobuoy arrays with vertical aperture were developed and tested to exploit the noise anisotropy for enhanced array gain [Ferat and Arvelo, J. Acoust. Soc. Am. **114**, 2462–2463 (2003)]. Afterwards, an inexpensive and conformal piezoelectric wire material was employed in the design and construction of a volumetric array [Arvelo *et al.*, J. Acoust. Soc. Am. **116**, 2650 (2004)]. However, detection performance of air-deployed sonobuoys is severely affected by noise interference through the beamformers sidelobes. An improved sonobuoy system design with in-buoy processing for sidelobe suppression is introduced and its detection performance in anisotropic wind noise with nearby shipping interference is evaluated and compared against previous designs in selected environmental conditions. The high-frequency array gain for increased vertical aperture will be compared against the gain for increased horizontal aperture to explore coherence limitations. The effects of array tilt, as well as interelement amplitude, phase, and topology errors, on the performance deterioration of the new design will also be addressed. [This effort is supported by the Maritime Sensing (MS) Program of the Office of Naval Research (ONR Code 321).]

10:35–10:50 Break

10:50

1aUW4. Ambient noise directivity in the East China Sea: Model/data comparison. Daniel Rouseff and Dajun Tang (Appl. Phys. Lab., College of Ocean and Fishery Sci., Univ. of Washington, Seattle, WA 98105)

In general, the vertical directivity pattern of high-frequency ambient noise is anisotropic. For measurements made beneath a shallow-water mixed layer, the observed noise directivity often displays a trough in the horizontal called the ambient noise notch. This notch develops because noise sources near the sea surface generate sound that propagates at relatively steep grazing angles. The directivity pattern is also generally asymmetric about the horizontal notch. This asymmetry develops because the seabed is an imperfect reflector of sound relative to the sea surface. In the present work, ambient noise data taken in the 1- to 5-kHz band during the 2001 East China Sea Experiment are analyzed and compared to models. The noise notch model integrates three components: an internal wave model, a noise model, and an acoustic propagation model that uses transport theory. It is shown how the depth and width of the observed noise notch are consistent with model predictions when internal waves of moderate strength are included in the calculation. The observed asymmetry in the directivity pattern is consistent with predictions when the correct bottom reflectivity is used. Future ambient noise measurements and tests for the models are described. [Work supported by ONR.]

11:10

1aUW5. Low-frequency noise from a bubble plume formed by a plunging water jet. Michael J. Buckingham (MPL, SIO, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0238)

Wind-driven, low-frequency ambient noise in the ocean may be due to bubble clouds created by wave breaking. To investigate this idea, a laboratory experiment was performed in which a vertical jet of fresh water plunged into a fresh water pool, to form a conical plume of bubbles beneath the surface. Both the jet speed, uj, and the air entrainment ratio, q, were carefully controlled and monitored. The power spectra of the sound from the plume, measured with an adjacent hydrophone, exhibit a sequence of nonuniformly spaced peaks below 1 kHz, which are associated with the resonances of the bubble-plume cavity. Each eigenfrequency scales accurately as $(1/uj)^{1/2}$ and approximately as $(1/q)^{1/4}$. A theoretical analysis of these scalings, based on a momentum-flux argument, shows that the sound speed profile within the plume varies as the square-root of depth down the axis. Using this square-root profile in the wave equation, an analytical solution for the pressure field within the plume is obtained, from which an explicit expression for the eigenfrequencies is derived. These theoretical eigenfrequencies, predicted from the conical geometry and square-root sound speed profile, exhibit the same nonuniform spacing and inverse-fractional power-law scalings as seen in the data. [Research supported by ONR.]

Contributed Papers

11:30

1aUW6. Robustness and constraints of high-frequency ambient noise inversion for sonar performance estimation. Juan I. Arvelo, Jr. (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099) and Andrea Prosperetti (John Hopkins Univ., Baltimore, MD 21218)

Perhaps the most dominant source of error in sonar performance estimation in shallow waters is the geoacoustic description of the seafloor. Various investigators have studied the possible use of ambient noise to infer some key parameters such as the critical angle, geoacoustic properties, or bottom loss. Most operational sonar performance models and tactical decision aids require the bottom loss as a function of grazing angles and frequency to describe the reflected acoustic energy off the ocean floor. Harrison and Simons [Harrison & Simons, ''Geoacoustics inversion of ambient noise: A simple method,'' J. Acoust. Soc. Am. **112**, 1377 (2002)] described a simple measurement approach to infer the bottom loss from ambient noise measurement on a vertical line array. This approach is well suited as an environmental measurement approach during naval operations. The robustness and operational constraints of this measurement approach are quantified using a high-frequency ambient noise model that accounts for wind, shipping, and thermal noise [Ferat & Arvelo, ''Mid-Frequency Ambient Noise Directivity And Array Response in Range-Dependent Environments,'' High-frequency Ocean Acoustics Conference Proceedings, 2004)]. The effects of array tilt, nearby shipping noise, water absorption, thermal noise, surface-noise dipole pattern, and array beam response on sonar performance estimation are demonstrated. Mitigating factors to maximize estimated bottom loss accuracy will also be outlined.

11:45

1aUW7. Emergence of the Green's function from ambient noise in moving media. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Earth System Res. Lab., 325 Broadway, Boulder, CO 80305, oleg.godin@noaa.gov)

A rather general proof of a connection between two-point correlation function of ambient noise and the elastodynamic Green's function relies on the reciprocity principle [K. Wapenaar, Phys. Rev. Lett. **93**, 254301 (2004)]. In this paper, the flow reversal theorem [O. A. Godin, Wave Motion **25**, 143–167 (1997)] is applied to extend this result to moving, inhomogeneous fluids and problems involving flow-structure interactions. In the motionless case, the time derivative of the cross-correlation function of ambient noise gives an estimate of the sum of the causal and acausal time-dependent Green's functions between the two receivers. In the moving medium, the result is shown to be similar, but the acausal Green's function corresponds to sound propagation in a medium with reversed flow. Therefore, instead of redundant information, combination of the causal and acausal Green's functions contains additional useful data on flow velocity field in the medium. The data contained in the two-point cross-correlation function of ambient noise are analogous to the data that would be obtained in a reciprocal transmission experiment employing acoustic transceivers located at the two points. Possible applications of the theoretical results obtained to passive remote sensing of ocean and atmosphere are discussed. [Work supported in part by ONR.]

Session 1pAA

Architectural Acoustics and Noise: Acoustics of Multi-Family Housing

Gary W. Siebein, Chair
Univ. of Florida, Dept. of Architecture, P.O. Box 115702, Gainesville, FL 32611-5702

Chair's Introduction—1:00

Invited Papers

1:05

1pAA1. Comparison of measurement indices of noise intrusions in multi-family housing. Gary Siebein, Hyun Paek, and Chris Jones (Siebein Associates, Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607)

Methods for measuring sound levels associated with transportation, impact, and airborne sound intrusions in multi-family housing are well established in the field. This paper compares field testing of assemblies where complaints and retrofits have been involved where traditional assessments of intruding noises showed compliance with design criteria, but where residents and/or building owners perceived problems. Case study 1 involves noise from airplanes approaching a runway at a large international airport as heard in an all-glass high-rise condominium evaluated by LDNs and SELs of actual flyovers in computer models and in a full size mock-up of a typical unit built on site. Case study 2 involves noise from outdoor amplified entertainment propagating into neighborhoods as evaluated by various noise ordinance criteria. Case study 3 involves footstep noise through flooring systems comparing IIC ratings with actual sound pressure levels of people walking on floors above. Case study 4 compares noise and vibration levels for various pieces of mechanical equipment before and after retrofit with NC, RC, and other room criteria. Auralizations of the case studies will be shown along with measurement data to illustrate the diagnostics made in each case.

1:30

1pAA2. Legal and acoustical issues in multifamily housing. Jeff Lapin (Lapin and Leichtling, LLP, 200 South Biscayne Blvd., Ste. 3150, Miami, FL 33131)

There are a plethora of legal issues that arise with respect to the suitability of design criteria and acoustical performance of multifamily housing properties. When complaints arise from residents, field testing can be done to determine if the units as constructed meet the design criteria or condominium documents. Questions can arise about the reasonableness of the criteria, the scientific validity of the testing process, the evidentiary value of test results, the type and extent of reasonable remedies, and allocation of responsibility among interested parties in resolving the dispute. This paper will identify the legal issues involved in disputes between private parties regarding acoustical issues such as soundproofing, as well as the use of acoustical criteria and testing as forming the legal basis for complaints and as a vehicle for resolving disputes involving the acoustical performance of buildings once they are occupied. The rights and duties existing under the common law of nuisance which allows people to enjoy their property free from unresonable interference will be examined along with the impact prescriptive requirements in building codes or condominium documents has on disputes involving acoustical performance in multifamily housing. The specific laws of one state will be used in the case studies as examples of situations that can arise where acoustical consultants deal with legal issues in multifamily housing.

1:55

1pAA3. Field airborne sound isolation in multifamily construction evaluation of metrics within ASTM E336. John J. LoVerde and David W. Dong (Veneklasen Associates, 1711 Sixteenth St., Santa Monica, CA 90404, jloverde@veneklasen.com)

ASTM E336 defines several metrics describing airborne sound isolation for wall and floor/ceiling assemblies within multifamily dwellings. The standard mirrors the procedures for laboratory testing of sound transmission class (STC) in defining field STC (FSTC), which does not include flanking paths and is therefore virtually impossible to measure in the field. The current standard includes provisions for the measurements of apparent STC (ASTC) which include flanking paths, and field measurements within multifamily dwellings typically report ASTC even though applicable building codes require the reporting of FSTC. The standard also provides two noise reduction metrics, noise isolation class (NIC) and normalized NIC (NNIC). Analysis of the relationships of ASTC, NNIC, and NIC will be presented along with a survey of field test results and correlation with subjective reactions of occupants within multi-family dwellings. The analysis shows that the additional work required to determine the ASTC or NNIC is not warranted because it provides no additional useful data, and that the proper metric for evaluation of acoustical performance in the field for multifamily dwellings is NIC.

2:20

1pAA4. Design parameters for impact sound insulation in lightweight wood and steel structures. John J. LoVerde and David W. Dong (Veneklasen Assoc., 1711 Sixteenth St., Santa Monica, CA 90404, jloverde@veneklasen.com)

One of the challenges in the acoustical design of multifamily wood and steel joist buildings is reducing low-frequency sounds from footfalls on both carpeted and isolated hard-surfaced floors. Theoretical calculations and field-testing experience indicate that low-frequency impact noise in lightweight structures is clearly audible, and that traditional isolation methods do not reduce low-frequency impact noise to inaudibility [Blazier and DuPree, J. Acoust. Soc. Am. **96**, 1521–1532 (1994)]. We report numerous measurements performed on wood-frame assemblies in the United States using the tapping machine method as defined in ASTM E1007. The impact sound-pressure level was measured down to the 20-Hz third-octave band, instead of 100 Hz as described in the standards. Representative case studies will be presented showing the effect of changes in construction on occupant complaints. The results indicate that the unweighted level in the 50–80 Hz third-octave bands correlates much better with occupant reaction than any of the other metrics [LoVerde and Dong, J. Acoust. Soc. Am. **112**, 2201 (2002); LoVerde and Dong, INCE Inter-Noise 2004 Proceedings, 167 (2004)]. Factors in the design of successful lightweight floor/ceiling assemblies, including structural design parameters, shall be presented along with corresponding acoustical performance and subjective reactions.

2:45

1pAA5. Old mills, new condos; sound isolation in mill building conversion projects. Benjamin Markham, Jonah Sacks, and Rose Mary Su (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

Up and down the New England Coast, century-old mill buildings are being converted into condominium and apartment buildings amidst one of the biggest housing booms in recent memory. While the inherent acoustical conflicts in multi-family dwelling are not new (oft-cited HUD guidelines are approaching their 40th birthday), the flood of mill conversions is bringing to light a number of new constructions that architects use to chop former factories into discrete living spaces. This paper will present field-collected data and case studies that illustrate some of the problems and common pitfalls associated with mill building conversion projects. Among the construction details discussed include exposed timber ceilings that run continuously across gypsum board partitions, the creative reuse of existing brick walls, and some acoustical properties of wood deck construction common to many mill buildings.

3:10–3:25 Break

3:25

1pAA6. Comparison of acoustical measurements of demising walls in luxury condominiums and townhomes to illustrate the importance of decoupling the mass layers. Robert M. Lilkendey and Gary W. Siebein (Siebein Associates, Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607)

Field Sound Transmission Class (FSTC) tests and acoustical measurements of impact noise from cabinet door and drawer impacts were made of demising walls in a number of multifamily luxury townhomes and luxury condominium tower units due to complaints from residents about noise intrusion from adjacent living units. The demising wall constructions included concrete or cmu with gypsum board attached, gypsum board on wood studs, and gypsum board on metal studs. The results of these measurements will be presented to illustrate the need to decouple the mass layers to maximize the acoustical performance. A case study of one project will be presented showing the effects of incremental changes made to a base wood stud wall with gypsum board on each side (STC 36) in which added mass (STC 44), added mass plus decoupling (STC 62), and decoupling with reduced additional mass (STC 58) were tested in an effort by the developer to strike a balance between acoustic performance and cost.

3:50

1pAA7. Condominium noise issues: Problem solving with flooring systems and solving flooring system problems. Jospeh Still (Proflex Products, Inc., 3406 Dean St., Naples, FL 34104, proflex@verizon.net)

With increasing frequency, condominiums are being designed without suspended ceilings. At the same time, the popularity of hard floor coverings such as tile continues to grow. These factors combine to create significant problems with impact noise in these buildings. Architects and acoustical engineers have looked to the flooring industry for noise-reduction solutions, and the response has been the production of a wide variety of sound-reducing systems. However, not all of these impact insulation sound reduction systems are acceptable for hard-surface flooring. The total performance of the floor/ceiling system including acoustical as well as durability tests should always be carefully scrutinized, evaluated, and verified by acoustical consultants. Furthermore, strict adherence to proper installation techniques by contractors is essential to achieving the sound reduction performance that should be expected of a given product. Among the major topic areas addressed in this paper are the current flooring industry standards for underlayment systems; realistic expectations for sound reduction and durability from various types of flooring and underlayment systems; and recommended changes in architectural specifications for flooring. This paper also includes practical guidelines for evaluating and verifying the claims made for flooring systems regarding impact sound reduction and several problem-solving examples from real-world applications.

4:15

1pAA8. Noise in a multifamily housing area in Mexico City. Sergio Beristain (P.O. Box 12-1022, Narvarte, 03001, Mexico, City, Mexico, sberista@hotmail.com)

Noise in a multifamily housing site is complex and hard to handle due to economic, cultural, educational, income, and many other reasons. The site studied here is surrounded by commercial and services areas, which increases the noise problem. Most of the complaints do not go too far due to lack of specific regulation and lack of knowledge of which existing one might be applicable, but even when legal action is taken, noise is hard to evaluate mainly because noise sources are off when authority shows up.

4:30

1pAA9. A jury study of noise isolation between condo units incorporating QuietZone® acoustic wall framing in the common wall. Dick Godfrey, Harry Alter, and Clarke Berdan, II (Sci. & Technol., Owens Corning, 2790 Columbus Rd., Granville, OH 43023)

Two identical side-by-side two-story condo units with basements were built in the newly built Owens Corning Building Science and Innovation Center. These units incorporated modern features such as multistory great rooms with adjoining kitchens. Home offices, bedrooms, and basements were located on opposite sides of the common wall. It is common practice to use double-stud walls between adjacent units in multifamily construction. It is also good practice to have a noncontiguous floor plate between units. Instead of a double-stud wall, these units incorporated QuietZone® acoustic wall framing (resilient stud), and other Owens Corning recommended noise control measures. The wall transmission loss characteristics were measured in the field and in the laboratory. A laboratory jury study was conducted to investigate the adequacy of perceived noise isolation between the units for various source and background noise types and levels.

4:45–5:15
Panel Discussion

MONDAY AFTERNOON, 5 JUNE 2006 ROOM 550AB, 1:00 TO 4:55 P.M.

Session 1pAB

Animal Bioacoustics: Acoustic Interactions in Animal Groups II

Joshua Schwartz, Chair
Pace Univ., Dept. of Biology, Pleasantville, NY 10570

Invited Papers

1:00

1pAB1. Signal detection theory and noisy communication. Lori Wollerman (Dept. of Biol., Hood College, 401 Rosemont Ave., Frederick, MD 21701, lwollerman@hood.edu)

Signal detection theory (SDT) has had little influence on thinking about the evolution of animal communication, despite its wide application in psychology. SDT is a model that describes the detection of signals in the presence of noise, a situation that clearly applies to animals that communicate in aggregations, such as many frogs and singing insects. Through SDT we can better understand fundamental issues about the evolution of communication. In many realistic signaling conditions receivers are unable to avoid making an error of one sort or another regardless of the criterion they use for responding to signals. Signals that are more exaggerated would be more detectable and receivers would be more likely to respond to these signals than to those that are less exaggerated and less detectable. Signal detection theory therefore provides a sufficient explanation for the exaggeration and elaboration of male displays used to attract females for mating. Moreover, principles of SDT apply to any communicatory interaction. Thus, SDT establishes universal conditions that any hypothesis regarding the evolution of signals must satisfy.

1:20

1pAB2. Discrimination of noise-contaminated communication signals by grasshoppers. Bernhard Ronacher (Biol.. Dept., Humboldt Univ., Berlin, Germany)

When using acoustic signals in order to find—and possibly select between—mates, grasshoppers face two major problems: the arriving signals are usually degraded by various kinds of noise, while as much information as possible should be extracted quickly. Major cues for signal recognition are the temporal patterns of amplitude modulations of grasshopper songs. Within the nervous system, signal classification is rooted in the spike trains provided by the sensory neurons and will be constrained by the variability of spiking responses. Several intrinsic sources further contribute to spike-train variability, which can be considerable even in the case where undistorted acoustic signals arrive at the receiver. Using van Rossum's [Van Rossum, M. Neural. Comp. **13**, 751–763 (2001)] method to compute spike-train distances, we quantified the variability of neural responses and showed that the temporal information

necessary to discriminate between songs of different conspecifics is preserved surprisingly well, even in the responses of single neurons. By comparing behavioral and neurophysiological data, we now aim at determining how well this information can be used by the animals, and how closely or remotely their nervous systems approach the performance of an ideal observer.

1:40

1pAB3. Mechanisms and evolution of chorusing interactions in acoustic insects and anurans. Michael Greenfield (Dept. of Ecol. & Evolutionary Biol., Univ. of Kansas, Lawrence, KS 66045)

Chorusing displays of acoustically signaling insects, and their counterparts among bioluminescent species, are acknowledged to include some of the great spectacles of the living world [Wilson, *Sociobiology*, Harvard University Press (1975)]. These communal displays are noteworthy for the sheer numbers of participating individuals, the volume of sound or brilliance of light produced, and a precise choreography in space and time, which often gives rise to striking alternation or synchrony between neighbors. Here, potential mechanisms that may generate such temporally structured choruses are reviewed and then the evolution of these communal displays is discussed. Under certain conditions alternation or synchrony may be adaptive *per se*, and the mechanisms that generate chorusing interactions are directly favored by selection. But, in other cases alternation and synchrony are merely epiphenomena, group effects that emerge as by-products of simple interactions between signaling neighbors; these interactions are, in effect, responses to psychoacoustic effects in perception. Simulation models inspired by various acoustic insects demonstrate how complex choruses may represent emergent properties of these simple responses. The models also show how feedback loops may arise by which these emergent properties can yet influence the simple interaction responses from which they originated.

2:00

1pAB4. Call synchronization in *Neoconocephalus nebrascensis* (Orthoptera, Tettigoniidae): Cooperation rather than competition? Johannes Schul, Katy E. Klymus, and Joshua A. Deily (Biological Sci., Univ. of Missouri, 207 Tucker Hall, Columbia, MO, 65211, schulj@missouri.edu)

Male calls of the katydid *Neoconocephalus nebrascensis* consist of 1000-ms-long verses separated by silent pauses of 800 ms. Verses of neighboring males exhibit remarkable synchrony during which leader-follower relationships often remain unchanged over extended periods of time. Such call synchrony is often the consequence of female preference for leading calls. We tested female phonotaxis on a walking compensator in response to male calls presented alone or as pairs. Females did not respond to continuous calls, but required a distinct verse structure for positive phonotaxis. Surprisingly, more than 18-dB modulation depth between verses and silent pauses was required to elicit female responses. When the delay between two simultaneously presented calls was between 0 and 600 ms, females did not exhibit any biologically relevant leader preference. Females did not show significant phonotaxis to alternating calls. Given that females require a modulation depth of 18 dB, call alternation even between distant males would effectively mask the attractive temporal pattern of both males' calls. Only through call synchronization do males achieve an attractive pattern. This, together with the absence of a biologically relevant female preference for leading calls, indicates that males do not compete but rather cooperate during call synchronization to preserve an attractive temporal call pattern.

2:20

1pAB5. Chorusing in fishes. David Mann and James Locascio (Univ. of South Florida, College of Marine Sci., 140 7th Ave. S., St. Petersburg, FL 33701, dmann@marine.usf.edu)

Many fishes in the family Sciaenidae engage in nightly chorusing events where sound production starts near dusk and lasts for several hours. This paper examines the periodicity and dynamics of chorusing in the Sciaenidae. Underwater acoustic recorders have been deployed in southwest Florida to record sound production with a high temporal resolution (10 s of recording every 10 min) over a large spatial area. Sound production was dominated by three sciaenid species: silver perch (*Bairdiella chyrsoura*), sand seatrout (*Cynoscion arenarius*), and spotted seatrout (*Cynoscion nebulosus*). Sounds produced by black drum (*Pogonias cromis*), gafftopsail catfish (*Bagre marinus*), and gulf toadfish (*Opsanus beta*) were also recorded. Sound levels in sciaenid choruses typically reach levels of approximately 110–120 dB *re* 1 μPa spectrum level, and are among the loudest sources of sustained ambient noise in temperate and subtropical coastal waters. Little is known about the behavior accompanying sound production, because it occurs in turbid water at night. The use of hydrophone arrays to document fine-scale spatial patterns of sound production will be discussed.

2:40–3:00 Break

3:00

1pAB6. Social complexity can drive vocal complexity: Group size and information in chickadee calls in Carolina chickadees. Todd Freeberg (Dept. of Psych., Univ. of Tennessee, Knoxville, TN 37996, tfreeber@utk.edu)

There is considerable variation in systems of vocal communication in animals. One hypothesis to explain this variation is that the complexity of the social group influences the groups vocal complexity. This social complexity hypothesis for communication is also central to recent arguments for human language origins, but experimental tests of the hypothesis are lacking. This study investigated whether group size, a fundamental component of social complexity, influenced the complexity of the chickadee call, a call system functioning in social organization of Carolina chickadees, *Poecile carolinensis*. This call system has been studied in chickadees and related species, and is structurally and functionally rich in its note composition. Complexity of calls was measured using uncertainty (diversity) related to note composition, based on Shannon's mathematical theory of communication. In unmanipulated field settings, calls of individuals in larger groups had greater complexity (more information) than calls of individuals in smaller groups. In aviary settings manipulating group size, individuals in larger groups used calls with greater complexity than individuals in smaller groups. These results indicate that social complexity can influence communicative complexity in this species.

3:20

1pAB7. Acoustic interactions in broods of begging nestling birds. Andrew G. Horn and Marty L. Leonard (Dept. of Biol., Dalhousie Univ., Halifax, NS B3H 4J1 Canada)

Begging by nestling birds has become a model system in the study of the evolution of reliable signaling. The transmission and reception of begging signals, however, remain largely unexplored. Especially puzzling is how, or indeed whether, the calls that accompany most begging displays can convey information to parents, given that several nestlings may be calling at once within the cramped confines of the nest. Most previous work on groups of acoustic signalers has involved signalers that are much more widely spaced, such as chorusing insects and frogs. Here, we review our research on how the begging calls of tree swallows, *Tachycineta bicolor*, convey information to parents. Various call features correlate with nestling need, and playing back calls to parents affects both how often they deliver food to the nest and which nestling they feed once they arrive. Nestlings adjust their calls in response to signal interference, albeit crudely, by adjusting call rate and amplitude. Begging calls and the contexts in which they occur vary enormously across species, so that other species likely use different strategies to overcome acoustic interference. Indeed, begging displays offer promising and barely tapped opportunities for developmental and comparative studies of acoustic signaling in groups.

3:40

1pAB8. Vocal exchanges in wild bottlenose dolphins. Nicola J. Quick and Vincent M. Janik (Sea Mammal Res. Unit, Univ. of St Andrews, Fife, Scotland, KY16 8LB, UK, njq@st-andrews.ac.uk)

This study investigated the behavioral context of stereotypic whistle exchanges in wild bottlenose dolphins off north-east Scotland to infer function from whistle usage. Concurrent acoustic and nonacoustic behavioral data sampling was conducted during focal follows. Passive acoustic localization using a four-element distributed array was used to identify vocal exchanges (i.e., whistles from different individuals within 3 s of each other). Vocal exchanges of stereotypic (two whistle types produced repeatedly by at least two animals) nature occurred during 12 focal follows from 10 separate days containing 108 whistles that could be localized. Exact randomization tests (10 000 iterations), using whistles from exchange and nonexchange periods within each follow, showed that the observed number of exchanges (i.e., whistles produced by different individuals within 5 s of each other) was significantly different than the value that resulted from randomization. Vocal exchanges occurred prior to dolphin subgroups joining in 9 out of 10 observed joins. Only 2 of 11 vocal exchanges were not followed by a join. Vocal exchanges of stereotyped whistles between individuals appear to be important in coordinating joins in wild dolphin groups. Further analysis of exchange whistles showed that their degree of stereotypy equalled that found in signature whistles.

4:00

1pAB9. Convergence of calls as animals form social bonds. Peter L. Tyack (Biol. Dept., Woods Hole Oceanogr. Inst., Woods Hole, MA 02543, ptyack@whoi.edu)

In many species capable of vocal production learning, the acoustic properties of calls from different individuals can converge when they are housed together in captivity or form social bonds in the wild. This has been demonstrated for animals as diverse as songbirds, parakeets, bats, cetaceans, and primates. In many species of songbirds, territorial neighbors have repertoires with several shared songs. Males can use the shared songs during territorial disputes, and birds escalate or de-escalate a countersigning interaction by choosing to reply with a song match, a repertoire match, or an unshared song. For most other species, call convergence is thought to reflect a group-distinctive identifier, with shared calls reflecting and strengthening social bonds. I will discuss examples of call convergence in primates, birds, elephants, and cetaceans. The breadth of the use of production learning for call convergence suggests that this may have been an important selective pressure for the evolution of production learning. Some species have well-developed capabilities of imitating distinctive calls of other individuals or groups. This capability of learning to imitate an arbitrary signal for an arbitrary referent may have evolved out of simpler forms of vocal convergence.

4:20

1pAB10. Bad vibrations: Mechanosensory risk assessment by red-eyed treefrog embryos. Karen M. Warkentin, Michael S. Caldwell (Dept. of Biol., Boston Univ., 5 Cummington St., Boston, MA 02215), and J. Gregory McDaniel (Boston Univ., Boston, MA 02215)

Red-eyed treefrogs, *Agalychnis callidryas*, lay gelatinous egg masses on vegetation overhanging ponds; tadpoles fall into the water upon hatching. Embryos hatch up to 30% prematurely in attacks by egg-eating snakes, but not in other intense but benign disturbances, such as storms. We used modal analysis and vibration recordings from clutches in natural disturbances to characterize the vibrational environment of embryos, and playbacks to assess if vibrations cue hatching, and how vibration characteristics affect hatching behavior. Escape hatching can be induced by snake vibrations, without actual snakes, and embryos distinguish between the vibrational patterns of snake attacks and rainstorms. Benign and dangerous disturbances overlap in individual temporal and frequency characteristics, but can be distinguished using combinations of features. Both frequency and temporal characteristics of vibrations affect the hatching response. Embryos use at least two temporal pattern elements—disturbance duration and interdisturbance interval—and respond to a limited range of low frequencies. These are all necessary elements of a composite cue. Moreover, higher frequencies, found in rainstorms but not attacks, reduce hatching. This is the most detailed examination of mechanisms underlying embryo decision making to date, and a demonstration of the importance of vibrational information in predator-prey interactions. [Work supported by NSF IBN-0234439.]

Contributed Paper

4:40

1pAB11. Attracting insects with an acoustic laser. Kent Lau and Steven Garrett (Schreyer Honors College and the Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA 16804-0030)

It has been demonstrated that female Mediterranean fruit flies (*Ceratitis capitata*) are attracted to a trap by a pure tone at approximately the spectral peak (330 Hz) of the males mating call [Mankin *et al.*, J. Econ. Entomol. **97**, 1299–1309 (2004)]. Recent research [Sueur *et al.*, "Sound radiation around a flying fly," J. Acoust. Soc. Am. **118**, 530–538 (2005)] suggests that a more complex tone that combines an amplitude-modulated fundamental and a second-harmonic component could produce more effective acoustic bait. We will describe a thermoacoustic sound source [S. L. Garrett and R-L. Chen, Echoes **10**, 4–5 (2000)] that combines two closely tuned quarter-wavelength resonators designed to produce a beating fundamental of adjustable modulation depth. A third source produces the second harmonic. The trio of sources generates the desired signal at levels of 90 to 100 dB 1 meter from the array. Since the sound is produced from electrically heated filaments, no oscillators, amplifiers, or electro-acoustic transducers are required. When the pair of fundamental resonators are placed side-by-side, the beating occurs uniformly in space. When placed back-to-back, the dipolar radiation pattern rotates to produce the beating in time at any fixed location. We hope to have results with insect subjects by the time of the meeting. [Work supported by an endowment from United Technologies Corporation.]

MONDAY AFTERNOON, 5 JUNE 2006 ROOM 552A, 1:00 TO 5:00 P.M.

Session 1pAO

Acoustical Oceanography: Acoustical Oceanography Inversion

Gopu Potty, Chair

Univ. of Rhode Island, Dept. of Ocean Engineering, Narragansett, RI 02882

Contributed Papers

1:00

1pAO1. Localization and geoacoustic inversion by using 20-Hz signals of finback whales on the New Jersey Shelf. Altan Turgut and Bruce Pasewark (Naval Res. Lab., Acoust. Div., Washington, DC 20375)

One-second duration 20-Hz signals of finback whales were recorded on three vertical line arrays (VLAs) and a bottomed horizontal line array (HLA) during the RAGS03 experiment conducted on the New Jersey Shelf. Modal analysis indicated that most signals, recorded over a 3-week period in December of 2003, contained a weak, early-arriving mode-2 and strong, late-arriving mode-1 component. These modal characteristics are exploited in a broadband matched-field algorithm to localize individual finback whales and to perform geoacoustic inversions. Bearing estimation of individual whales is obtained by performing horizontal beamforming on the HLA data. Range estimation results are verified by time-of-flight triangulation using single hydrophone data from each VLA location. Geoacoustic inversion results were comparable to those previously obtained from chirp sonar data. Localization results indicated that most finback whales traveled near the shelfbreak front where food might be abundant. [This work was supported by the ONR.]

1:15

1pAO2. A method to reduce geoacoustic inversion uncertainties caused by water column fluctuations. Ying-Tsong Lin and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

A linear method is developed to reduce the effects of small sound-speed fluctuations in the water column on geoacoustic inversions. Sound-speed fluctuations are considered as a source of physical model error, and their influence on the acoustic data used for determining geoacoustic parameters is estimated by linear perturbation theory. The idea of this method is to project acoustic data onto a space that is insensitive to water sound-speed fluctuations. To determine this space, the singular value decomposition is applied to the kernel of the linear perturbation equation, which links the fluctuations to the acoustic data. This approach will reduce the number of data points available for inversion. Because of this, empirical orthonormal function analysis is employed to represent the sound-speed fluctuations by a small number of modes, which effectively increases the number of projected data points. This method is demonstrated by a numerical test case, using linear internal waves with the Garrett-Munk spectrum to vary the water column in time and space. The influence of the higher-order terms neglected in the linear perturbation equation on the performance of this method is also discussed.

1:30

1pAO3. Tabu inversion in the East China Sea. Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102, michalop@njit.edu), James H. Miller, and Gopu R. Potty (Univ. of Rhode Island, Narragansett, RI 02882)

Tabu, a global optimization method that exploits search history to explore the parameter space in a structured manner, has been shown to perform well in geoacoustic inversion with both real and synthetic data. To investigate the efficacy and adaptability of Tabu search to various ocean environments, we conducted Tabu inversion for broadband data (at frequencies between 20 and 600 Hz) collected in the East China Sea during the ASIAEX experiment. A simple incoherent Bartlett objective function was considered. Localization results and geoacoustic estimates obtained via Tabu closely match ground truth information and parameter estimates obtained with other inversion methods. Appropriate hierarchical ordering of the unknown parameters ensured identification of the global maximum of the objective function within only a few forward model calculations. [Work supported by ONR.]

1:45

1pAO4. Validation of statistical estimation of transmission loss in the presence of geoacoustic inversion uncertainty. Chen-Fen Huang, Peter Gerstoft, and William S. Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, California 92093-0238)

Often the ocean acoustic environment is not well known and sonar performance prediction will be affected by this uncertainty. Here, a method for characterizing transmission loss (TL) is proposed which incorporates these environmental uncertainties. The statistical estimation of TL based on the posterior probability density of environmental parameters obtained from the geoacoustic inversion process is the main theme of this

study. First, a Markov chain Monte Carlo procedure is employed to sample the posterior probability density of the geoacoustic parameters. Then, these parameter uncertainties are mapped to the transmission loss domain, where a full multidimensional probability distribution of TL as a function of range and depth is obtained. In addition, TL is also characterized by its summary statistics including the median, percentiles, and correlation coefficients. The approach is illustrated using a data set obtained from the ASIAEX 2001 East China Sea experiment. Based on the geoacoustic inversion results, the predicted TL and its variability are estimated and then compared with the measured TL. In general, the results show that approximately 80% of the observed TL data fall within 90% of the TL probability distribution.

2:00

1pAO5. Passive matched-field inversion for thick-sediment characterization. Donald R. DelBalzo, Robert S. Linzell, and Lisa A. Pflug (Planning Systems, Inc., 40201 Hwy. 190 E., Slidell, LA 70461, ddelbalzo@plansys.com)

Acoustic field estimation in shallow-water, littoral areas is severely limited by environmental uncertainty, especially regarding sediment property characteristics. Direct methods of obtaining geophysical data are expensive, so inverse methods have recently become popular. Most techniques rely on inverting signals from controlled active sources. This work is focused on inverting signals with unknown source levels from surface ships of opportunity at unknown ranges. A range-staggered technique is employed to create a range-dependent sediment description, with ever-increasing confidence as additional surface ships are analyzed. Matched-phase techniques are applied in a simulated shallow-water environment with a single vertical array and high signal-to-noise ratios. The inversion process is evaluated as a function of frequency bandwidth, number of hydrophones, and array length. The simulations indicate significant potential for passive, range-dependent, thick-sediment characterizations out to ranges of tens of water depths in shallow water. [Work sponsored by SPAWAR.]

2:15

1pAO6. Bayesian layer-stripping inversion of seabed reflection data. Jan Dettmer, Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria, BC, Canada), and Charles W. Holland (Appl. Res. Lab., The Penn State Univ., State College, PA)

This paper develops a Bayesian inversion technique for recovering multilayer geoacoustic profiles using seabed reflection data. The measured data originate from acoustic time series windowed for a single bottom interaction, which are processed to yield spherical reflection coefficients (i.e., a response function of frequency and angle analogous to plane-wave reflection coefficients). Replica data are computed using a wave number-integration model (OASES) to calculate the full complex acoustic pressure field, which is processed to produce a similar seabed response function. The inversion results are compared to those obtained using plane-wave reflection coefficients. To address the high computational modeling costs, the Bayesian algorithm is implemented for a massively parallel computer. Further, the data are time windowed and divided into several layer packets, wherein each packet contains the seabed response to a certain depth. This layer-stripping approach uses the results of the previous layer packet as prior information for subsequent packets. The resulting posterior probability density for the final packet is considered the full solution to the inverse problem, and is interpreted in terms of optimal parameter estimates, marginal distributions, credibility intervals, and parameter correlations.

2:30

1pAO7. Estimation of poroelastic parameters from measured reflectivity. Tracianne Neilsen (Brigham Young Univ., N283 ESC, Provo, UT 84602) and Marcia Isakson (Appl. Res. Labs., Univ. of Texas at Austin)

The reflectivity of sound at a water-sediment interface contains information about the properties of the sediment. Estimates of these properties can be obtained by optimizing the match between the magnitude of the measured reflectivity, as a function of grazing angle and frequency, and values computed using a poroelastic sediment model. A traditional simulated annealing approach yields such parameter estimates but little information about the uncertainty associated with those estimates. An alternate approach computes the posterior probability distributions (PPDs) for the parameters using a fast Gibbs' sampler approach. The PPDs not only indicate the most probable values for each parameter, but the widths of the distributions also provide an understanding of the uncertainty and the dependence of the data-model mismatch on each parameter. Both of these methods are applied to reflectivity measured in a laboratory tank over a frequency band of 77–147 kHz and grazing angles of 7–49 deg. [Work suppported by ONR, Ocean Acoustics.]

2:45–3:00 Break

3:00

1pAO8. Time-reversal operator target focusing using optimal beam sets. G. F. Edelmann, J. F. Lingevitch, D. M. Fromm, C. F. Gaumond, and R. Menis (U.S. Naval Res. Lab., 4555 Overlook Ave., SW, Code 7145, Washington, DC 20375, edelmann@nrl.navy.mil)

Time-reversal operator (TRO) methods can be used to focus sound on an ocean target using a time-reversal mirror. In order to increase the signal-to-noise ratio, the TRO can be measured by transmitting simultaneously the entire array instead of transmitting source by source. The so-called beam-space. TRO is measured by transmitting a set of orthogonal beams. Acoustic propagation will be simulated in a realistic range and time-dependent ocean environment measured during the TREX04 experiment [Gaumond, J. Acoust. Soc. Am. (in press)]. This paper will compare four different sets of beams: Hadamard, phased planar, spatially phased, and modal. The first two methods require no *a priori* environmental knowledge, the third requires only waveguide depth, and the final requires a sound-speed profile and geoacoustics measured at the TRM. Results will be discussed in terms of signal gain, complexity to implement, and coherence in an inhomogeneous ocean environment varying in both time and space. [Work supported by the Office of Naval Research.]

3:15

1pAO9. Perturbation of arrival times in sound pulses propagating in shallow water through moving internal solitons. Mohsen Badiey (Ocean Acoust. Lab., Univ. of Delaware, Newark, DE 19716), Valery Grigorev, Boris Katsnelson (Voronezh State Univ., Voronezh 394006, Russia), and James Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Sound pulse propagation in shallow water is studied. We assume a geometry where an acoustic track crosses the wave-front of a moving internal soliton (IS) or train of solitons. During propagation, sound pulses undergo perturbation due to mode coupling. For each individual mode selected at the receiver, we observe a set of incoming pulses, each having different arrival times. Time-frequency analysis is carried out for the modal arrival times for different frequencies, emphasizing the dependence on position of the moving solitons. It is shown that the maximum amplitude of the incoming pulses created at the ISs due to mode coupling has approximately the same arrival time, independent of frequency. This invariant arrival time is determined by waveguide parameters and corresponds to maximal coupling of different pairs of modes. Analysis is carried out both for a model shallow water channel and for real experiment

data (SWARM95, NRL acoustic track). Satisfactory agreement is demonstrated, and physical interpretation of observed effect is discussed. [Work is supported by ONR and RFBR.]

3:30

1pAO10. Sediment models for the East Coast based on *a priori* information, geological data, and sediment inversions. Gopu R. Potty, James H. Miller (Dept. of Ocean Eng., Narragansett, RI 02882), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

A preliminary sediment model for the U.S. East Coast with particular emphasis on the Shallow Water-2006 experimental site is presented. This sediment model will describe the bottom as layers of sediments with different properties. The layering information will be based on geological, geophysical, stratigraphy models and inversion results. The surface sediment description will be based on the Effect of Sound on Marine Environment (ESME) sediment models for the U.S. East Coast. This sediment model, developed as part of the ESME initiative, draws from a variety of data sources including the sediment texture database maintained by the United States Geological Survey. This database coupled with Hamilton, Biot-Stoll (with reduced number of parameters), or Buckingham models (to define the sediment geoacoustic properties) has the potential for providing an efficient surface sediment representation with large spatial coverage. Other sources of data, which will be used to develop the preliminary sediment model, include published sediment core data from the region and inversions results. The use of empirical orthogonal function-based sediment parametrization for inversions will be explored. This study will provide sediment data for the preliminary experimental design efforts prior to the experiment in 2006. [Work supported by Office of Naval Research.]

3:45

1pAO11. Estimating geoacoustic parameters from broadband data from the New Jersey shelf. Yong-min Jiang, N. Ross Chapman (School of Earth and Ocean Sci., Univ. of Victoria, P.O. Box 3055, Victoria, BC V8W 3P6, Canada, chapman@uvic.ca), and Mohsen Badiey (Univ. of Delaware, Newark, DE 19716)

This paper presents the geoacoustic inversion of broadband data to estimate the sea bed parameters of SWARM95 site. The data set used in this study contains air gun signals that were transmitted at the depths of 25 and 45 m, respectively, below the sea surface, while the source ship was maneuvering at small speed (towing); the signals then were collected at the ranges around 5 km by a vertical line array. A canonical geoacoustic model was generated first according to previous experimental research results, for the purpose of a parameter sensitivity study; then both hybrid optimization and Bayesian inversion techniques were applied to the real data. The data error covariance matrix is estimated to resolve the spatial correlated data error in Bayesian inversion approach. Since the ratio of range and water depth is rather large in this study, only the geoacoustic parameters at the top of the sediment are sensitive in the inversion. The optimization results show the consistency of sensitive geoacoustic/geometric parameter estimations over the ranges and different source depths. Bayesian inversion gives the estimation uncertainties of the geoacoustic properties by marginal probability distributions. [Work supported by ONR.]

4:00

1pAO12. Comparison of transmission loss data with predictions based on inverted seabed parameters from the Yellow Sea '96 experiment. Lin Wan, Ji-Xun Zhou, and Peter H. Rogers (School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332-0405, gtg446p@mail.gatech.edu)

Transmission loss is useful in sonar performance prediction. In this paper, the transmission loss data as a function of range, frequency, and propagation direction are obtained from the Yellow Sea '96 experiment. The experiment was conducted along two radial directions up to 55 km and a quarter of a circle with a radius of 37 km. About 300 broadband explosive sources were detonated above and below the thermocline in the experiment. The seabed acoustic parameters are inverted from the broadband explosive source data. Based on the inverted seabed parameters, the transmission loss is predicted and compared with the experimental data. [Work supported by ONR.]

4:15

1pAO13. The Ionian Sea rainfall experiment. Jeffrey Nystuen (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Eyal Amitai (George Mason Univ., Fairfax, VA), Emmanouil Anagnostou, and Marios Anagnostou (Univ of Connecticut, Storrs, CT 06269)

Rainfall on the sea surface generates a loud and distinctive sound underwater. This sound propagates downward and is attenuated, producing an effective listening area, or an equivalent "catchment basin" for a listening device that is a function of depth and frequency. Acoustical measurements of rainfall are reported from 60, 200, 1000, and 2000-m depths from a mooring in the Ionian Sea off the southwestern coast of Greece (37N, 21.5E) from January to April 2004. These measurements are compared to colocated high-resolution coastal X-band radar (XPOL) rainfall reflectivity measurements. The XPOL reports the spatial distribution of the rain over the listening areas of the acoustic rain gauges (ARGs). Roughly ten rainfall events, including drizzle, squall line, and heavy rainfall were recorded. The results demonstrate the inherent spatial averaging of the rainfall signal with depth. This spatial averaging is desirable when comparisons are made to satellite measurements of oceanic rainfall as such measurements also have inherent spatial averaging. The data also verify that a viable signal for acoustical rainfall measurement is present at 2000 m, allowing this technology to be transferred to deep-ocean instrumentation platforms such as ARGO floats and bottom-mounted cabled systems. [Work supported by NSF Oceanography.]

4:30

1pAO14. Haro Strait geometry from time-domain data. A. Tolstoy (A. Tolstoy Sci., 1538 Hampton Hill Circle, McLean, VA 22101)

This paper presents results which relate to the geometry of various Haro Strait scenarios. In particular, from the time-domain data we can estimate the source range, source depth, individual phone ranges, individual phone depths, and water depths at the source and at the array. These parameters are optimally estimated based on the time-domain signals alone. We assume that the bottom is sloping, consider up to four reflections (surface, bottom, surface-bottom, and bottom-surface), and allow for up to 16 phones of data. We see that nonuniqueness can still be an issue even using a great deal of time-domain data. However, these parameters are nearly uniquely estimated if additional constraints are considered, particularly on the arrays. Some of the optimal parameters are "confirmed" by simulating the signals for the predicted geometries by means of a pulse PE (RAMGEO) propagation code (courtesy of M. Collins) and then comparing those simulated signals with the observed data.

4:45

1pAO15. Modeling sound progation in shallow water including dynamic internal wave fields. Kyle M. Becker (Appl. Res. Lab., Penn. State Univ., P.O. Box 30, State College, PA 16804-0030)

Shallow-water acoustic fields are dominated by effects due to multiple interactions with the bottom. Particularly at low frequencies, the acoustic field can be modeled as a sum of normal modes that satisfy local waveguide properties. For a point source in the water column, the acoustic field is characterized in depth and range by constructive and destructive interference of the propagating modes in range. When including a range-dependent sound-speed field in the water column, such as that caused by internal waves, mode coupling occurs which can impact the observed interference pattern. In previous work, these effects have been considered by modeling acoustic propagation through a moving solibore, or by con-

sidering fixed 2D realizations of the sound-speed field. In this work, acoustic fields are generated for the case of a propagating internal wave field. In this case, the modal source function and local normal modes are constantly changing. The impact of these changes on the resulting modal interference pattern is examined for frequencies between 50 and 1000 Hz. In addition, fields are compared as would be measured on fixed versus synthetic aperture horizontal arrays. [Work supported by an ONR Ocean Acoustics Entry-Level Faculty Award.]

MONDAY AFTERNOON, 5 JUNE 2006 BALLROOM A, 1:00 TO 5:00 P.M.

Session 1pBB

1p MON. PM

Biomedical Ultrasound/Bioresponse to Vibration: Brown Tadpoles and Red Herrings II: Boiling Inertial Cavitation and Nonlinearity in High Intensity Focused Ultrasound Lesion Formation

Peter J. Kaczkowski, Chair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Invited Papers

1:00

1pBB1. Vapor bubbles and sound. Andrea Prosperetti (Dept. of Mech. Eng., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218, prosperetti@jhu.edu)

The paper focuses on the differences between the behavior of gas and vapor bubbles in a sound field. The stiffness of a gas bubble derives from the compression of the gas, while that of a vapor bubble is due to the release of latent heat and the consequent heating of the bubble contents. As a consequence, while the resonance frequency of a gas bubble is approximately inversely proportional to its radius, that of a vapor bubble varies like the inverse radius to the power 2/3. At sufficiently low frequencies, vapor bubbles have no stiffness at all. An oscillating vapor bubble may pick up by diffusion some gas dissolved in the liquid, which has a strong effect on the dynamics. These and other effects will be illustrated in the presentation. [Work supported by NASA Grant NNC05GA47G.]

1:30

1pBB2. Cavitation-enhanced ultrasound heating *in vivo*: Therapy protocols, mechanisms, and acoustic and MRI monitoring. S. D. Sokka (Philips Research North America, 345 Scarbarough Rd., Briarcliff Manor, NY 10510, sham.sokka@philips.com), N. Vykhodtseva, and K. Hynynen (Brigham and Women's Hospital, Harvard Med. School, Boston, MA 02115)

Focused ultrasound is currently being developed as a noninvasive thermal ablation technique for benign and cancerous tumors in several organ systems. Although these therapies are designed to ablate tissue purely by thermal means, cavitation can occur. These bubbles can be unpredictable in their timing and location and often interfere with thermal therapies. However, gas bubbles *in vivo* have some potential useful features for therapy. In this research, we design and test *in vivo* ultrasound exposures that induce cavitation at appropriate times and take advantage of their absorption enhancing properties while yielding reliable lesion sizes and shapes. In addition, MRI and acoustic methods to monitor and potentially control cavitation induction and the associated therapy are investigated. Finally, histology of the resulting cavitation-enhanced heating lesions is performed to assess the type of tissue damage. We demonstrate that cavitation-enhanced heating can be reliable and useful with the appropriate therapy protocol. Properly timed cavitation during thermal therapy yields larger lesions at lower powers taking advantage of bubbles' higher absorption. In addition, we show that bubble clouds that result from cavitation can block ultrasound wave propagation, providing a means to protect vital tissues during ablation of nearby pathology.

2:00

1pBB3. Nonlinear mechanisms of heating by high-intensity focused ultrasound. Vera A. Khokhlova (Dept. of Acoust., Faculty of Phys., Moscow State Univ., Moscow 119992, Russia, vera@acs366.phys.msu.ru), Michael R. Bailey, Michael S. Canney, Peter J. Kaczkowski, and Lawrence A. Crum (Univ. of Washington, Seattle, WA)

Two major nonlinear mechanisms are known to influence HIFU heating: acoustic nonlinearity and cavitation. Heating may also result in formation of boiling vapor bubbles that grow much larger than the cavitation bubbles. The relevant role of these phenomena was investigated experimentally and numerically in a gel phantom. HIFU pressure thresholds for shock formation, cavitation, and boiling were measured using a fiber-optic probe hydrophone, passive cavitation detection, ultrasound and optical imaging, and thermocouples. The KZK and Bio-heat equations were employed to simulate experimental conditions. Elevated static pressure was applied to suppress bubbles and increase the boiling temperature, thus isolating the pure effect of acoustic nonlinearity in comparison of heating between short, high-amplitude and long, low-amplitude pulses of equal average intensity. The experimental results indicated that both nonlinear mechanisms accelerated lesion production with acoustic nonlinearity responsible for the greater effect. It was observed that lesion distortion and migration was due to boiling detected in as little as 40 ms within the center of the lesion, in agreement with nonlinear acoustic simulations. These data indicate that acoustic nonlinearity and the boiling play a significant role earlier in HIFU treatments than previously anticipated. [Work supported by the NSBRI-SMS00402, RFBR, NIH FIRCA and DK43881.]

2:30

1pBB4. The multiple facets of bubble-enhanced high-intensity focused ultrasound (HIFU) lesion formation. Ronald A. Roy, Caleb H. Farny, Tianming Wu, and R. Glynn Holt (Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Exposure to high-intensity focused ultrasound can in some cases result in bubble formation. These bubbles can accelerate heating and contribute to the formation of irregularly shaped lesions. The relevant mechanical and thermal effects depend critically on several factors, such as temporal peak and temporal average acoustic intensity, pulse duration and duty cycle, bubble equilibrium sizes, tissue temperature, and tissue acoustic, rheological, and thermal properties. We will present a summary description of what physical effects matter, when, and why. Implications for HIFU treatment monitoring through the active and passive bubble detection are discussed. [Work supported by the Dept. of the Army (award No. DAMD17-02-2-0014) and the Center for Subsurface Sensing and Imaging Systems (NSF ERC Award No. EEC-9986821).]

3:00–3:15 Break

Contributed Papers

3:15

1pBB5. Measurement and modeling of nonlinear waveforms in high-intensity focused ultrasound fields. Michael S. Canney, Michael R. Bailey (Ctr. for Industrial and Medical Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), Vera A. Khokhlova, Misha A. Smagin, Oleg A. Sapozhnikov (Moscow State Univ., Moscow 119992, Russia), and Lawrence A. Crum (Univ. of Washington, Seattle, WA)

Direct measurement of HIFU fields *in situ* is important for the accurate prediction of thermal and mechanical bioeffects, as well as for the development of standards for medical systems. An experimentally validated numerical model can be an effective tool in both laboratory and clinical settings when direct measurements are not possible. Calculations with a KZK-type model and measurements with a fiberoptic probe hydrophone were employed together to characterize HIFU fields in water and in a tissue-mimicking gel. To determine the boundary conditions for simulations, the normal velocity distribution on the transducer surface was reconstructed using acoustic holography and combined with acoustic power measurements. At the focus, highly nonlinear waveforms ($+700$ and -150 bars peak pressures) were obtained both experimentally and numerically, which differed significantly from waveforms linearly extrapolated from low-amplitude results. Strongly distorted shock waveforms were localized in an axial region much smaller than the half-maximum beamwidth of the transducer excited at low level. At the highest excitation levels, the simulations predicted frequency content higher than was measurable in our configuration. Simulations also show that if these frequencies are not included, predicted heating rates are significantly lower. [Work is supported by NIH DK43881, NSBRI SMS00402, and RFBR.]

3:30

1pBB6. What is boiling during high-intensity focused ultrasound. Wayne Kreider, Michael R. Bailey, and Lawrence A. Crum (Ctr. for Industrial & Med. Ultrasound, APL, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105, wkreider@u.washington.edu)

For treatments that use high-intensity focused ultrasound (HIFU), it is important to understand the behavior of bubbles in the context of both large acoustic pressures and elevated temperatures in the surrounding medium. Based upon clinical and experimental observations, any preexisting cavitation nuclei in tissue or blood are likely to be less than 1 micron. For HIFU conditions characterized by megahertz frequencies and pressures on the order of megaPascals, gas bubbles less than a micron in radius can grow explosively. Calculations for a single, spherical bubble imply that the resulting bubble motions are significantly influenced by evaporation and condensation processes. Consequently, at both high and low ambient temperatures, HIFU-driven bubbles may best be described as gas-vapor bubbles that can exhibit rectified transfer of both heat and noncondensable gases. Moreover, increased vapor pressures associated with ambient temperatures at or above ''boiling'' may not lead to unbounded bubble growth as expected for a quasistatic bubble in a superheated medium. Instead, calculations suggest that growth of boiling bubbles can be confined. [Work supported by NIH T32 EB001650, NIH DK43881, and NSBRI SMS00402.]

3:45

1pBB7. Modeling the inertial cavitation threshold for tissue-mimicking gels seeded with laser-irradiated gold nanoparticle nuclei. Tianming Wu, Ronald A. Roy, and R. Glynn Holt (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

Previous experiments demonstrated the acoustic cavitation threshold in a tissue-mimicking phantom was lowered from 4.5 to 1 MPa by seeding the optically transparent phantom with light-absorptive gold nanoparticles and irradiating these absorbers with ns-pulsed laser light with radiant exposure less than 10 mJ/cm^2. The absorption of laser energy deposits a remarkable amount of heat into the particle and surrounding liquid. Under appropriate acoustic conditions, explosive vaporization occurs and the resultant vapor cavity expands many times its initial size and collapses inertially. A three-stage model is developed to account for prenucleation heating, nucleation and formation of the vapor cavity, and follow-up vapor bubble dynamics. The acoustic pressure and laser-radiant exposure thresholds are determined by examining the radius-time evolution of the cavity. The acoustic threshold decreases when laser exposure increases but the efficiency of light depends on its pulse length and the particle size. A strategy for choosing optimal particles to generate inertial cavitation with acoustic and optic energy inputs is suggested. [Work supported by the Dept. of the Army (Award No. DAMD17-02-2-0014) and the Center for Subsurface Sensing and Imaging Systems (NSF ERC Award No. EEC-9986821).]

4:00–5:00
Panel Discussion

Session 1pNS

Noise and Underwater Acoustics: Marine Noise Topics

Michael Bahtiarian, Chair
Noise Control Engineering, 799 Middlesex Turnpike, Billerica, MA 01821

Contributed Papers

2:00

1pNS1. The acoustic design of the research vessel SHARP. Michael Bahtiarian (Noise Control Eng., 799 Middlesex Turnpike, Billerica, MA 01821)

The R/V HUGH R. SHARP is a general purpose research vessel that is owned and operated by the University of Delaware. It has been designed to achieve the International Council for the Exploration of the Seas (ICES) underwater noise limit for Fisheries Research Vessels. This paper describes the design process, construction details, and final results with regard to the acoustic design of this ship. Both compartment airborne noise and underwater radiated noise predictions were carried using a proprietary shipboard noise prediction program. Noise control treatments that were used include: vibration isolated machinery, double-stage isolated diesel generators, damping tiles, special acoustic insulation, and floating floors. After the vessel was completed a series of at-sea tests was performed to check compliance to vibration, compartment noise, and underwater noise. Two problems were discovered with underwater noise levels. One was related to a gearbox and the other was related to the propellers. A discussion of both issues and the final results are given in the paper. [Support provided by the University of Delaware, College of Marine Studies, in particular Mr. Matthew Hawkins, Director of Marine Operations.]

2:15

1pNS2. Results of flow-field quantities from computational fluid dynamics (CFD) analysis of a flow-excited Helmholtz resonator. Paul Zoccola, Joseph Slomski, and Theodore Farabee (Carderock Div. Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817)

A flow-excited Helmholtz resonator was modeled using a hybrid RANS/LES CFD solver. Solutions were obtained for a speed where the excitation was fully coupled to the response of the cavity, and for a lower speed, where it was not. The cavity and speeds modeled were chosen to replicate a prior experiment by one of the authors. Flow-field quantities analyzed include the energy production term and the energy flow vector field. These were compared to experimental data where available. Results at resonant and nonresonant conditions were also compared.

2:30

1pNS3. Performance of unstructured hybrid Reynolds averaged Navier-Stokes/large-eddy simulation methods in simulating shear-driven cavity flows. Joseph Slomski, Paul Zoccola, Michael Ebert (Carderock Div., Naval Surface Warfare Ctr., 9500 MacArthur Blvd., West Bethesda, MD 20817), and Srinivasan Arunajatesan (Combustion Res. and Flow Technol., Inc., Pipersville, PA 18947)

Two separate, unstructured hybrid RANS/LES methods are used to simulate both resonant and nonresonant shear-driven cavity flows. The first method uses an upwind-biased discretization for the inviscid flux calculations in the governing equations, along with a nonlinear $k-e$ turbulence closure for RANS regions, and the Smagorinsky subgrid-scale closure for LES regions. The second method uses an upwind-biased discretization for the inviscid flux terms which is modified to reduce the inherently high dissipation in the associated Riemann solver when applied to cell faces not orthogonal to the flow direction. The second method uses a $k-e$ closure for RANS regions. In LES regions, the second method solves a transport equation for subgrid turbulent kinetic energy, relating this energy to a spectrum for the energy-inertial-dissipation range, which allows calculation of a less dissipative eddy viscosity. Both methods are applied to a three-dimensional, deep cavity problem, at resonant and nonresonant flow conditions. Resulting pressure-time series are compared to experimental measurements.

2:45

1pNS4. Radiated noise signature of a ship in maneuvers. Boris Vasiliev (DRDC-Atlantic, 9 Grove St., P.O. Box 1012, Dartmouth, NS B2Y 3Z7, Canada, boris.vasiliev@drdc-rddc.gc.ca)

Underwater acoustic radiated noise of ships is usually measured for constant course and speed. The measurements are often averaged over the ship's aspect to yield the radiated noise signature. Consequently, dependence of the radiated noise on the ship's aspect or maneuvers is generally unknown. This talk will present measurements of the broadband and demodulated noise signatures of two research vessels. Variation of the signatures with ship aspect, acceleration, and turn rate will be discussed.

3:00

1pNS5. Reducing underwater radiated noise from gravel islands. Jesse Spence (Noise Control Engineering, 799 Middlesex Tpke., Billerica, MA 01821, jesse@noise-control.com)

Methods for reducing underwater noise radiated from gravel islands used for oil and natural gas production have been investigated. This study focused on operations for one gravel island, the Northstar, located in Prudhoe Bay, Alaska, during standard oil production activities in open water conditions. Various transmission paths have been examined, including direct airborne transmission, vibratory paths, and other secondary paths. Direct measurements of airborne noise, vibration, and underwater noise were used to determine the prominence of each noise path. General treatment options and details are given for those noise paths deemed to be significant. [The author would like to thank the U.S. Minerals Management Service for their funding and support, and BP Exploration (Alaska) Inc. and Greeneridge Sciences Inc. for their support and assistance.]

3:15

1pNS6. Vertical directionality of noise at two shallow water sites. Timothy F. Duda, James F. Lynch, Arthur E. Newhall (Woods Hole Oceanograph. Inst., AOPE Dept., MS 11, Woods Hole, MA 02543), and Philip A. Abbot (Ocean Acoust. Services and Instumentation Systems, Inc., Lexington, MA 02421)

Noise recorded with 16-element vertical line arrays moored at two sites near 80-m water depth has been analyzed. One site is south of New England (Shelfbreak PRIMER, June 1996), the other south of China in the South China Sea (ASIAEX, May 2001). Both sites exhibited downward

refracting conditions, so that a noise notch is realizable, and both exhibit internal gravity waves. The waves were more plentiful and of larger amplitude at the South China Sea site. Sound in a 40-Hz band centered at 135 Hz was extracted and separated into vertical beams using delay and sum beamforming. At each location, noise lower than the median level exhibited a peak near zero degrees for each data set (near-horizontal arrival). Noise higher than the median level in the New England data set exhibited a 5-dB noise notch, with peak levels at about ± 15 deg. Noise higher than the median level in the South China Sea data set did not show the notch and was essentially flat between 10 and -10 deg. The results are consistent with observed fluctuating propagation between fixed sources and receivers at the sites, and with differences in coupled-mode propagation through the two internal-wave fields.

3:30

1pNS7. Vertical directionality of midfrequency surface noise in downward-refracting environments. Cathy Ann Clark (NUWC, Bldg. 1320, Rm. 457, 1176 Howell St., Newport, RI 02841)

The vertical directionality of ambient noise due to surface agitation in the midfrequency range (0.5 to 10 kHz) for a subsurface receiver has a characteristic shape, knowledge of which may enhance shallow-water operations. In general, the noise level is highest at upward-looking angles and attenuated at downward-looking angles depending on the nature of the bottom. In environments with a negative profile gradient, the noise level is also greatly reduced in a low-angle shadow zone or ''notch'' at angles around 0 deg. This paper reviews the character of vertical noise directionality by examining a small number of measured data sets and considering the underlying physical mechanisms that drive the form of the distribution. A discussion of the implications for receiving sonar systems is followed by model reproduction of the measured data sets.

MONDAY AFTERNOON, 5 JUNE 2006 ROOM 551AB, 1:00 TO 4:35 P.M.

Session 1pPA

Physical Acoustics: Celebration of the Work of Brown University

Murray S. Korman, Cochair
U.S. Naval Academy, Physics Dept., Annapolis, MD 21492

Humphrey J. Maris, Cochair
Brown Univ., Dept. of Physics, Providence, RI 02912

Stephen V. Letcher, Cochair
Univ. of Rhode Island, Dept. of Physics, Kingston, RI 02881-0817

Chair's Introduction—1:00

Invited Papers

1:05

1pPA1. A short history of physical acoustics at Brown. Robert T. Beyer (Dept. of Phys., Brown Univ., Providence, RI 02912)

Modern acoustics at Brown began with Bruce Lindsay. Lindsay had been a student of atomic theory, but switched to acoustics during and after WWII. He found strong support from the Office of Naval Research (ONR), and brought to the department Arthur Williams, Peter Westervelt, Jordan Markham, Bob Morse, Rohn Truell, Wes Nyborg, and many others. Honors in the Society among this group included three presidencies, two gold medals, and several silver awards, as well as Lindsay's role of 20 years as Editor-in-Chief. The author of this paper recently found himself described in a German publication as ''Schatzmeister'' (look it up) of the Acoustical Society of America. Rohn Truell moved to Applied Mathematics at Brown and developed a second acoustics group there, which included Charles Elbaum and Humphrey Maris. From acoustic filters to nanosecond technology, the subject moved forward. But, we were all following Bruce Lindsay.

1:35

1pPA2. Reminiscences of Brown University during the 1950s. Wesley L. Nyborg (Phys. Dept., Univ. of Vermont, Burlington, VT 05405)

After World War II, Brown University was known internationally as one of several major centers for fundamental research in physical acoustics, with faculty and students engaged in studying properties of sound fields in gases, liquids, and solids, under linear and nonlinear conditions. Students came to Brown to benefit from this atmosphere and went on to establish their own programs at

other institutions. In the Physics Department the emphasis among areas of acoustics changed with time, following nationwide trends. For a few faculty and students, there was increased activity in areas of physical acoustics which were proving to be important to the emerging fields of biophysics and biomedical engineering. These areas included ultrasound field description, ultrasound propagation, radiation pressure, and acoustic streaming, to which major contributions had been made by R. B. Lindsay, A. O. Williams, Jr., R. T. Beyer, and P. J. Westervelt, among others. In this talk, specific examples will be discussed.

1:50

1pPA3. Acoustic methods used in condensed matter physics from the mid-1950s and beyond. Charles Elbaum (Dept. of Phys., Brown Univ., Providence, RI 02912)

A brief overview will be presented of various acoustic methods that were used at Brown University to study some aspects of condensed matter physics, in the period from the mid-1950s and beyond. The main emphasis will be directed to studies of dislocation dynamics in the context of mechanical properties (elastic and plastic) of solids, to phase transitions in solids, to metallic glasses, and to properties of liquid and solid helium.

2:05

1pPA4. The Metals Research Laboratory and the development of instrumentation for acoustics research at Brown University. Bruce Chick (RITEC Inc., 60 Alhambra Rd., Ste. 5, Warwick, RI 02886)

People I worked with and instrumentation that was developed during 30+ years at the Metals Research Laboratory at Brown University, and what that original instrumentation has led to.

2:20

1pPA5. The development of picosecond ultrasonics. Humphrey J. Maris (Dept. of Phys., Brown Univ., Providence, RI 02912, humphrey_maris@brown.edu)

Brown University has been a center for acoustics research for a very long time. Over the past 20 years, efforts have been directed towards the development of a new measurement technique that makes possible experiments on very thin films and other small structures of current interest in the computer industry. This method has also made it possible to make sound attenuation measurements over the frequency range from 10 GHz up to approximately 1000 GHz. The method will be described and some examples of its application will be presented.

2:35–2:50 Break

2:50

1pPA6. Professor Peter Westervelt and the parametric array. Thomas G. Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, One Coliseum Dr., University, MS 38677) and Murray S. Korman (U. S. Naval Acad., Annapolis, MD 21402)

Work on the interaction of large-amplitude waves in fluids, leading up to the discovery of the parametric array, is mentioned, starting with Thuras, Jenkins, and O'Neil ["Extraneous frequencies generated in air carrying intense sound waves," J. Acoust. Soc. Am. **6**, 173–180 (1935)]. Westervelt's theory for the scattering of sound by sound ["Scattering of sound by sound," J. Acoust. Soc. Am. **29**, 934–935 (1957)], is reviewed. The discovery of arraylike and directivity properties of colinear waves interacting nonlinearly, as remembered by Peter Westervelt (private communication to Muir, 2005), is discussed. The formulation of the parametric array, illustrated from the original manuscript [Westervelt, P. J., "Parametric acoustic array," J. Acoust. Soc. Am. **35**, 535–537 (1963)], is discussed, expounding on its elegance and utility in opening new avenues for research and development, as well exploring its assumptions and limits of applicability. Subsequent benchmark experiments and theoretical mode relaxing the assumptions and extending the range of applicability, are mentioned. The First and Second International Symposia on Nonlinear Acoustics (1968 and 1969) are remembered, photographically, with mention of several discoveries made by the luminaries at that time. Numerous practical applications, developed over the past 45 years, are then briefly illustrated. Westervelt's mentorship with graduate students at Brown is discussed. [Work in large part, supported by the U.S. Navy.]

3:05

1pPA7. Graduate level acoustics at Brown University 1950–1955. Samuel A. Elder (U. S. Naval Acad. (Ret.), 308 Halsey Rd., Annapolis, MD 21401)

When Sam Elder considered graduate school he had in mind going into some exotic field like low-temperature physics, but when Dr. Lindsay offered him a good deal on a teaching assistantship, he could not turn it down. Then, when he started there in 1950, he began to realize that Brown was becoming known as a powerhouse of research in acoustics, so he happily changed his direction in order to study under that great man. As editor of JASA, Lindsay and the other acoustics experts he had gathered around him were at the very heart of the Society and it was an exciting time to be there.

3:20

1pPA8. Acoustics at Brown during the late 1950s and early 1960s and its impact on two of its students. John L. Butler (Image Acoust., Inc., 97 Elm St., Cohasset, MA 02025) and Stephen V. Letcher (Univ. of Rhode Island, Kingston, RI 02881)

The Brown University Physics Department researched a wide range of acoustics covering the frequency band from infrasonic to ultrasonic and linear to nonlinear compression in gases and liquids. Although JLB's research at Brown focused on the attenuation of sound in various liquids, he was mostly fascinated by the means for generating and receiving sound and still is today. SVL's research also involved measuring sound absorption in various liquids. Except for one experiment we were two nonlinear holdouts in the heyday of the field. As a means for appreciating and understanding the activities in the acoustics lab at Wilson Hall, the paths of these two students will be traced. We start from their desire to go to Brown, through their experiences and research there, as well as their interaction with the other students and professors and on to their present positions and more recent research. Some of the research will be illustrated with photos, results, and FEA animation.

3:35

1pPA9. Ah well!—The unflappable A.O. Williams, Jr.. Peter H. Rogers (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332)

A student of Art Williams in the late sixties reflects on his experiences at Brown and on the career of his mentor.

3:50

1pPA10. Parametric acoustic array and P. J. Westervelt: A tribute. Kenneth G. Foote (Woods Hole Oceanogr. Inst., 98 Water St., Woods Hole, MA 02543)

Discovery of the parametric acoustic array, that is, its principle, by P. J. Westervelt [J. Acoust. Soc. Am. **35**, 535–537 (1963)] is justly famous. Reasons are given for this claim, including its unprecedented character and its usefulness, which is witnessed by a number of operational acoustical devices whose applications span subbottom profiling, marine archaeology, mine detection, and fish swimbladder resonance absorption spectroscopy, among others. New applications continue to be found, as to backscattering by swimbladder-bearing fish for sizing and quantification. These applications may be supported and extended by current efforts to develop a standard-target calibration protocol. While the larger community has pursued such applications, while also defining limits to the performance of the parametric acoustic array, the achievement represented by the essential discovery remains undiminished. In this tribute, some comments are also made of a personal sort on Westervelt's approach to science and teaching.

4:05

1pPA11. Nonlinear interactions in the 1970s: Tribute to the work of Robert T. Beyer. Murray S. Korman (Phys. Dept., U. S. Naval Acad., Annapolis, MD 21402) and Timothy K. Stanton (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Professor Beyer was our thesis advisor and supported, motivated, and directed both T.K.S's and M.S.K's Ph.D. research in the area of nonlinear underwater acoustics at Brown University. T.K.S performed experiments in an effort to measure the nonlinear interaction of sound with noise, while M.K.S did experiments involving the scattering of sound by sound in the presence of turbulence. Along the way T.K.S designed and built a complex intensity meter and measured the growth of the nonlinearly generated sidebands with propagation distance, while M.S.K designed and built a hot-film anemometer and measured the angular dependence of the spectral broadened sum frequency versus angle. We also learned ultrasonics and nonlinear acoustics from our mentor, who wove in the works of Rayleigh, Lighthill, and Westervelt along the way. Beyer also gave us a passion for the history that goes beyond the science. In the spirit of ''once a graduate student, always a graduate student,'' and ''don't just stand there go write a thesis,'' in our tribute to Beyer we will try to tell the ''vintage story'' that is spun around Professor Beyer the scientist, author, and teacher who established B/A, and made the earliest measurements on the parametric array.

4:20

1pPA12. Study of parametric excitation of nonlinear vibrations in strings. David A. Brown and Xiang Yan (Acoust. Res. Lab. and Dept. of Phys., Adv. Technol. & Manufacturing Cntr., Univ. of Massachusetts Dartmouth, 151 Martine St., Fall River, MA 02723)

The vibration of a string apparatus can easily be modified to demonstrate and study the parametric excitation of nonlinear acoustic vibrations. As faculty at Brown University were pioneers in both the study of nonlinear acoustic phenomena and introducing clever teaching apparatus to illustrate the principles of this science, it seems fitting to present the results of our study in this special session. The direct drive inductive transducer generates an alternating force, which drives a fixed-free aluminum thin beam that is tuned at resonance with the addition of mass. The string is attached to the free end of the beam and a dual pulley system and micro-positioning rail allows for the continuous adjustment of tension. The orientation of longitudinal vibrations in the string modulates the tension and excites transverse resonance modes that match the condition of half the drive frequency. The classic results of highly bent tuning curves are illustrated in this system by adjusting the tension (sound speed) in the string such that the resonance versus wave number spectra can be observed. Thus such a system may be useful in introducing the student to the study of parametric excitation of nonlinear vibrations in simple systems.

Session 1pPP

Psychological and Physiological Acoustics: Amplitude Modulation, Binaural Hearing, Decisions, and Speech (Poster Session)

Emily Buus, Chair
Univ. of North Carolina, 1115 Bioinformatics Bldg., Chapel Hill, NC 27599

Contributed Papers

All posters will be on display from 1:00 p.m. to 5:00 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:00 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 5:00 p.m.

1pPP1. Frequency discrimination in amplitude-modulated maskers. Emily Buss, Joseph W. Hall, III, and John H. Grose (Univ. of North Carolina at Chapel Hill, 130 Mason Farm Rd., CB 7070, 1115 Bioinformatics Bldg., Chapel Hill, NC 27599, ebuss@med.unc.edu)

Under some conditions, the detection threshold for a pure tone signal masked by a bandpass Gaussian noise can be reduced by the introduction of amplitude modulation (AM) to the masker. Previous data suggest that this advantage may be reduced for discrimination tasks, such as frequency discrimination of a masked tone [Hall *et al.*, J. Acoust. Soc. Am. **102**, 1746–1754]. In the present study, frequency discrimination thresholds were estimated for a 500-Hz pure tone presented at a range of levels (3, 6, and 12 dB *re*: masked detection threshold). The masker was either a Gaussian noise or an AM noise (bandpass filtered 50–4000 Hz), with a range of square wave AM rates (10–40 Hz). Masking release was greater for 10-Hz AM than 40-Hz AM, but frequency discrimination was better for the 10 than the 40-Hz rate when measured at equal sensation levels. This result suggests that temporal resolution and integration may have important contributions to frequency discrimination measured for tones presented in modulated maskers. [Work supported by NIH NIDCD Grant R01 DC000418.]

1pPP2. Sensitivity to across-frequency detection cues in binaural hearing. Joseph W. Hall, III, Emily Buss, and John H. Grose (Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7070)

The detectability of an $S\pi$ signal in a narrow-band No masking noise can be improved by the addition of comodulated No flanking noise bands [Hall and Cokely, J. Acoust. Soc. Am. **83**, 1839–1845(1988)]. This effect is relatively unreliable, however, occurring in only about 50% of listeners. The present study examined whether the effect can be made more reliable by slightly reducing the interaural correlation of the masking noise. For such a masker, interaural correlation fluctuates as a function of time. For the stimuli used here, the on-signal and the comodulated flanking bands shared the same pattern of interaural correlation as a function of time. The rationale was that the flanking bands could potentially be used as templates to help the auditory system distinguish the ongoing interaural differences of the slightly decorrelated on-signal noise band from the interaural differences introduced by the $S\pi$ signal. Results from all six of the listeners tested showed improved $S\pi$ detection in conditions where the on-signal and flanking bands had the same pattern of interaural correlation. This effect can be considered a form of binaural profile analysis wherein signal detection is aided by an across-frequency analysis of binaural cues. [Work supported by NIH NIDCD R01 DC00397.]

1pPP3. The contribution of envelope and carrier to cues for tone-in-noise detection in narrow-band reproducible noises. Sean A. Davidson (Dept. of Biomed. and Chemical Eng., Inst. for Sensory Res., Syracuse Univ., 621 Skytop Rd., Syracuse, NY, 13244), Robert H. Gilkey (Wright State Univ., Dayton, OH 45435), H. Steven Colburn (Boston Univ., Boston, MA 02215), and Laurel H. Carney (Syracuse Univ., Syracuse, NY 13244)

To elucidate the roles of envelope and carrier cues in detection of a 500-Hz tone in 50-Hz bandwidth noise under N_0S_π conditions, hit- and false-alarm rates were collected for 4 sets of reproducible stimuli. The envelopes and carriers from two sets of stimuli (each with 25 signal-plus-noise and 25 noise-alone waveforms) were extracted and recombined to form 2 new sets (each having the carriers from one of the original sets and the envelopes from the other). Noises were preselected to produce minimal spectral splatter in the recombination processes. Preliminary results for signal-plus-noise trials suggest that carrier cues (e.g., temporal fine structure) are dominant over envelope cues (i.e., response patterns were more similar between sets with the same carriers but different envelopes, than between sets with the same envelopes but different carriers); however, carrier cues cannot completely explain the data. Results for noise-alone trials suggest that both envelope cues and carrier cues contribute to listener responses. Further analyses will explore potential interactions between monaural and binaural stimulus properties, as well as interactions between interaural envelope cues and carrier cues. [Supported by NIH DC007798(SAD), DC001641(LHC), and DC00100(HSC).]

1pPP4. Responses of cochlear nucleus units recorded from chinchillas trained to discriminate iterated rippled noise from flat-spectrum noise. William P. Shofner (Dept. of Speech and Hearing Sci., Indiana Univ., 200 S. Jordan Ave., Bloomington, IN 47405)

Behavioral responses obtained from chinchillas using a stimulus generalization paradigm show that the perception of pitch strength is dependent on the stimuli used for training. When animals are trained to discriminate a cosine-phase harmonic tone complex (COS) from a flat-spectrum, wideband noise (WBN), generalization gradients indicate that animals rely primarily on temporal information in the envelope. Gradients from animals retrained to discriminate infinitely iterated rippled noise (IIRN) from WBN indicate animals rely more on information in the fine structure. Thus, chinchillas learn to use the information in the fine structure when trained with an appropriate stimulus. In order to begin to examine whether there exists a neural correlate of this learning at the level of the cochlear nucleus, single-unit responses to IIRNs and WBN were recorded from three anesthetized chinchillas that had received considerable behavioral training in discriminating IIRNs from WBN. Preliminary data were ob-

tained for 19 single units consisting of primarylike, chopper, and onset response types. All-order interspike-interval histograms obtained for IIRNs and WBN were qualitatively similar to those obtained from naive, untrained animals for each unit type. Quantitative comparisons among units recorded from trained and untrained animals will be presented. [Work supported by NIDCD R01 DC005596.]

1pPP5. On the rate of decay of forward masking and its relation to cochlear compression. Enrique A. Lopez-Poveda and Ana Alves-Pinto (Instituto de Neurociencias de Castilla y Leon, Universidad de Salamanca, Avda. Alfonso X El Sabio s/n, 37007 Salmanca, Spain)

Recent psychoacoustic studies [e.g., Lopez-Poveda *et al.*, J. Acoust. Soc. Am. **113**, 951 (2003)] have suggested that cochlear compression extends to a wider range of frequencies in the apex than in the base of the cochlea. This conclusion was drawn from the analysis of the slopes of temporal masking curves (TMCs) on the assumption that forward masking decays at the same rate for all probe and masker frequencies. This study aims to verify the conclusion stated above using a different assumption. TMCs for normal hearing listeners were measured for probe frequencies (Fp) of 500 and 4000 Hz and for masker frequencies of 0.4, 0.55, and 1.0 times the probe frequency. Measurements were made for probes of 9- and 15-dB sensation level. The assumption was that given a 6-dB increase in probe level, linear cochlear responses to the maskers should lead to a 6-dB vertical shift of the corresponding TMCs, while compressive responses should lead to bigger shifts. Results are consistent with compressive cochlear responses to the 0.4Fp and 0.55Fp maskers at 500 Hz but linear responses at 4000 Hz, in line with the conclusions from earlier studies. [Work supported by the Spanish FIS (PI020343, G03/203), MEC (CIT-390000-2005-4), and FEDER.]

1pPP6. Amplitude modulation detection by infants and adults. Lynne Werner (Dept. Speech & Hearing Sci., Univ. of Washington, 1417 N.E. 42nd St., Seattle, WA 98105-6246)

Infants' and adults' intensity processing was examined by estimating the modulation depth required to detect amplitude modulation in a broadband noise. The infants were either 3 or 6 months old. An observer-based response method was used. Modulation depth was varied adaptively to estimate thresholds. The noise was presented continuously at 70 dB SPL. Listeners detected 4-Hz modulation presented for 4 s. Adults thresholds were similar to those reported in the literature for trained listeners. Infants' thresholds were substantially worse than those of adults, but 3- and 6-month-olds did not differ. This result is consistent with previous reports that 6–9-month-olds are immature in intensity discrimination. However, it is surprising that little change in performance occurs between 3 and 6 months, because infants of these ages differ in other measures of intensity processing. [Work supported by NIDCD.]

1pPP7. Across-channel gap detection in cochlear implant listeners. John H. Grose, Emily Buss, and Joseph W. Hall, III (Dept. Otolaryngol.-HNS, Univ. N. Carolina at Chapel Hill, 1115 Bioinformatics CB# 7070, Chapel Hill, NC 27599-7070)

In normal hearing, gap detection thresholds are elevated when the leading and trailing markers of a gap differ widely in frequency or are presented to different ears; thresholds are also elevated by the introduction of amplitude modulation (AM) to the markers. This study assessed whether similar effects can be observed in a group of high-performing, MED-EL COMBI 40+ cochlear implant users, a subset of whom are bilaterally implanted. In experiment 1, markers were equal-amplitude 3-kHz pulse trains of random duration (40–80 ms) delivered through a research interface. Thresholds were low (<1 to 6 ms) when the two markers were presented on a single electrode channel and increased when the markers were on different channels within an ear or across ears. In experiment 2, markers were AM pulse trains of random duration (100–150 ms). These were generated by multiplying a 3-kHz, equal-amplitude pulse train with the envelope of a 50- or a 300-Hz bandpass filtered Gaussian noise; the envelope was either uncompressed or compressed with an exponent of 0.2. Thresholds were elevated by AM at both rates, with larger effects for the uncompressed envelope. These patterns of results are similar to those observed in normal hearing. [Work supported by NIDCD 5R01DC1507.]

1pPP8. Effects of aging and masker-level variability on backward-masked frequency discrimination. Blas Espinoza-Varas (Commun. Sci. & Disord., OU Health Sci. Ctr., Oklahoma City, OK 73190), Hyun-sook Jang (Hallym Univ., Chuncheon, South Korea), and Praveen Jajoria (OU Health Sci. Ctr., Oklahoma City, OK 73190)

Backward-masked frequency discrimination thresholds (FDTs) were measured to probe the ability to ignore irrelevant auditory information in young and older adults. For 1500-Hz, 40- or 60-ms, 75-dB-SPL target tones, FDTs were measured unmasked and in three backward-masked conditions in which a 1500-Hz, 80-ms, 75-dB-SPL masking tone followed the target with 20–240-ms ISI. In two masked conditions, there was within-trial irrelevant variability in masker level, with negative or positive polarity (i.e., 75 ± 5 or 10 dB SPL) but not in the third. The no-variability condition yielded more backward masking in older than in young adults. Irrespective of polarity, masker-level variability elevated the FDTs of both age groups, but relative to those of the young, the elders' elevations were larger and did not decrease with training; elders had difficulty learning to ignore irrelevant variability. In young adults, the decrease in FDTs with training tended to be steeper with negative than with positive level variability. Memory capacity and long-term memory constraints seem to limit the elders' ability to ignore irrelevant information.

1pPP9. Detection thresholds for amplitude modulation of noise with simultaneous modulating tone. Jonathan Pincas and Philip J. B. Jackson (CVSSP, Dept. Electron. and Elec. Eng., Univ. of Surrey, Guildford, GU2 7XH, UK, jon@pincas.co.uk)

In speech, the combination of amplitude-modulated (AM) noise and a low-frequency periodic source characterizes voiced fricative sounds. To investigate the modulation's contribution to fricative auditory quality, AM white noise, with simultaneous sinusoidal component at the modulating frequency, provided stimuli for perceptual tests. Two AM detection-threshold experiments were conducted to establish the effect of varying the relative amplitude and phase of the tone. In the first experiment, tone and noise stimuli were separated within each trial by short pauses; in the second, the tone played continuously throughout the trial. AM detection thresholds when the tone was absent were similar to previously published thresholds for AM noise with no tone (averaging approximately -22 dB). The effect of the quieter tones (10–30 dB tone-to-noise ratio, TNR) was dependent on test condition. In the broken-tone experiment, a slight increment in threshold occurred for all relative phases. The constant-tone experiment revealed impaired detection when tone and noise modulation were in phase, and the converse when in antiphase (5-dB difference in threshold at 30-dB TNR). As TNR increased further (40–50-dB TNR), thresholds rose in all cases. Future work is needed to evaluate how this detectable AM percept influences quality and categorization of fricatives.

1pPP10. Static and dynamic spectro-temporal gap detection. David A. Eddins (Dept. of Otolaryngol., Univ. of Rochester, 2365 S. Clinton Ave., Ste. 200, Rochester, NY 14618) and Christine E. Carlson (Univ. at Buffalo, Buffalo, NY 14514)

Frequency-dependent measures of temporal processing are often based on temporal gap detection measured using sinusoids or bandpass noise. Some natural sounds, such as speech, are better characterized as relatively broadband sounds with low-amplitude portions restricted to narrow frequency regions. The present study reports the perception of analogous

stimuli: static and dynamic spectro-temporal gaps embedded in broadband noise. In the first experiment, the temporal gap created a static spectral notch, the duration of which was the independent variable (gap duration in ms) and the width of which was fixed, in separate conditions being either 800 Hz or equal to one or three ERBs (equivalent rectangular bandwidth). The spectral notch was centered at either 800, 1600, 3200, or 6400 Hz. Gap onset relative to stimulus onset varied from 10 to 320 ms. In the second experiment, the temporal gap created a dynamic spectral notch, the lower and upper cutoff frequencies of which varied over the duration of the temporal gap much like a frequency sweep. Experimental parameters included the bandwidth of the notch in octaves and notch center frequency at gap onset and gap offset. The data will be considered in terms of temporal resolution, frequency resolution, and across-channel processing.

1pPP11. Psychophysical and physiological assessment of the representation of high-frequency spectral notches in the auditory nerve. Ana Alves-Pinto (Instituto de Neurociencias de Castilla y Len, Universidad de Salamanca, Avda. Alfonso X El Sabio, 37007 Salamanca, Spain), Alan R. Palmer (MRC Inst. of Hearing Res., Nottingham NG7 2RD, UK), and Enrique A. Lopez-Poveda (Universidad de Salamanca, 37007 Salamanca, Spain)

High-frequency spectral notches such as those generated by the pinna are useful for sound localization. Because their center frequency (6–8 kHz) is beyond the cutoff of phase locking, the common view is that the internal representation of such notches does not involve the timing of action potentials and that, instead, they must be exclusively represented in terms of auditory nerve rate profiles. Recently, we [A. Alves-Pinto and E. A. Lopez-Poveda, J. Acoust. Soc. Am. **118**, 2548 (2005)] have shown that the threshold notch depth for discriminating between a flat-spectrum noise and a noise with a square spectral notch centered at 8 kHz varies nonmonotonically with increasing sound level, and that discrimination is worst for levels around 80 dB SPL. Here, we show that the information in the auditory nerve rate profiles is insufficient to account for this psychophysical result, and that temporal information is almost certainly used for discriminating between noise bursts with and without high-frequency spectral notches.

1pPP12. Spatiotemporal coding based on the responses of individual auditory-nerve fibers. Michael G. Heinz (Dept. of Speech, Lang., Hearing Sci. and Biomed. Eng., Purdue Univ., 500 Oval Dr., West Lafayette, IN 47907, mheinz@purdue.edu)

Spatiotemporal coding has been suggested to provide a robust representation of vowel spectra, pitch, and tones in noise. However, much of the quantitative support has come from modeling studies because of the inherent variability in physiological population studies. A spectro-temporal manipulation procedure (STMP) was developed to simulate spatiotemporal patterns from the responses of individual auditory-nerve (AN) fibers to frequency-shifted stimuli with features placed around the fibers' best frequency. Individual AN fiber responses from anesthetized cats were measured for vowellike stimuli as a function of vowel and background-noise level. Similar to previous population studies, the STMP results show that spatiotemporal coding of vowel spectra (quantified using ALSR) is more robust than average-rate coding as vowel or background-noise level increases. The STMP results also suggest that the response phase transition near low-frequency formants may not be sharp enough to provide as robust encoding as previously suggested. This method of simulating spatiotemporal patterns from individual AN fiber responses may be particularly useful for studying the effects of sensorineural hearing loss because it is particularly difficult to obtain large populations of responses from animals with hearing loss. Preliminary data have been collected from cats with noise-induced hearing loss. [Work supported by NIH/NIDCD R03DC007348.]

1pPP13. Individual differences in dynamic binaural-masking-level differences at 500 Hz in listeners older than 50 years. Michael A. Akeroyd (MRC Inst. of Hearing Res., Scottish Section, Glasgow Royal Infirmary, 16 Alexandra Parade, Glasgow G31 2ER, UK)

One demonstration of binaural sluggishness is that the binaural masking level difference (BMLD) for a tone burst is strongly dependent upon the rate of modulation of the interaural correlation of a masking noise [D. W. Grantham and F. L. Wightman, "Detectibility of a pulsed tone in the presence of a masker with time-varying interaural correlation;" J. Acoust. Soc. Am. /bf 65, 1509–1517 (1979)]. This paradigm was used to measure such dynamic BMLDs for a group of 18 listeners, aged 50–78 and with low-frequency hearing losses of 7–43 dB. The signal was a 20-ms, 500-Hz tone placed in the temporal center of a 500-ms, 100–2000-Hz, 63-dB overall-level noise. The interaural correlation of the noise was either fixed at 1.0 (static conditions) or 100% modulated (dynamic conditions). As expected, the dynamic BMLD reduced as the modulation rate increased; the across-listener means were 11, 8, 6, 2, 0, and 0 dB at rates of 0.5, 1, 2, 4, 8, and 16 Hz. But the individual differences were substantial: e.g., for the 11 listeners who showed a static BMLD of 15 dB or more, their 2-Hz dynamic BMLDs varied between 2 and 10 dB. Those listeners with poor dynamic thresholds (and so more sluggish than the others) may have increased difficulties in detecting sounds in changing spatial environments.

1pPP14. The role of monaural spectral cues in modeling the precedence effect. Pavel Zahorik (Dept. of Psychol. and Brain Sci., Univ. of Louisville, Louisville, KY 40292, pavel.zahorik@louisville.edu)

Various evidence suggests that the precedence effect may generalize to spatial locations of sound source and echo that minimize binaural information available to the listener. In these situations, it is likely that monaural spectral cues to the direction of the source+echo composite. Here a model of directional sound localization that incorporates both binaural temporal cues and monaural spectral cues is described and evaluated. The binaural portion of the model relies on interaural cross-correlation to extract binaural cues to lateral direction. The monaural spectral portion uses a spectral correlation metric to evaluate the similarity of the incoming sound spectrum with a catalog of spectral shapes resulting from a listener's own head-related transfer functions (HRTFs) to known source directions. The model was used to predict sound localization performance of seven human listeners in 16 spatial configurations of source and echo (presented using virtual auditory space techniques with individualized HRTFs), including six configurations where binaural information was minimized (four on the median plane). In general, precedence effects that relied primarily on spectral cue information were considerably less predictable than those that relied on binaural information. [Work supported by NIDCD.]

1pPP15. Interaural correlation and loudness. Barrie A. Edmonds and John F. Culling (School of Psych., Tower Bldg., Park Pl. Cardiff Univ., Cardiff, CF10 3AT, UK)

When the same sound is presented to both ears rather than to just one, a phenomenon called binaural summation occurs, in which the overall loudness is greater than when either ear is individually stimulated. The present investigation focused on the effect of interaural correlation on this phenomenon. An adaptive matching paradigm was used to measure the difference in intensity between monotic, diotic, uncorrelated, and anticorrelated bands of noise needed to equate them in loudness. The results for 460–540-Hz bands showed (1) consistent with binaural summation, diotic noise was matched to monotic noise that was 4.6 dB more intense and (2) uncorrelated noise was matched to diotic noise that was 2.0 dB more intense. Anticorrelated and diotic noises did not differ significantly in intensity when matched in loudness. Similar results were observed for 100–900-Hz bands: (1) was extended to 6.5 dB and (2) remained at 2.0 dB, but

anticorrelated noise was now matched to diotic noise that was 1.0 dB more intense. For 100–5000-Hz bands results showed (1) was extended to 5.5 dB, (2) was reduced to 1.0 dB, and anticorrelated noise was now matched to diotic noise that was 0.5 dB more intense.

1pPP16. The inclusion of time-varyng jitter in the equalization-cancellation model of binaural processing. Rui Wan, N. I. Durlach, and H. Steven Colburn (Hearing Res. Ctr. and Dept. of Biomed. Eng., Boston Univ., 44 Cummington St., Boston, MA 02215)

A more realistic version of the analytically defined equalization-cancellation(EC) model of binaural processing can be simulated in the time domain by introducing time-varying jitters, both in time and amplitude. Although the classic EC model can predict a wide range of binaural-masking-level-difference (BMLD) data in binaural detection, it is more reasonable physiologically to assume that internal noise is time varying, as opposed to the original assumption that the internal time and amplitude errors are different from waveform to waveform but stay constant within a waveform. In the current study, the BMLD predictions for both the analytical model and the simulation are compared for the representative conditions(N_oS_π, $N_\pi S_o$, N_oS_m). It is also proved theoretically that the simulation results converge to the analytical prediction in the $N_{\tau_n}S_{\tau_s}$ case. Differences between the EC model and the simulation are discussed. Applying this simulation to speech perception based on the speech intelligibility index(SII) in anechoic environments predicts binaural advantages consistent with results described in the literature (Zurek,1993). The simulation is also being applied to the speech-transmission-index (STI) prediction in complex environments. [Work supported by NIDCD R01 DC00100.]

1pPP17. Effects of correlated and uncorrelated interfering noise on interaural time delay (ITD) detection. Andrew J. Kolarik and John F. Culling (School of Psych., Cardiff Univ., Tower Bldg., Park Pl., Cardiff, CF10 3AT, UK, kolarika@cf.ac.uk)

Results are reported from two binaural discrimination experiments. The first measured ITD discrimination threshold for broadband noise in the presence of interaurally correlated interfering noise. Six listeners performed an adaptive 2I-2AFC discrimination task with 100-ms stimuli composed of delayed and diotic noises (0–3 kHz) mixed in various proportions (1, 0.75, 0.563, 0.422, 0.316, 0.237, and 0.178). A regression analysis showed that thresholds doubled for every halving of the proportion of delayed noise power in the stimulus. The second experiment measured ITD discrimination thresholds for a range of interaural correlations obtained by mixing correlated and uncorrelated noise before applying the delay. Three listeners performed 2I-2AFC discrimination tasks at correlations of 1, 0.75, 0.5, 0.25, 0.2, 0.15, 0.1, and 0.05, and durations of 100, 500, and 1000 ms. Seventy-one percent of the threshold were obtained using psychometric functions. Thresholds more than doubled for each halving of the correlation at all three stimuli durations tested.

1pPP18. Monaural and binaural cues to signal detection. John F. Culling (School of Psych., Cardiff Univ., Tower Bldg., Park Pl., Cardiff, CF10 3AT, UK)

When a signal is added to random noise in the NoSπ configuration it has two effects on the stimulus at the signal frequency: the correlation is reduced and the energy is increased. Both changes may contribute to the detection and discrimination of signals. A signal-level discrimination task (2I-FC) was used to measure the strength of each cue and their combination across a wide range of signal-to-noise ratios (-20–0 dB). The signal was a band of noise (462–539 Hz). The masker was a broadband noise (10–3000 Hz) of fixed sound level. Conditions were NoSπ, NoSo, and correlation manipulation within the target band equivalent to NoSπ. At low signal-to-noise ratios, performance was determined by the correlation change, but at higher signal levels both cues became ustable and discrimination was better in NoSπ than in the correlation-manipulation condition. Analysis using signal-detection theory indicated that these data could not be explained by assuming the two cues were independent.

1pPP19. Informational masking: Effects of similarity and uncertainty for detection of a high-frequency amplitude-modulated tone. Frederick J. Gallun, Christine R. Mason, and Gerald Kidd, Jr. (Dept. of Speech, Lang. and Hearing Sci. and Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215)

Detection of a 3-kHz target tone with 125-Hz amplitude modulation (AM) was measured in the presence of maskers that varied in their similarity to the target. Spectro-temporal similarity was varied in that maskers were tones, narrow-band noises, or AM tones. Similarity in terms of interaural configuration was varied by presenting maskers diotically while target was presented (1) diotically, (2) monotically, or (3) with its envelope delayed at one ear. All maskers had four narrow-band components with center frequencies (CFs) between 1 and 5 kHz. Masker CFs were varied on every block, trial, or interval of a 2IFC procedure, resulting in three levels of masker-frequency uncertainty. As expected, the spectro-temporally similar AM maskers caused greater masking than pure-tone or noise maskers. Surprisingly (and at variance with previous results), varying similarity by interaural configuration had no effect. Increasing masker uncertainty resulted in more masking only for the AM masker. These results suggest that (1) the effects of uncertainty are most pronounced in conditions where maskers and targets are similar and (2) more study is needed of the factors responsible for binaural release from informational masking. [Work supported by NIH/NIDCD and AFOSR.]

1pPP20. Improving listener decision strategy in the perception of complex auditory events. Jesse Flint (Psych. Dept., Binghamton Univ., PO Box 6000, Binghamton, NY 13902), Richard Pastore, and Jeremy Gaston (Binghamton Univ., Binghamton, NY 13902)

Natural auditory events are complex. In perceiving these events, listeners tend to use various, and not necessarily optimal, samplings of information. Prior research examined static and dynamic source and sound characteristics for a posture contrast (upright versus stooped) in human gait, identifying several acoustic cues that sometimes differentiate the posture contrast (e.g., pace, spectral slope and amplitude), but with no one acoustic cue always doing so. Successful listener decision strategy thus needs to be a complex process of attending to and evaluating several cues in a flexible manner. With a long-term goal of determining how to improve listening skills, the current research examines ways to enhance attention to and decisions about specific acoustic cues to the posture distinction. For the stimulus property of pace, both information (or knowledge) of cue relevance and experience making relative decisions about the cue, as opposed to experience with sample stimuli, significantly improve listener decision strategy in making the posture judgment. The current study provides a similar evaluation for the acoustic cue of spectral slope. [Research supported by grant from NSF.]

1pPP21. Effect of decision weights and internal noise on the growth of d' with N. Erica Lenhart and Robert A. Lutfi (Dept. of Communicative Disord., Univ. of Wisconsin, Madison, WI 53706)

A general finding of psychoacoustic studies is that detectability d' of a noisy signal grows less than optimally with the number N of independent observations of the signal. Traditional accounts attribute the nonoptimal growth to a form of central internal noise that accumulates with each observation. A more recent interpretation, however, implicates an upper limit on the number of observations that listeners can process at any one time in a given stimulus. The latter account was clearly favored in the present study by the results of a COSS analysis [Berg "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158 (1990)] applied to the detectability of level increments added

to simultaneous and sequential N-tone patterns. Estimates of decision weights showed a predominant reliance of listeners on no more than 2–3 tones in each pattern independent of N. Measures of weighting efficiency systematically decreased with N while noise efficiency remained constant across N. The results are discussed in terms of their implications for the interpretation of early experiments on the effects of the bandwidth-duration product on signal detection. [Research supported by NIDCD Grant 5R01DC1262-14.]

1pPP22. Individual differences in source identification for synthesized impact sounds. Robert A. Lutfi (Dept. of Communicative Disord., Univ. of Wisconsin, Madison, WI, 53706) and Raman Arora (Univ. of Wisconsin, Madison, WI, 53706)

Impact sounds were synthesized according to standard textbook equations given for the motion of freely vibrating bars, membranes, and plates. In the two-interval, forced-choice procedure highly practiced listeners identified from these sounds various physical attributes of these sources and the manner in which they were driven to vibrate. A COSS analysis of the trial-by-trial data [Berg, "Observer efficiency and weights in a multiple observation task," J. Acoust. Soc. Am. **88**, 149–158 (1990)] revealed large differences across listeners in the acoustic cues used within each identification task, but similarity in the cues used within listeners across tasks. Only in conditions in which the acoustic information for identification was highly constrained did listeners use identical cues. Despite the large individual differences in listening strategy, identification accuracy was similar across listeners. Where there were differences in identification accuracy the differences appeared large, related to differences in internal noise and not listening strategy. [Research supported by NIDCD Grant 5R01DC006875-01.]

1pPP23. Using multidimensional scaling to determine distances in the perceptual space of rolling sounds. Christophe Stoelinga, Dik Hermes (Technische Universiteit Eindhoven, Den Dolech 2, P.O. Box 513, 5600MB Eindhoven, The Netherlands), and Armin Kohlrausch (Philips Res. Labs. Eindhoven, 5656AA Eindhoven, The Netherlands)

We recorded sounds generated by a ball rolling over a wooden plate. Ball diameter, rolling speed, and plate thickness were varied. In a difference-scaling experiment, the important dimensions of the space spanned by these sounds were determined. All sounds were combined in pairs and subjects were asked to quantify the difference they heard in each pair. In this way all possible perceptual differences could play a role in the difference scalings without a need to interpret these differences in terms of underlying object properties such as size or speed. Using multidimensional scaling, a good fit for the scaled differences was obtained with a minimum of three dimensions. None of the dimensions corresponded directly with one of the three mechanical parameters varied in the rolling sounds. Differences in speed lead to a smaller distance in the thus-obtained perceptual space than differences in size, for the ranges we used. This corresponds to results from experiments where the subjects were directly asked to judge size and speed differences using the same sounds. Remarkably, the plate thickness resulted in large distances in perceptual space, while this mechanical property is poorly recognized when participants are directly asked for it.

1pPP24. Optimizing information-transfer rates for human communication. Charlotte M. Reed, Nathaniel I. Durlach (Res. Lab. of Electron., MIT, Cambridge, MA 02139, cmreed@mit.edu), and Hong Z. Tan (Purdue Univ., West Lafayette, IN 47907)

This research is concerned with investigating the factors that contribute to optimizing information-transfer (IT) rate in humans. This issue has important ramifications for the design of human-machine interfaces in a wide variety of applications, including virtual environment and teleoperator systems as well as sensory aids for persons with impaired vision and/or hearing. Based on the results of several early studies from the 1950s, a general rule of thumb has arisen in the literature which suggests that IT rate is dependent primarily on the stimulus delivery rate and is optimized for presentation rates of 2–3 items/second. Thus, the key to maximizing IT rate is to maximize the information in the stimulus set. Recent data obtained with multidimensional tactual signals, however, appear to be at odds with these conclusions. In particular, these current results suggest that optimal delivery rate varies with stimulus information to yield a constant peak IT rate that depends on the degree of familiarity and training with a particular stimulus set. We discuss factors that may be responsible for the discrepancies in results across studies including procedural differences, training issues, and stimulus-response compatibility. [Research supported by Grant No. R01-DC00126 from NIDCD, NIH.]

1pPP25. A miniature, low-cost, user-friendly personal sound-level dosimeter. Harold A. Cheyne, II and Thomas E. von Wiegand (Sensimetrics Corp., 48 Grove St., Ste. 305, Somerville, MA 02144-2500, harold@sens.com)

The objective of this research is to develop a miniature, low-cost, user-friendly personal sound-level dosimeter that can meet the ANSI S1.25-1991 standard. This dosimeter is intended for people whose work-related or recreation-related activities expose them to noises that cause hearing loss. In contrast to typical dosimeters that are controlled by a manager or clinician, the wearer of the miniature dosimeter will also control it and receive its data output to better appreciate his or her sound exposure dose. To significantly miniaturize the dosimeter and lower its cost as compared to typical dosimeters, it will have a reduced feature set, simplified signal processing, an inexpensive microphone capsule, and a microcontroller rather than a digital signal processor. Preliminary tests, using a Texas Instruments embedded controller and an LCD display, suggest that approximations to the standard dosimeter function blocks can be made while maintaining the ANSI tolerances. Tests with inexpensive microphone capsules show acceptable drift due to temperature and humidity for a dosimeter. Following further testing for compliance with the ANSI standard, a limited set of field tests is planned, with wearers using both the miniature dosimeter and a commercial dosimeter, to compare their dosimetry performance.

1pPP26. Intimacy in performance venues: Towards a socioecological methodology. Bobby Gibbs, II (Rensselaer Polytechnic Inst., 110 8th St., Troy, NY 12180), Jonas Braasch, and T. Krueger (McGill Univ., Montreal, QC, Canada)

Intimacy can be subjectively defined as the feeling of closeness to performance. Intimacy has been widely correlated with the initial time delay gap, the delay between the direct sound and early reflections in a venue [Beranek, *Concert and Opera Halls: How They Sound* (Acoustical Society of America, New York, 1996)]. Contemporary research suggests that visual stimuli have considerable bearing upon intimacy perception [e.g., Cabrera and Nguyen, J. Acoust. Soc. Am. **116**, 2475 (2004)]. Building upon paradigms of ecological psychology, the author hypothesizes that social affordances, perceived offerings of social propinquity in a venue, will significantly enhance the auditory impression of intimacy. A socioecological approach to weigh the hypothesis is proposed wherein auditory perception is evaluated in the context of both the physical and the social environment. The methodology will employ interviews, ethnographic analyses, field studies, and laboratory tests. The research results will likely have immediate ramifications for the tools of visualization and auralization, its acoustical analogue. The study has broader ramifications for the design process, engendering the possibility of more humane design of artistic listening spaces through understanding reciprocal relationships between physical, acoustical, and social domains. [Work supported by ASA Minority Fellowship and RPI Humanities, Arts, Architecture, and Social Sciences Fellowship.]

1pPP27. How do task instructions influence speech-on-speech masking? Karen S. Helfer and Richard L. Freyman (Dept. of Commun. Disord., Univ. of Massachusetts, 7 Arnold House, Amherst, MA 01003-0410)

This study examined how the nature of instructions to the listener affects masking in speech perception. A new set of stimuli (the TVM sentences) was recorded from three male talkers. These sentences, which are similar to the widely used CRM sentences but are open-set and therefore more difficult, have the format *NAME discussed the *** and the *** today*, where *NAME* is Theo, Victor, or Michael and *** is a one- or two-syllable noun used for scoring (e.g., *Theo discussed the airplane and the textbook today*). On a given trial, three sentences were presented simultaneously (one beginning with each of the three names), each spoken by a different talker. Subjects participated in 6 Name conditions (where they were told to repeat the sentence beginning with a certain name) and in 12 voice conditions (where they were told to listen for a certain voice with which they had been familiarized). Stimuli were presented in both spatially coincident and spatially separated conditions at 3 S/N ratios. Analyses of differences in percent-correct performance among the listening conditions and of error patterns (e.g., errors related to confusion between target voice and masking voices versus errors related to energetic masking) will be shown. [Work supported by NIH DC01625.]

1pPP28. Segregation of vowel sequences having spectral cues reduced using a noise-band vocoder. Etienne Gaudrain, Nicolas Grimault (UMR CNRS, 5020 50 av. T. Garnier 69366 Lyon Cedex 07, France), Eric W. Healy (Univ. of South Carolina, Columbia, SC 29208), and Jean-Christophe Bera (Inserm UMR 556, Lyon, France)

Previous studies have shown that a reduction in spectral cues can be associated with a deficit in pitch-based sequential segregation. The difficulties experienced by hearing-impaired listeners in cocktail party situations may then be explained in part by a segregation deficit due to the broadening of auditory filters. The reduced speech-recognition abilities of cochlear implant users in noisy environments could also potentially be related to a loss of spectral cues leading to a segregation deficit. The current study was designed to test the ability of listener's hearing cochlear implant simulations to segregate sequences of vowels having alternating $F0$ values simulating concurrent voices. The vowels were either unprocessed or processed using noise vocoders having limited numbers of channels. Performance on the segregation task was evaluated using a paradigm in which listeners judged the temporal order of items within the sequence. By relating systematic reductions in the number of channels to performance on the segregations task, the current results have the potential to clarify potential limitations placed upon auditory scene analysis by cochlear implant processing.

1pPP29. Acoustic cues for improved speech recognition in combined acoustic and electric hearing. Ying-Yee Kong and Robert P. Carlyon (MRC—Cognition & Brain Sci. Unit, 15 Chaucer Rd., Cambridge, CB2 2EF, England)

Speech recognition in noise improves with combined acoustic and electric hearing compared to electric hearing alone. Here, we investigate the contribution of four low-frequency hearing cues (fundamental frequency ($F0$), first formant ($F1$), voicing, and glimpsing) to speech recognition in combined hearing. Normal-hearing listeners heard vocoded speech in one ear and low-pass (LP) filtered speech in the other. Three conditions (vocode alone, LP alone, and combined) were tested. Target speech with an average $F0$ of 120 Hz was mixed with a time-reversed masker sentence (average $F0 = 172$ Hz) at three SNRs (5,10,15 dB). The LP speech aided performance at all three SNRs. $F1$ cues were then removed by replacing the LP speech with an LP equal-amplitude harmonic complex with the same $F0$ contour as the target speech, and an amplitude that was modulated by that of the target's voiced portions. The benefits of combined hearing disappeared at 10- and 15-dB SNR, but persisted at SNR = 5 dB. The same happened when, additionally, $F0$ cues were removed by fixing the $F0$ of the harmonic complex at 150 Hz. The results are consistent with a role for the $F1$ cue and of voicing and/or glimpsing cues, but not with a combination of $F0$ information between the two ears.

1pPP30. Listening in the valleys with cochlear implants: Speech perception versus psychophysics. Monita Chatterjee and Shu-Chen Peng (Hearing and Speech Sci., 0100 LeFrak Hall, Univ. of Maryland, College Park, MD 20742)

Unlike listeners with normal hearing who can benefit from masking release with modulated noise, cochlear implant users show little release from masking when identifying words in sentences. Little is known about basic mechanisms underlying masking release in speech perception. In the present study, phoneme recognition was evaluated in implant users with tense vowels (/a, i, u, e, o/) and plosives (/p, t, k, b, d, g/) in quiet, steady-state noise and modulated noise at different signal-to-noise ratios (SNRs). In parallel psychophysical experiments, both modulation sensitivity and electrode discrimination were measured in the presence of on- and off-channel maskers, with different fluctuating envelopes. Results from three implant users indicated that unlike the previous findings with sentence materials, these individuals showed some masking release in phoneme recognition. All three participants exhibited masking release with the tense vowels at some SNRs, but only one showed release with the plosives. Modulation detection interference (MDI) was observed in these participants to varying degrees. Whereas MDI was observed in the two participants exhibiting little release with the plosives, no MDI was observed in the participant exhibiting release with the plosives. The relationship between implant users' phoneme recognition and psychophysical performance will be discussed. [Work supported by NIDCD-R01DC04786.]

1pPP31. A theoretical model of cochlear processing improves spectrally degraded speech perception. Evan C. Smith and Lori L. Holt (Dept. Psych. & The Ctr. for the Neural Basis of Cognition, Carnegie Mellon Univ., Pittsburgh, PA 15213, evan@cnbc.cmu.edu)

Smith and Lewicki, Neural Comp. **17**, 19–45, 2005a; Adv. Neural. Inf. Process. Syst. **17**, 1289–1296, 2005b; Nature, **439**, 7079, 2006, demonstrated that mammalian hearing follows an efficient coding principle (Barlow, Sensory Communications, 217–234, 1961; Atick, Network, **3**(2), 213–251, 1992; Simoncelli and Olshausen, Ann. Rev. Neurosci., **24**, 1193–1216, 2001; Laughlin and Sejnowski, "Communications in Neuronal Networks," Science, **301**, 1870–1874 2003). Auditory neurons efficiently code for natural sounds in the environment, maximizing information rate while minimizing coding cost (Shannon, Science, **270**, 303–304, 1948). Applying the same analysis to speech coding suggests that speech acoustics are optimally adapted to the mammalian auditory code (Smith and Lewicki, Neural Comp. **17**, 19–45, 2005a; Adv. Neural. Inf. Process. Syst. **17**, 1289–1296, 2005b; Nature, **439**, 7079, 2006). The present work applies this efficient coding theory to the problem of speech perception in individuals using cochlear implants (CI), for which there exist vast individual differences in speech perception and spectral resolution (Zeng *et al.*, Auditory Prostheses and Electric Hearing, **20**, 1–14, 2004). A machine-learning method for CI filterbank design based on the efficient-coding hypothesis is presented. Further, a pair of experiments to evaluate this approach using noise-excited vocoder speech (Shannon *et al.*, Bell Systems Technical Journal **27**, 379–423, 623–656, 1995) is described. Participants' recognition of continuous speech and isolated syllables is significantly more accurate for speech filtered through the theoretically-motivated efficient-coding filterbank relative to the standard cochleotopic filterbank, particularly for speech transients. These findings offer insight in CI design and provide behavioral evidence for efficient coding in human perception.

1pPP32. Processing of speech in complex listening environments by individuals with obscure auditory dysfunction. Kim Block, Sandra Gordon-Salant (Dept. of Speech and Hearing Sci., Univ. of Maryland College Park, 0100 Lefrak Hall, College Park, MD 20742), Monica Hawley, and Craig Formby (Univ. of Maryland School of Medicine, Baltimore, MD 21201)

Obscure auditory dysfunction (OAD) is a disorder for which patients report excessive amounts of difficulty understanding speech in the presence of background noise, despite no obvious evidence of hearing loss. It has been hypothesized that OAD may be the result of mild cochlear dysfunction, central auditory processing deficits, and/or psychological disorders. We evaluated 10 persons with self-reported problems understanding speech in noisy environments and 10 controls. Both groups had normal-hearing sensitivity and excellent word recognition in quiet. The following tests were administered to each participant: DPOAE and ABR assessment, central auditory processing screening, personality inventory, and tests involving speech recognition in complex environments. The speech recognition measures were a subset of tests reported in Hawley *et al.* [J. Acoust. Soc. Am. **115**, 833–843 (2004)] and were designed to probe binaural processing ability, as well as spectral and temporal processing of speech in complex environments. The OAD participants had poorer overall speech recognition ability than did the controls in both single- (1.5 dB, $p=0.017$) and two-competitor environments (0.9 dB, $p>0.05$). However, the reported problems in these OAD participants are not attributable to poor binaural processing. [Funding provided by NIDCD.]

1pPP33. Predictions of speech quality under conditions of noise and distortion. James Kates (GN ReSound and Univ. of Colorado, Dept. Speech Lang. and Hearing Sci., 409 UCB, Boulder, CO 80309, jkates@gnresound.dk), Kathryn Arehart, Lewis Harvey, Jr., and Melinda Anderson (Univ. of Colorado, Boulder, CO 80309)

Noise and distortion reduce speech intelligibility and quality in audio devices such as hearing aids. The purpose of this study is to quantify and model the effects of signal degradation on the perception of speech quality in normal-hearing and hearing-impaired listeners. The stimuli were sentences subjected to eight different levels of additive noise, peak clipping, and center clipping distortion. The subjects listened to all possible comparisons of pairs of the 24 degraded sentences, and for each comparison indicated their preference. Multi-dimensional analysis is used to model the subject decision spaces, and the data are also modeled using a one-dimensional metric to predict the subject quality judgments. The one-dimensional metric is an extension of the procedure developed by Kates and Arehart [J. Acoust. Soc. Am. **117**, 2224–2237 (2005)] to model speech intelligibility. That approach is based on dividing the speech signal into three amplitude regions, computing the coherence for each region, and then combining the three coherence values across frequency in a calculation based on the speech intelligibility index (SII). The one-dimensional and multi-dimensional analyses are able to accurately model the quality perception of both normal-hearing and hearing-impaired listeners. [Work supported by GN ReSound and Centers for Disease Control.]

1pPP34. Cognitive effects of noise reduction strategies. Anastasios Sarampalis (Dept. of Psych., Univ. of California at Berkeley, 3210 Tolman Hall, Berkeley, CA 94720, asaram@berkeley.edu), Sridhar Kalluri, Brent W. Edwards (Starkey Hearing Res. Ctr., Berkeley, CA 94704), and Ervin R. Hafter (Univ. of California at Berkeley, CA 94720)

Noise reduction algorithms that attempt to improve speech intelligibility in noise have been used in hearing aids for decades. Yet, the surprising result has been that they are not effective in improving the speech reception threshold (SRT). One explanation is that the artificial processing is redundant, providing the same information as that extracted by the listener's own auditory system. Our working hypothesis is that natural processing is effective only if there are sufficient cognitive resources to support it and that a reduction in those resources may increase the functional value of the computational algorithm. In the present experiments, subjects performed in a dual-task paradigm in which they listened to and repeated sentences presented in a noisy background while doing a competing visual task, such as a simple driving game. Using driving performance as an indicator of mental effort, we evaluate the effects of different noise reduction algorithms on speech reception. The logic is that, while there will be a loss in performance due to doing two difficult tasks at once, this loss will be smaller when auditory sentences are processed by noise reduction.

1pPP35. Strategies for phonological learning. McNeel G. Jantzen and Betty Tuller (Ctr. for Complex Systems and Brain Sci., Florida Atlantic Univ., 777 Glades Rd., Boca Raton, FL 33431)

An adult's ability to perceive speech sounds that do not occur in their native language is influenced by the relationship between the new and native language phonology and by the individual's capacity to perceive acoustic differences. The study investigated how learning a non-native speech sound occurs over time. First, a perceptual mapping procedure was performed using a synthetic continuum that ranged from the Malayalam voiced, unaspirated, dental stop consonant to the American English alveolar [d]. Subjects identified the stimuli (2AFC) and judged the goodness of stimuli as exemplars of each of the two categories. Next, subjects received 15 sessions (one/day) of 2AFC training with feedback, using natural stimuli produced by native speakers of Malayalam and performed the difference ratings on selected pairs of synthetic stimuli. The perceptual mapping procedure was repeated immediately following training and at two weeks post. Results are consistent with previous work [Case, *et al.*, "The dynamics of learning to hear new speech sounds," Speech Pathol., Nov. 17 (2003)]. Moreover, comparison of results from the different tasks suggests that some subjects learn the acoustic differences among stimuli, while others treat the stimuli as belonging to two phonological categories. [Work supported by NSF and ONR.]

1pPP36. Perception of roughness by people with normal hearing and sensorineural hearing loss. Jennifer B. Tufts (Univ. of Connecticut, 850 Bolton Rd., Unit 1085, Storrs, CT 06269, jennifer.tufts@uconn.edu) and Michelle R. Molis (Portland VA Medical Ctr., Portland, OR 97207)

Anecdotal and laboratory evidence indicates that many people with sensorineural hearing loss have difficulty perceiving music clearly, even when the musical sounds are loud enough to be audible, or are amplified through hearing aids. A better understanding of the music perception difficulties faced by hearing-impaired people is required in order to improve assistive listening technology for music listening, performance, and appreciation. One important aspect of music appreciation is the perception of dissonance between combinations of tones (e.g., the tritone). A recent study [J. B. Tufts, M. R. Molis, and M. R. Leek, J. Acoust. Soc. Am. **118**, 955–967 (2005)] showed that people with sensorineural hearing loss perceive differences in musical dissonance to be less salient than do normal-hearing people. Various acoustic and perceptual factors contribute to musical dissonance, but one of the most important is auditory roughness. In the present study, roughness regions for amplitude-modulated tones at several carrier frequencies, two different modulation depths, and several presentation levels were established using magnitude estimation techniques in normal-hearing and hearing-impaired listeners. Differences in the perception of roughness between the listener groups will be discussed with regard to implications for music perception in people with sensorineural hearing loss.

1pPP37. Automatic perceptual tuning of hearing aids for multidimensional perceptual preferences. Shih-Hao Lin, Pei-Chun Li, and Shuenn-Tsong Young (Inst. of Biomed. Eng., Natl. Yang-Ming Univ. No. 155, Sec. 2, Linong St., Taipei City 112, Taiwan, young@bme.ym.edu.tw)

The numerous parameters of modern digital hearing aids and the multi-dimensional perceptual preferences of a subject present the audiologist with a difficult fitting task. A system was designed for integrating various real-world stimuli selected from the questionnaire of the most important hearing aid benefits for people in Taiwan and common fitting guidelines, and an automatic modified simplex procedure to perform efficient search for the optimal preference structure in a subject's perceptual space. Three experiments were conducted using six normal and six hearing-impaired subjects. Four parameters, the low-frequency gain, high-frequency gain, compression threshold, and compression ratio, were varied to find the desired tonality in the first experiment, the desired Mandarin speech intelligibility in the second, and the overall satisfaction in the third. The results showed that the modified simplex procedure worked efficiently for all cases, however, the test-retest reliabilities of the last two were lower than that of the first experiment. Besides, the preferred parameters were different by comparing the intersubject judgments for the same objective and the within-subject judgments for different objectives. The results indicated that there are multiple peaks in one's response surface and further studies are required to improve reliability and satisfaction when fitting more complicated parameter sets.

Session 1pSC

Speech Communication: Speech Perception (Poster Session)

Michael J. Kiefte, Chair
Dalhousie Univ., Human Communication Disorders, 5599 Fenwick St., Halifax, NS, B3H 1R2, Canada

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

1pSC1. The role of vowel-to-consonant transitions in the perception of phonological consonant length. Benjamin B. Hansen (Dept. of Linguist., Univ. of Texas, CAL 501, Austin, TX 78712)

It has been noted [S. Kawahara, Univ. of Mass. Occasional Papers in Linguistics 32: Papers in Optimality Theory (2005)] that more sonorous consonants are less likely to occur as geminates in the world's languages. Two experiments were performed to test whether a phonetic parameter, the vowel-to-consonant formant transition duration, relates to the possible neutralization of the length distinction for more vowel-like consonants. In a production experiment the consonant constriction and onset transition durations are measured for Persian geminate and singleton consonants spoken at three speaking rates. Stops, fricatives, nasals, liquids, glides, and glottal consonants are studied. Segments that are predicted to be less likely to occur as geminates have longer transitions at all speaking rates. A perception experiment based on manipulated palatal glide stimuli demonstrates that listeners have more difficulty categorizing a consonant as long or short if the transition is longer. Difficulty of categorization is measured both by reaction times and the slope of the curve-fitted cumulative identification model, where a flatter slope indicates a less distinct perceptual boundary.

1pSC2. Auditory distinctiveness and the perception of voiceless stop consonants in continuous speech. Jason Bishop (Eastern Michigan Univ., 612 Pray-Harrold Hall, Ypsilanti, MI 48197, jbishop@emich.edu)

Explaining the way in which listeners cope with the apparent lack of acoustic invariance as well as the phonetic ambiguity frequently encountered in continuous speech has remained problematic for models of speech perception. The voiceless stop consonants /p t k/, for example, are frequently unreleased word-finally. Despite the reduction in articulation and presumably in the distinctiveness of the acoustics signal, listeners appear quite accurate in identifying the correct stop consonant during conversation. Sets of words such as {shop, shot, shock} and {cheap, cheat, cheek} should, given the absence of release of the final stops, be difficult (if not impossible) to distinguish in isolation. In this study, listeners were presented with such lexical stimuli (in isolation as well as in semantically priming sentences) and tested for their perception of the word-final [p t k]. It was found that the amount of distinctive information carried in the acoustic signal for this class of sounds in the environment of interest was insufficient to consistently disambiguate listener impressions; further evidence for lack of perceptual salience is provided by an extremely weak resistance to semantic context effects. The results of this study highlight the value of an auditory dimension to perceptual studies in linguistic phonetics.

1pSC3. Perception of pitch and prominence in a speech context. Alejandro C. Renato (Inst. Genius of Technol., Av. Aa 875, Distrito Industrial, Manaus, Brazil)

Different researchers have proposed a dominance region that refers to the band of harmonics whose frequencies determine the perceived pitch. This work examined the different cues appearing in the dominance region in order to discriminate the pitch and the prominence in a speech context. It has been done as a reiterant speech task. Native Spanish speakers were instructed to imitate the intonation and prominence perceived in the stimuli presented. The stimuli were composed by trisyllabic words of the form CV1CV2CV2, where C was the voiced lateral /l/, V1 = /a/, and V2 the five vowels of Spanish /a/, /e/, /i/, /o/, and /u/. The degree of prominence was varied from the first to the last syllable, and the final tone. The stimuli were passband filtered between 250 and 900 Hz, their harmonics extracted and synthesized in three ways: (a) The unresolved harmonic more close to 600 Hz; (b) The three resolved harmonics around 600 Hz; and (c) The two resolved harmonics with larger amplitud by phonetic

segment, next to the vowel formant. Results indicate that the stimuli (c) described above presents a better relationship between pitch and prominence discrimination.

1pSC4. Effects of reliable spectral-temporal characteristics of listening context on perception of monophthongs and diphthongs. Keith R. Kluender (Dept. of Psychol., Univ. of Wisconsin, Madison, WI 53706-1696) and Michael Kiefte (Dalhousie Univ., Halifax, NS B3H 1R2 Canada)

Brief experience with reliable characteristics of listening context, such as a precursor sentence, alters perception of subsequent vowel sounds [M. Kiefte and K. Kluender, J. Acoust. Soc. Am. **112**, 2248 (2002)]. For example, an appropriately filtered precursor sentence serves to perceptually cancel effects of second formants (F_2) of vowel sounds, forcing listeners to rely solely on spectral tilt for identification. Auditory systems effectively absorb spectral characteristics, including formant peaks, if those characteristics are reliable (redundant) properties of the listening context. Present studies investigate whether this perceptual calibration extends to reliable spectral-temporal properties. Listeners identified a series of stimuli with incremental increases in slope of F_2 that varied perceptually from /u/ to /ui/. When repeated formant transitions, exactly matching those of the target vowels, were added to precursor sentences, /i/ percepts predominated despite no stimuli being appropriate for /i/ in isolation. When repeated formant transitions were reversed in precursors (high to low frequency), the same results were obtained. Because perceptual absorption of reliable acoustic characteristics was indifferent to F_2 trajectory, it appears that processes that calibrate to reliable properties of a listening context are sensitive only to spectral, not temporal, composition of listening context. [Work supported by SSHRC and NIDCD.]

1pSC5. Perceptual absorption of listening context when perceiving musical instruments. Christian E. Stilp and Keith R. Kluender (Dept. of Psych., Univ. of Wisconsin—Madison, 1202 W. Johnson St., Madison, WI 53706)

Brief experience with reliable characteristics of listening context, such as a precursor sentence, alters perception of subsequent vowel sounds. In experiments focused on perception of vowels, effects of preceding context do not depend on sentences being intelligible speech [K.R. Kluender and M. Kiefte, J. Acoust. Soc. Am. **114**, 2335 (2003)]. In the present studies, listeners identified musical instruments (tenor saxophone versus French horn) following a context of speech sentences that are modified by spectral envelope difference filters created to capture differences between saxophone and horn spectra. Precursor sentences were resynthesized renditions of a female talker saying "Please identify this instrument." For a morphed series of sounds with 196 Hz fundamental frequency (G3) varying perceptually from saxophone to horn, listeners are more likely to report hearing a saxophone following a sentence filtered to emphasize spectral characteristics of the horn and vice versa. Consistent with previous studies, these effects are contrastive and indicative of the auditory system absorbing reliable (redundant) properties in the service of emphasizing change and maximizing transmitted information. Perceptual calibration to reliable spectral characteristics does not depend upon sounds being perceived as arising from the same source. [Work supported by NIDCD.]

1pSC6. Compensation for coarticulation: Three theories compared. Navin Viswanathan and James S. Magnuson (Univ. of Connecticut, Storrs, CT 06269)

Classification of a /da/-/ga/ continuum shifts towards more /ga/ responses following /al/ and more /da/ responses following /ar/. Theories explain this as compensation for coarticulation: in natural speech, if speakers must transition from relatively front (/l/) to back (/g/) places of articulation, they will be unlikely to reach the optimal place of articulation. Thus, after /l/, /g/ is likely to be produced further forward than usual and, following /l/, ambiguous tokens in a /d/-/g/ continuum will often be perceived as /g/ with place of articulation shifted by coarticulation. The current study compares the perceptual effects of American English /r/ and /l/ with those of Tamil /R/ (trilled) and /L/ (retroflex). /R/ shares its place of articulation with /l/ but is categorized as an allophone of /r/ by American English speakers. Similarly, /L/ shares its place of articulation with /r/ but is classified as an instance of /l/. However, the acoustic relationships between the Tamil and American English phones are complex, potentially leading to conflicting predictions from auditory contrast and other theoretical explanations of compensation. Our acoustic analyses and experiments examine whether compensation depends upon place of articulation, acoustic cues that might produce contrast effects, or perceptual classification of a phoneme.

1pSC7. Audiovisual asymmetries in speech and nonspeech. Julie M. Brown (Haskins Labs., 300 George St., New Haven, CT 06511, julie.brown@uconn.edu)

The individual contribution of the various perceptual systems to multimodal perception is typically examined by placing information from the various perceptual systems in conflict and measuring how much each contributes to the resulting perception. When one perceptual system dominates perception more than another, it is called intersensory bias. The present research examines intersensory bias using a tapping task in which participants were asked to tap to the frequency of either an auditory or visual stimulus. The stimuli that were used were either nonspeech stimuli, such as a tone and a flashing ellipse, or speech, such as a visual or auditory syllable. In past experiments that have used nonspeech stimuli, audition has been shown to dominate visual temporal perception. Based on the influence of visible speech on auditory speech perception found in the McGurk effect, it was thought that vision might influence temporal perception more when using speech stimuli than nonspeech stimuli. It was found that the speech and nonspeech distinction did not affect intersensory bias. Rather, the amount of information provided by the auditory and visual stimuli predicted the occurrence of intersensory bias. [This research was supported by a NIH grant through Haskins Laboratories, New Haven, CT.

1pSC8. Parsing time and rate normalization versus durational contrast. Eve Brenner-Alsop (Ling. Dept., Univ. of Massachusetts, Amherst, MA 01003-9274)

An acoustic interval is perceived as shorter next to a long neighboring interval than next to a short one. This finding has been attributed to rate normalization (Miller and Liberman, Percept. Psychophys. **25**, 457–465, 1979) and to durational contrast (Diehl and Walsh, J. Acous. Soc. Am. **85**, 2154–2164, 1989). In the first experiment, we present listeners with a vowel varying in duration from short to long in a context where it could be interpreted as one vowel or two in sentences manipulated to sound as though they were spoken at fast, moderate, and slow rates. We predict that the vowel is more often perceived as two vowels in the fast than the slow sentence. The second experiment tests the effect of parsing the vowel into one versus two vowels on judgments of a [b-w] transition duration continuum and a [b-p] voice onset time continuum. If the vowel is parsed into two vowels, a given transition duration or VOT will sound longer and thus more [w] or [p]-like than if it is parsed into just one. Such a result would reflect durational contrast between the transition duration or VOT and the vowel's perceived duration. It would not reflect rate normalization directly but only indirectly via its effect on parsing. [Work supported by NIH.]

1pSC9. Sequential contrast or compensation for coarticulation? Della Chambless and John Kingston (Ling. Dept., Univ. of Massachusetts, Amherst, MA 01003-9274)

English listeners identify a stop ambiguous between /t/ and /k/ more often as "k" after [s] than [sh] (Mann and Repp, J. Acous. Soc. Am. **69**, 548–558, 1981). Judgments shift similarly after a fricative ambiguous between [s] and [sh] when its identity is predictable from a transitional probability bias but perhaps not from a lexical bias (Pitt and McQueen, J. Mem. Lang. **39**, 347–370, 1998, cf. Samuel and Pitt, J. Mem. Lang. **48**, 416–434, 2003). In replicating these experiments, we add a discrimination task to distinguish between the predictions of competing explanations for these findings: listeners respond "k" more often after [s] because they compensate for the fronting of the stop expected from coarticulation with [s] or because a stop with an F3 onset frequency midway between [t]'s high value and [k]'s low value sounds lower after the [s]'s high-frequency energy concentration. The second explanation predicts listeners will discriminate [s-k] and [sh-t] sequences better than [s-t] and [sh-k] sequences because sequential contrast exaggerates the spectral differences between [s-k]'s high-low intervals and [sh-t]'s low-high intervals and distinguishes them more perceptually than [s-t]'s high-high intervals and [sh-k]'s low-low intervals. Compensation for coarticulation predicts no difference in discriminability between the two pairs because it does not exaggerate differences between the two intervals. [Work supported by NIH.]

1pSC10. Sequential contrast and assimilation in the perception of neighboring vowels and consonants. John Kingston, Della Chambless, Daniel Mash, and Jonah Katz (Ling. Dept., Univ. of Massachusetts, Amherst, MA 01003-9274)

Sequential contrast exaggerates perceptual differences between neighboring segments. Our experiments measure this effect in consonant-vowel-stop and stop-vowel-consonant strings. Listeners identify and discriminate vowels from a front-to-back continuum in next-to-coronal versus labial stops, or stops from a coronal-to-labial continuum in next-to-front versus back vowels. If these intervals contrast sequentially, a vowel midway between front and back should sound backer next to a coronal than a labial stop because a coronal concentrates energy high in the spectrum. Similarly, a stop midway between coronal and labial should sound more labial next to a front vowel, because it too concentrates energy high in the spectrum. Exaggeration of spectral differences would also make a front vowel + labial stop versus a back vowel + coronal stop more discriminable than a front vowel + coronal stop versus a back vowel + labial stop. The strings' lexical status is manipulated orthogonally via the other consonant to test whether sequential contrast effects are independent of lexical biases. Both identification and discrimination results show independent sequential contrast effects in stop judgments after front versus back vowels, but assimilation in vowel judgments before labial versus coronal stops. The stop-vowel-consonant strings test whether this difference reflects segment order or a consonants-vowel difference. [Work supported by NIH.]

1pSC11. Perceptual salience and acoustic distance: Evidence from Tagalog nasal-place contrasts. Chandan Narayan (Dept. of Linguist., Univ. of Michigan, 1190 USB, 204 Washtenaw Ave., Ann Arbor, MI 48109-2215)

The acoustics and perception of three Tagalog (Austronesian) nasal-onset places of articulation ([ma], [na], and [nga]) are investigated. In experiment 1, native speakers of Tagalog ($n=13$) were presented with the [ma]-[nga] and [na]-[nga] contrasts (as recorded by a native speaker of Tagalog) in an AX task. As expected, listeners performed according to Tagalog phonology, with successful discrimination of both contrasts ($>90\%$ correct). Listener performance on the [na]-[nga] contrast, however, was significantly lower than performance on [ma]-[na]. In a second experiment, designed to assess the perceptual salience of the [na]-[nga] contrast relative to the [ma]-[na] contrast, native-Tagalog listeners were tested (in an AX task) with the contrasts in signal-dependent (though uncorrelated) noise. Findings are discussed in light of acoustic measurements of multiple repetitions of [ma], [na] and [nga] as produced by three speakers of Tagalog. These acoustic results show that, in F2×F3 space (measured at the NV juncture), [na] is more similar to [nga] than it is to [ma]. In addition, the acoustic and perceptual findings will be interpreted relative to recent findings in the development of nasal-place perception in infancy. [ng=velar nasal.]

1pSC12. Experience-driven effects of visual cues in speech perception. Joseph D. W. Stephens and Lori L. Holt (Dept. Psych. & Ctr. for the Neural Basis of Cognition, Carnegie Mellon Univ., 5000 Forbes Ave., Pittsburgh, PA 15213, jds2@andrew.cmu.edu)

The integration of information across modalities is a key component of behavior in everyday settings. The current study examined the extent to which experience drives multimodal speech integration. Two groups of participants were trained on combinations of speech sounds with corresponding videos of an animated robot, whose movements and features bore no resemblance to speech articulators. Participants' identification of acoustically presented consonants was influenced by simultaneous presentation of learned visual stimuli in a manner that reflected the correlation structure of auditory and visual cues in training. The influence of novel non face visual cues on speech perception developed over the course of training, suggesting that experience altered the perceptual mechanisms used in combining this cross-modal information. Pairings of auditory and visual cues given to two groups of participants resulted in patterns of bimodal perception that differed in systematic ways. Perceptual integration of the newly learned visual cues with auditory speech was not optimal, reflecting a bias toward the sound cue. The findings are discussed in reference to current models of speech perception and multimodal integration.

1pSC13. Parsing voicing cues: A context effect predicted by feature cue parsing theory. Donna Van Worthe, Jillian Celi, and David Gow (Salem State College, Salem, MA 01970)

Factors including lawful assimilation conspire to produce overlapping temporal distributions of phonetic feature cues for neighboring segments. Gow's feature cue parsing model [Gow, Perception and Physics, **65**, 575–590 (2003)] suggests that listeners rely on perceptual grouping mechanisms to integrate multiple cues and align them with correct segmental positions. While this account was developed to address progressive and regressive context effects found in the perception of lawfully assimilated speech, its strong form predicts that similar context effects may be found in the perception of speech where there is similar overlap in feature cue distribution, even if such overlap is not the result of a lawful phonological process. This hypothesis was examined in a series of speeded 2AFC experiments in which voicing cues were manipulated in adjacent fricatives and stops. A regressive context effect was found in the perception of the fricative in a peace-peas voicing continuum, which was strongly affected by the voicing of a following stop (e.g., pay versus bay). Similarly, listeners showed a progressive context effect in the perception of stops in a pay-bay VOT continuum when these items were preceded by voiced versus voiceless fricatives. These results are discussed in terms of competing interpretations of assimilation context effects. [Work supported by NIH R01DC3108.]

1pSC14. The time course of cross-voice speaking rate normalization. James R. Sawusch (Dept. of Psych., Univ. at Buffalo, Park Hall, Buffalo, NY 14260, jsawusch@buffalo.edu) and Rochelle S. Newman (Univ. of Maryland, College Park, MD 20742)

Individuals vary their speaking rate, and listeners use the speaking rate of precursor sentences to adjust for these changes [Kidd, G. R., "Articulatory-rate context effects in phenone identification," J. Exp. Biol.: Human Percep. Perfor. **15**, 736–748 (1989)]. This speaking rate adjustment is not limited to speech from a single talker [Sawusch & Newman,

''Perceptual normalization for speaking rate II: Effects of signal-discontinuities,'' Percep. Psychophys. **62**, 285–300 (2002)]; if one speaker stops talking and another begins, listeners will use rate information from the first talker to interpret speech in the second. We examined how long this cross-talker effect lasts. A male talker produced a carrier sentence at fast and slow speaking rates. A female talker then produced one or more subsequent words at a constant speaking rate. The final word in the female voice contained a duration-based contrast. We examined how much intervening speech in the female voice was necessary before the male voice's speaking rate no longer influenced perception of the final target word. Results suggest that listeners require approximately 1 s of intervening speech or of silence before they stop using rate information from the male talker. This suggests that rate normalization may be based on a 1-s temporal window, with any speech information that occurs within 1 s in advance of a target phoneme influencing perception of that phoneme.

1pSC15. Contextual effects on the perception of duration. Shigeto Kawahara (Dept. of Linguist., Univ. of Massachusetts, Amherst, 226 South College, Amherst, MA 01003, kawahara@linguist.umass.edu)

English speakers perceive silent intervals with the same duration as longer when preceded by shorter vocalic intervals [Kluender *et al.* ''Vowel length differences before voiced and voiceless consonants: An auditory explanation,'' J. of Phonetics, **16**, 153–169 (1988)]. This may be due to sequential contrast effects which exaggerate perceptual differences between neighboring segments. In Japanese, vowels are longer before geminates than before singletons [Han ''Acoustic manifestations of more timing in Japanese,'' J. Acoust. Soc. Am. **96**, 73–82 (1994)]. Reflecting this correlation, Japanese speakers perceive silent intervals with the same duration as longer (as more ''geminatelike'') when preceded by longer vowels. The same pattern holds when non-speech analogs are used in place of speech stimuli: silent intervals are judged as longer when preceded by longer anharmonic buzzes. These results suggest that the perception of silent duration by Japanese speakers follows the pattern predicted from the acoustic properties of Japanese, but goes in the opposite direction from that predicted if durational contrast determines judgments, as it did for Kluender *et al.*'s (1988) listeners. This suggests that the segmental properties of the listener's native language may influence nonspeech temporal judgments [cf. Moreton and Kusumoto ''Native language determines parsing of nonlinguistic rythmic stimuli,'' J. Acoust. Soc. Am. **102**, 3204 (1997)]. English speakers will be tested on the same tasks to determine whether the perception of the temporal properties of nonspeech signals are generally determined by those of the listeners' native language. [Work supported by NIH.]

1pSC16. Spectral auditory integration and virtual cues to place-of-articulation in stops. Robert Allen Fox, Marc Smith, and Ewa Jacewicz (Speech Percept. and Acoust. Labs., Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210)

The spectral ''center-of-gravity'' (COG) hypothesis proposed by Chistovich and colleagues for static vowels suggested spectral auditory integration of two or more formant peaks within 3.5 barks. More recently, studies have examined dynamic COG effects by modifying the amplitude of steady-state formants over time in order to create ''virtual'' diphthongs. The present study looks at ''virtual frequency change'' cues to the alveolar-velar place distinction in stops [an extension of Fox and Smith, J. Acoust. Soc. Am. **117**, 2423 (2005)]. First, [d≥-g≥] and [t≥-k≥] series were synthesized with a 50-ms *F*3 transition (with a voiced or voiceless source). Next, matched stimulus series were created with virtual *F*3 transitions by changing the amplitudes of either a pair of sine waves (in [d≥]-[g≥]) or a pair of narrow band noise signals (in [t≥]-[k≥]) whose frequencies did not change. Another matched [t≥]-[k≥] series was produced with both actual and virtual stops bursts (with no *F*3 transitions). Identification of these stimulus series by listeners with normal hearing show that a dynamically changing COG in the initial 50 ms of the token can effectively cue place distinctions in stops as do formant transitions and bursts; however, virtual spectral cues are less salient than formant or burst cues. [Work supported by NIDCD R01DC00679-01A1.]

1pSC17. Listener sensitivity to intensity variation in vowels in coarticulatory context. Ewa Jacewicz and Robert Allen Fox (Speech Percept. and Acoust. Labs., Dept. of Speech and Hearing Sci., The Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210)

There are two sources of acoustic variation in the relative formant amplitude of a vowel which are unrelated to emphasis-oriented changes in articulatory effort: (1) vowel-specific intrinsic formant frequency pattern, which differs across vowel categories, and (2) the vowel's immediate phonetic (coarticulatory) context. Listener sensitivity to the latter source of variation is examined here to establish the extent to which distortion in the distribution of acoustic energy across the vowel spectrum affects the perception of vowel quality. Synthetic versions of CVC syllables in symmetrical consonant environments were created (using stops or fricatives differing in place-of-articulation and voicing). Formant frequencies and amplitude values were modeled after those found in a large corpus of CVC acoustic data [E. Jacewicz and R. A. Fox, J. Acoust. Soc. Am. **117**, 2623 (2005)]. Initial amplitude values for the first four formants in synthetic CVCs reflected the intensity distribution (and spectral tilt) typical of a vowel coarticulated with these consonants. Subsequent stepwise adjustments to selected formant levels were made to create a range of permissible and not permissible formant amplitude variations in coarticulatory context. Preliminary results indicate that listeners are sensitive to amplitude variation across vowel spectrum as a function of a specific consonant environment. [Work supported by NIDCD R03 DC005560-01.]

1pSC18. The feature [sonorant] in auditory word identification. Danny R. Moates, Russell Fox (Dept. of Psychol., Ohio Univ., Athens OH 45701) Chao-Yang Lee and Sara Kellgreen (Ohio Univ., Athens, OH 45701)

The distinctive feature [sonorant] separates resonants and obstruents. Does this feature mark a difference in the perceptual processing of these consonants? Marks *et al.* [Marks *et al.*, ''Word reconstruction and consonant features in English and Spanish,'' Linguistics, **40**, 421–438 (2002)] showed such an influence in a word reconstruction task. Target words containing an obstruent segment were harder to reconstruct when the obstruent was replaced by a resonant (mismatch condition) than when it was replaced by another obstruent (match condition). Resonant target words showed no such effect. The present study extended Marks *et al.* (2002) using an auditory priming lexical decision task. Participants heard a nonword prime 50 msec prior to another word or nonword target. The prime differed from the target item in just one consonant which either matched the target segment on the feature [sonorant] or mismatched it. Listeners judged whether the target item was a real word or nonword. The obstruents were divided into fricatives and stops, which were compared with resonants in both cases. Latencies for the fricative target words were significantly slower in the mismatch condition than in the match condition but not for stops. The present study validates the effect found by Marks *et al.* (2002) with an online task but shows that it is limited to fricatives.

1pSC19. Perception of acoustically modified Mandarin tones. Chao-Yang Lee (School of Hearing, Speech and Lang. Sci., Ohio Univ., Athens, OH 45701), Liang Tao, and Zinny S. Bond (Ohio Univ., Athens, OH 45701)

How well do listeners recognize tones when only partial acoustic information is available to them? Twelve minimal tone pairs including all six Mandarin tonal contrasts were digitally processed to generate four syllable modifications: silent-center, center-only, onset-only, and intact syllables. The syllables were recorded in two carrier phrases such that the offset of the carrier and the onset of the target were either a match or

mismatch in fundamental frequency. The stimuli were presented in the original carrier phrase, excised from the carrier phrase, or excised and cross spliced with another carrier phrase. Forty native speakers of Mandarin Chinese were asked to identify the tone associated with each syllable. Response time and accuracy were measured. Listeners identified the tones at better than 90% correct with and without context except when they heard the syllable onset only; then, their identification accuracy fell to approximately 80%. A match or mismatch in fundamental frequency between the carrier and target tones made no difference in either reaction time or accuracy. Cross-spliced carrier phrases also did not compromise accuracy or reaction time. Tone 2 was identified least accurately. Listeners also took more time to identify tone 1 and tone 2 than the other two tones.

1pSC20. An acoustic and perceptual investigation of the geminate and singleton stop contrast in Japanese. Kaori Idemaru (Dept. of East Asian Lang. and Lit., Univ. of Oregon, Eugene, OR 97403-1248) and Susan Guion (Univ. of Oregon, Eugene, OR 97403-1290)

Segmental durations are known to be heavily influenced by changes in speech rate such that the durations of members of two contrasting length categories may overlap [Pickett *et al.* "Effects of speaking rate on singleton/geminate contrast in Italian," Phonetica **56**, 135–157 (1999); Hirata & Whiton, "Relational acoustic invariance in the single/geminate stop distinction in Japanese," J. Acoust. Soc. Am. **117**, 2569 (2005)]. Given this, questions arise as to how category distinction is achieved for contrasts signaled by differences in segmental duration. This study investigates the question by examining the production and perception of the Japanese singleton versus geminate stop contrast by adult native speakers. Six native Japanese speakers produced disyllabic words with medial singleton and geminate stops at three speaking rates. Duration of stop closure and a number of relational timings between the closure duration and other segments were examined as potential rate-independent measures distinguishing the two categories. The ratios of stop closure to the preceding mora as well as stop closure to the following vowel remained stable across speaking rate and accurately classified naturally produced tokens. Ten native Japanese listeners participated in an identification task to examine the perceptual relevance of these relational timings. The results indicated an influential role of preceding mora duration and a small role of following vowel duration. The results also suggested that nondurational factor(s) influence perception of the stop length contrast.

1pSC21. Could backward recognition masking play an important role in speech contrasts? Jeremy R. Gaston and Richard E. Pastore (Dept. of Psych., Bingahmton Univ. SUNY, Binghamton, NY 13901)

Auditory backward recognition masking (BRM) is the interference in recognition by a following stimulus for some aspect of a clearly detectable target. BRM has been studied for stationary targets, with its conjectured relevance largely limited to the sensory storage and encoding of simple (nonspeech) auditory properties. Could BRM play an important role for more complex stimuli—specifically speech contrasts? The current study systematically investigates BRM interference for speechlike analogs of $F2$ transitions in varying spectral context that mimic the properties of speech contrasted by place of articulation. Initial and final transitions in isolation show equivalent recognition. Adding a vowel $F2$ equivalent significantly increases thresholds for initial, but not final, transitions. The further addition of a vowel $F1$ equivalent results in an additional significant increase in threshold for initial position transitions, and a significant decrease for final position transitions. The possible importance of BRM in defining speech contrasts is discussed. [Work supported by NSF.]

1pSC22. Acoustic and phonetic discrimination in speech using event-related potentials: A mismatch negativity paradigm. Erin K. Robertson, Randy Lynn Newman, and Marc F. Joanisse (Dept. of Psych., Univ. of Western Ontario, 1151 Richmond St., London, ON, N6A 5C2 Canada)

The interaction of acoustic and phonetic discrimination in speech was studied using event-related potentials (ERPs). Participants consisted of 14 neurologically normal right-handed adults. Data were recorded in response to strong and weak acoustic changes signaling consonant distinctions, using a 64-channel encephalogram and a mismatch negativity (MMN) paradigm. The MMN is an automatic response evoked by stimulus change and is elicited when a train of repeated stimuli (standard) is interrupted by an oddball (deviant) stimulus. Stimuli consisted of one standard and two deviants that differed in their acoustic difference from the standard. All three stimuli were synthetic speech syllables in which $F2$ was manipulated to create a continuum between the syllables "ba" and "da." The standard was the endpoint item "da." The "strong deviant" was the "ba" endpoint item; the "weak deviant" was a more intermediate "ba" item, and therefore acoustically closer to the standard. MMNs were identified for both deviants, but the amplitude was smaller for the weak deviant condition. The data suggest that MMNs are sensitive to both phonetic change and also the magnitude of the acoustic difference within this change. The results are discussed with respect to theories of phonetic and acoustic discrimination in speech perception.

1pSC23. Formant deflection directions of the voiced alveolar stop consonant in different vowel contexts. Kang Li and Brad Story (Speech and Hearing Sci., Univ. of Arizona, P.O. Box 210071, Tucson, AZ 85721, lik@u.arizona.edu)

A parametric model of the vocal-tract area function was used to synthesize a series of alveolar stop consonant constrictions imposed on 9 $/V_1V_2/$ transitions (e.g., /idae/). The constrictions differed in configurations, places of occurrence along the vocal-tract length, and the onset and release times to create different on-glide and off-glide formant deflection patterns. It was hypothesized that the formant deflection directions caused by the onset and release of the consonant constriction relative to the underlying vowel-to-vowel formant transitions (i.e., without consonant perturbations) provide information about the identity of the alveolar stop. Perceptual tests were conducted to assess the phonemic identities of the formant patterns produced by the model. The model parameter settings used to create the consonant constrictions that were perceived as /d/'s were analyzed to study the coarticulation effect between /d/ and different vowel contexts.

1pSC24. Is nonmodal phonation at the end of utterances a cue for speaker recognition by humans? Tamás Böhm (MIT-RLE, Cambridge, MA and BME-TMIT, Magyar Tudósok krt. 2., H-1117 Budapest, Hungary, bohm@tmit.bme.hu)

Many studies of nonmodal phonation report a high rate of interspeaker variation. These experiments also showed that the rate and type of glottalization are characteristic to the speaker. Still, it remains an open question whether human listeners use this information as a cue for recognizing familiar voices. A listening experiment was conducted to investigate this issue, concentrating on utterance-final irregularities. Short utterances produced by four speakers were copy synthesized with the Klatt synthesizer in two conditions: with modal and with nonmodal voice quality in their final regions. Because two of the speakers were reliable "glottalizers" and two seldom "glottalized," one of the two synthesis conditions reflected the speaker's usual utterance-final voice quality and the other did not. These pairs of stimuli were used to test the effect of the presence or absence of utterance-final nonmodal phonation on the identification of voices familiar to the subjects. When asked to choose which one of the two variants is the speaker's voice, listeners preferred the stimuli with the speaker's typical

voice quality at the end. Results will be discussed in comparison with a well-known perceptual cue on speaker identity, the F0 contour. [Work supported by a Fulbright Fellowship.]

1pSC25. Recognition of Mandarin words heard through single and multiple spectral bands. Tien Ho, Pei-Chun Li, and Shuenn-Tsong Young (Inst. of Biomed. Eng., Natl. Yang-Ming Univ., No. 155, Sec. 2, Linong St., Taipei City 112, Taiwan, whosweet@gmail.com)

Two listening experiments were designed to evaluate the applicability of two macroscopic models—the articulation index and the speech recognition sensitivity (SRS) model—towards Mandarin speech recognition. The experiments aimed to identify if there is any synergetic and redundant effect of spectral frequency bands in monosyllabic Mandarin words. In experiment 1, 20 normal-hearing subjects listened to monosyllables through 20 bandpass filters, which have center frequency ranging from 125 to 8000 Hz, 1/3-octave bandwidth, and filter slope of 100 dB/octave. The purpose of experiment 1 was to identify the intelligibility of spectral frequency bands in Mandarin speech. In experiment 2, 10 subjects listened to monosyllables with the combination of 2 or 3 individual bands for investigating the synergetic or redundant interactions. The perceptual outcomes of the second experiment were fitted to both models by statistical regressions, and the results indicated that the contribution of each band to recognition of Mandarin monosyllables was not independent, and the recognition of Mandarin tones, consonants, and vowels showed different redundant and synergetic effects. The SRS model appears to be able to predict the differential effect of linguistic entropy on filter condition.

1pSC26. jTRACE: A user-friendly reimplementation and extension of the TRACE model of speech perception and spoken word recognition. Ted J. Strauss, James S. Magnuson (Dept. of Psych., Univ. of Connecticut, 406 Babbidge Rd., Unit 1020, Storrs CT, 06269), and Harlan D. Harris (New York Univ., New York, NY 10003)

jTRACE is a cross-platform Java reimplementation and extension of the TRACE model of spoken word recognition. TRACE (McClelland and Elman, 1986) is a three-layer interactive-activation model of speech perception that takes a pseudoacoustic input representation and accurately simulates many phenomena of human speech processing. The goals of the reimplementation are to provide a user-friendly version of the model that can run on any current computer platform, extend the model with graphing and scripting tools, and thus provide a tool that is powerful enough for intensive research while easy enough to use in a classroom setting. New features in jTRACE include simulation visualization, results graphing, powerful scripting, internal and external noise parameters, adjustable lexical and phonological characteristics, word recognition metrics, data exporting, and saving/archiving tools. A set of historically relevant simulations is bundled with the program. This feature is ideal for educational applications. Any simulation can be saved in a similar format, promoting sharing and replication. jTRACE was validated quantitatively against the original implementation by replicating key simulations reported in the literature. jTRACE can be downloaded freely from <http://maglab.psy.uconn.edu/jtrace.html> [Work supported by National Institute on Deafness and Other Communication Disorders.]

1pSC27. Nonspeech sounds prime acoustically similar words. Andrew B. Wallace and Sheila E. Blumstein (Dept. of Cognit. and Linguist. Sci., Box 1978, Brown Univ., Providence, RI)

Recent experiments show that nonspeech sounds can shift the perceptual boundaries between the phonetic categories of speech [e.g., Holt, Psych. Sci. **16**, 305–312 (2005)]. The current experiment investigates whether nonspeech sounds will also affect the speed of access to words. Subjects received prime-target pairs in which the target was one of the three words deed, dad, and dude, preceded either by tone complexes which matched the frequencies of the first two formants of each target or by a neutral tone midway between the ranges of F1 and F2. Every prime was paired with every target in a word-identification task. Results demonstrated a priming effect of the nonspeech sounds on acoustically similar speech. Significant differences were found in reaction time, with subjects responding faster when targets were preceded by matched primes than when they were preceded by mismatched or neutral primes. These results support a role for general auditory processes in speech perception and lexical access. [Work supported by NIDCD.]

1pSC28. Acquisition of phonological variants in an artificial lexicon: The role of variant frequency. Eleni Pinnow and Cynthia Connine (Psych. Dept., Binghamton Univ., P.O. Box 6000, Binghamton, NY 13902)

A series of experiments examined the role of exposure frequency in the acquisition of phonological variants (schwa deletion) using an artificial lexicon. During training participants heard either two- (experiment 1) or three- (experiment 2) syllable stimuli produced in both schwa-present and schwa-deleted forms. Frequency of the schwa-present or deleted form was manipulated with high-frequency forms presented eight times and low-frequency forms presented one time. An orthographic version of the stimuli was presented during training as feedback. A lexical decision task performed after training showed a reaction time advantage for the high-frequency variant form for two syllable stimuli. No frequency effect was found for three-syllable stimuli. Experiment 3 presented the three-syllable stimuli across 2 days of training. On each day of training, high-frequency variants were heard 15 times and low-frequency variants were heard once. A reaction time advantage for high-frequency forms was found on day 1 that was reduced on day 2; both days revealed more accurate identification for high-frequency forms. Results will be discussed in terms of previous findings for recognition of schwa deletion variants and possible mechanisms for the lexicalization of phonological variants.

1pSC29. A lexical locus for the integration of asynchronous cues to voicing: An investigation with natural stimuli. Joseph Toscano and Bob McMurray (Dept. of Psych., Univ. of Iowa, E11 SSH, Iowa City, IA 52242)

Speech perception requires integrating multiple asynchronous acoustic cues. Two questions pertinent to the integration of VOT and vowel length in word recognition were examined. First, this study asked whether lexical activation is sufficient to integrate these cues or if a sublexical buffer is required. Second, the role of stimulus type was examined since studies have found different results for synthetic and natural stimuli (Shinn *et al.*, 1985; Miller and Wayland, 1993; Utman, 1998). An eye tracking methodology that is sensitive to timing and small differences in activation was used (McMurray *et al.*, 2002). Stimuli were recordings of words (e.g., ''beach'' and ''peach'') whose initial VOT was manipulated along a nine-step continuum using splicing. The length of the following vowel was modified to create long and short vowels. Participants selected the correct object from screens containing the target, competitor, and two filler objects while eye movements were monitored. A small vowel length effect was found on voicing perception. This effect occurred substantially later in the fixation record than the effect of VOT. Thus, effects of vowel length can be seen in natural speech. Further, these cues each affect lexical activation as soon as they are available; no sublexical integration is required.

1pSC30. Gradient effects of continuous acoustic detail revealed by event-related potentials. Joel Dennhardt, Bob McMurray, Steven Luck, and Joseph Toscano (Dept. of Psych., E11 SSH, Univ. of Iowa, Iowa City, IA 52242, bob-mcmurray@uiowa.edu)

A growing body of research has demonstrated that lexical activation during online word recognition is systematically sensitive to continuous detail in the speech signal (Andruski *et al.*, ''The effect of subphonetic differences on lexical access,'' Cognition **52**, 163–187 1994; McMurray *et al.*, ''Gradient effects of within-catagory phonetic variation on lexical

access,'' Cognition **26**, B33–B42 2002). However, investigations of within-category sensitivity to continuous phonetic cues using electrophysiological measures have typically been consistent with categorical perception (Sharma *et al.*, ''Acoustic versus phonetic representation of speech as reflected by the mismatch negatively event-related potential,'' Electroencephalography & Clinical Neurophys.: Evoked Potentials, **88**, 67–71 1993; Phillips *et al.*, ''Auditory cortex accesses: Phonological categories: An MEG mismatch study,'' J. Cogn. Neurosci. **12**, 1038–1055 2000). This raises the possibility that these findings of gradiency are artifacts of the small number of comparisons made in priming experiments or the eye-movement tasks themselves. This issue was addressed with event-related potentials in a P3 categorization paradigm. Subjects heard tokens from beach/peach or dart/tart continua (9 steps, 0–40 ms of VOT). In each block of trials, one of the four words was the target; one button was pressed for the target word and a different button for any of the nontarget words. The amplitudes of the P2 and P3 components decreased continuously as VOT departed from the target. This was the case even when trials in which subjects identified the token as nontarget were excluded. These results favor a gradient view of lexical activation: lexical processing is gradiently sensitive to continuous speech cues and such cues may facilitate word recognition.

MONDAY AFTERNOON, 5 JUNE 2006 ROOM 556AB, 1:00 TO 4:20 P.M.

Session 1pSP

Signal Processing in Acoustics, Underwater Acoustics, and Acoustical Oceanography: Adjoint Modeling in Acoustics II

Jean-Pierre Hermand, Chair
Univ. Libre de Bruxelles, Dept. of Optics and Acoustics, av. F.D. Roosevelt 50, B-1050 Brussels, Belgium

Chair's Introduction—1:00

Invited Papers

1:05

1pSP1. Time reversal and the adjoint imaging method with an application in underwater communication. Oliver Dorn (Universidad Carlos III de Madrid, Avda de la Universidad, 30, 28911 Leganes, Madrid, Spain)

A direct link is established between the time-reversal technique and the so-called adjoint method for imaging. The latter one has been used successfully in medical imaging applications, for example for the early detection of breast cancer in ultrasound tomography. In this application, ultrasound waves are sent through the unknown medium of investigation, and the differences between detected and predicted measurements are computationally back-propagated into the estimated medium in order to determine the correct acoustical parameters. In the application of underwater acoustic communication, the goal is to optimize a given signal, emitted by a base station, such that it focuses on the target user in the desired fashion without interfering with other users who might share the same environment. It turns out that efficient optimization schemes can be derived for achieving this goal, which are very similar to those applied in the above-mentioned medical imaging application, but which can be realized physically in the true (unknown) ocean by iterated physical time-reversal experiments. The close relationship, as well as some differences, between these two applications will be highlighted in this talk. [Financial support for this work from the U.S. Office of Naval Research under Grant N00014-02-1-0090 is gratefully acknowledged.]

1:25

1pSP2. Estimating sound-speed perturbations with full-wave environmental sensitivity kernels. Jit Sarkar, Bruce Cornuelle, W. A. Kuperman (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093-0238, bsarkar@ucsd.edu), and Philippe Roux (LGIT-Maison des Geosciences, Grenoble Cedex 9—France)

The acoustic field at a receiving array due to an individual source is affected by changes in the intervening sound-speed profile. This sensitivity has been calculated for broadband signals centered at 3500 Hz and 4-km range using a Born approximation to perturbations in the wave equation. These sensitivity kernels have been verified against range-dependent propagation model results. The range of validity for the linear approximation has been explored as part of these tests. Least-squares methods can be used to estimate the environmental changes from the measured differences in the received signal(s) and prior knowledge of the statistics of the variability. Sample estimates and error statistics, including resolution, will be presented.

1:45

1pSP3. Applications of adjoint models in underwater acoustics. Paul Hursky, Michael B. Porter (Heat, Light, and Sound Research, Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130), Bruce D. Cornuelle, W. S. Hodgkiss, and W. A. Kuperman (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

Complementing a forward model with its adjoint has some nice features that can be helpful in inverse problems. We will review our previous work on constructing an adjoint for forward propagaton using an implicit finite-differences parabolic equation model and some simulated results in inverting for the sound speed in the water column. We will also review how an adjoint can aid in model identification, or the question of how the inverse problem should be modeled (e.g., in inverting for the properties of the bottom, how

many layers should be used to parametrize the bottom). We will demonstrate this for a range-dependent inversion, in which in one case there is a bottom inclusion (bottom properties are abruptly different over some range interval between the source and receiver), and another case where there is an internal wave. After reviewing these examples, we will discuss some practical issues in applying adjoint models to real data.

2:05

1pSP4. Adjoint approach to the physical characterization of a shallow-water environment. Matthias Meyer, Jean-Pierre Hermand (Dept. of Optics and Acoust., Université libre de Bruxelles, av. F. D. Roosevelt 50, CP194/05, B-1050 Bruxelles, Belgium), Mohamed Berrada (Université de Pierre et Marie Curie, 75252 Paris Cedex 05, France), and Mark Asch (Université de Picardie Jules Verne, 80039 Amiens Cedex 1, France)

In underwater acoustics a variety of different applications of adjoint models has been proposed in recent years. Adjoints have been derived for normal modes and for both the standard parabolic equation and Claerbout's wide-angle approximation. This paper reviews the analytic nonlocal boundary control approach proposed in an earlier paper by the authors [Meyer & Hermand, "Optimal nonlocal boundary control of the wide-angle parabolic equation for inversion of a waveguide acoustic field," J. Acoust. Soc. Am. **117**, 2937–2948 (2005)] and presents a numerical extension that allows direct inversion of the geoacoustic parameters that are embedded in a discrete representation of the nonlocal boundary condition at the water-sediment interface. The effectiveness of this numerical adjoint approach for the physical characterization of a shallow-water environment is illustrated with applications for geoacoustic inversion and ocean acoustic tomography. In particular, it is shown how a joint inversion across multiple frequencies can enhance the performance of the optimization process, especially for the case of a sparse receiver array spanning part of the water column. In an additional example we combine the two applications and discuss the feasibility of geoacoustic inversion in the presence of an uncertain sound-speed profile.

2:25

1pSP5. Bathymetry inversion in shallow water using an adjoint method. Isabelle Charpentier (Project Idopt (CNRS, INRIA, UJF, INPG), BP 53, F-38041, Grenoble Cedex 9, France) and Philippe Roux (Universite Joseph Fourier, Grenoble, France)

Bathymetry inversion is performed at low frequency on a kilometer-scale shallow water waveguide between two vertical arrays of sources and receivers. The inversion is based on the use of the adjoint of a wide-angle Pade PE code. The adjoint is obtained using automatic differentiation software. The bathymetry is decomposed under a limited number of spline functions. Inversion results improve with the number of source/receiver pairs and when a multiple-frequency approach is used. Application to experimental at-sea data is discussed.

Contributed Paper

2:45

1pSP6. Time-reversal-based noniterative exact inverse scattering of multiply scattering point targets. Edwin A. Marengo and Fred K. Gruber (Dept. of Elec. and Comput. Eng., Northeastern Univ., Boston, MA 02115)

This research investigates within exact scattering theory time-reversal-based approaches for the inverse scattering in a known but rather arbitrary embedding background medium of an unknown total scatterer composed of a number of dominant subwavelength (pointlike) scattering centers (targets) that are located at unknown positions and have unknown scattering strengths or reflectivities (characterizing a scattering potential). Of particular interest is the formulation of a new inverse scattering method (comprising both inverse support and inverse medium aspects) based on a two-step procedure. The first step (support inversion) consists of the localization of the targets via signal subspace approaches such as time-reversal MUSIC and a new, high-dimensional signal subspace method that is shown to be linked to maximum likelihood estimation. The second step (nonlinear medium inversion) is a new, noniterative scattering strength inversion formula for the reconstruction of the target scattering strengths or reflectivities. Importantly, the derived method is valid even in the presence of multiple scattering between the targets, which is remarkable since the mapping from the reflectivities to the field data is then nonlinear. The theory is illustrated with several computer examples. [Work supported by AFOSR under Grant FA9550-06-01-0013.]

3:00–3:20 Break

Invited Papers

3:20

1pSP7. Application of the acoustic adjoint equation to normal mode propagation in a waveguide. Aaron Thode (Marine Physical Lab., Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238)

This presentation reviews work that applied the acoustic adjoint equation to situations where the propagating acoustic field in an oceanic waveguide can be expressed as a sum of normal modes. The acoustic adjoint equation [Norton, "Inertive inverse scattering algorithms: Methods of Computing Fréchet derivatives," J. Acoust. Soc. Am. **106**, 2653–2660 (1999)], Hurksy *et al.*, "Adjoint modeling for acoustic inversion, J. Acoust. Soc. Am. **115**, 607–619 (2004)] which has a form very similar to the Born approximation, provides an expression for multiple-order derivatives of the received acoustic field with respect to an infinitesimal change in an environmental parameter, which in turn can be used to derive the gradient of an error or ambiguity function without resorting to numerical derivatives. For the situation of a range-independent environment, closed-form analytical expressions can be obtained for pressure derivatives with respect to range-independent perturbations in refractive index and density in the water column and sediment, as well as derivatives with respect to frequency. Expressions have also been derived for derivatives with respect to three-dimensional refractive index perturbations, which permits the evaluation of certain horizontal refraction effects of environmental perturbations without resorting to three-dimensional propagation codes. Current work on obtaining derivatives with respect to perturbations in waveguide boundary conditions (bottom bathymetry or sea surface shape) will also be discussed. [Work supported by the ONR Uncertainty Program.]

3:40

1pSP8. Wave-theoretic kernels for long-range ocean travel-time measurements. Bruce Cornuelle, Matthew Dzieciuch (Scripps Inst. of Oceanogr., UCSD, 9500 Gilman Dr., Mail Code 0213, La Jolla, CA 92093-0213, bdc@ucsd.edu), and Emmanuel Skarsoulis (Inst. of Appl. and Computational Mathematics, GR-71110 Heraklion, Greece)

Travel-time measurements have been a mainstay of long-range acoustic measurements, but the high sensitivity of ray paths to sound-speed perturbations makes the ray theory questionable. In the ray approximation travel times are sensitive to medium changes only along the corresponding eigenrays. Wave-theoretic modeling can be applied to obtain travel-time sensitivity kernels representing the amount by which ray travel times are affected by localized sound-speed variations anywhere in the medium. Propagation codes using the parabolic equation (PE) approximation allow the efficient calculation of these travel-time sensitivity kernels (TSK) in range-dependent environments that obey reciprocity. At high frequencies, the TSK becomes similar to the ray path. This work will explore the spatial frequency content of the TSK in comparison to expected ocean variability. It will also examine the stability of the TSK for range-dependent environments which produce strong sensitivity of ray paths to initial conditions. [Work supported by ONR.]

4:00

1pSP9. Tangent linear and second-order models from the Born and Rytov approximations. Emmanuel Skarsoulis, George Piperakis (Inst. of Appl. & Comput. Math., FORTH, P.O. Box, 1385, 711 10 Heraklion, Greece, eskars@iacm.forth.gr), and George Makrakis (Univ. of Crete, Crete, Greece)

Adjoint and higher-order methods for the solution of inverse problems in underwater acoustics rely on tangent linear and higher-order models of the forward propagation problem, with respect to the underlying parameters of interest. The use of Born and Rytov approximations is examined for the representation of perturbations of the Green's function and perturbations of the phase, respectively, caused by range-dependent variations of the sound-speed distribution about a range-independent background environment. Using the normal-mode representation for the background Green's function and applying the stationary-phase approach in combination with analytic evaluation at singular points, closed-form expressions are derived for the first- and second-order perturbations of the Green's function, generalizing previous perturbation results derived under the assumption of range independence. The resulting expressions for the phase show the second-order character of the phase (and the associated travel times) in the case of range-dependent, zero-mean sound-speed perturbations.

Separate registration fee required to attend Tutorial

MONDAY EVENING, 5 JUNE 2006 BALLROOM D, 7:00 TO 9:00 P.M.

Session 1eID

Interdisciplinary: Tutorial Lecture

Ralph A. Stephen, Chair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543-1542

Chair's Introduction—7:00

7:05

1eID1. The 2004 Sumatra earthquake and tsunami: Multidisciplinary lessons from an oceanic monster. Emile Okal (Dept. of Geological Sci., Northwestern Univ., 1850 Campus Dr., Evanston IL 60208-2150)

The 2004 Sumatra earthquake was the largest seismic event recorded in 40, perhaps 44, years. The death toll from its tsunami was unprecedented in documented history. Remarkably, the earthquake and tsunami were recorded by many technologies designed for strikingly different purposes, including hydroacoustic sensors of the International Monitoring System of the CTBT, which captured standard (Rayleigh) waves, acoustic (*T*) phases, as well as the pressure component of the tsunami wave itself. I will discuss the contributions of such records, and other systems, to the investigation of the earthquake's source and to the nature of the "high"-frequency (10–20 MHz) components of the tsunami wave in the far field.

Session 2aAAa

Architectural Acoustics, Noise, Physical Acoustics, and Engineering Acoustics: Microperforated Acoustical Absorbing Materials

Christian Nocke, Cochair
Akustikbuero Oldenburg, Katharinenstr. 10, 26121 Oldenburg, Germany

Ning Xiang, Cochair
Rensselaer Polytechic Inst., Architecture, 110 8th St., Troy NY 12180

Chair's Introduction—8:00

Invited Papers

8:05

2aAAa1. The perforated panel resonator with flexible tube bundle. Yadong Lu (Inst. of Acoust., Chinese Acad. of Sci., 17 Zhongguancun St., P.O. Box 2712, Beijing 100080, PROC, yadong@mail.ioa.ac.cn), Huide Tang (Tianjin Hearing-aid Factory, Tianjin 300161, PROC), Jing Tian (Chinese Acad. of Sci., Beijing 100080, PROC), and Hongqi Li (Beijing Univ. of Technol., Beijing 100022, PROC)

This paper deals with a new type of resonant absorber, i.e., the perforated panel resonator with flexible tube bundle. The acoustical properties of the new sound-absorbing structure are measured and analyzed. From the measurement results and theoretical analysis, the following conclusions can be made. (1) The incorporation of the tube bundle into the resonator absorber is helpful to improve its low-frequency sound absorption. Because the artificial prolongation of air column length inside a hole neck will increase acoustic mass, resonance frequency can be shifted to a lower frequency correspondingly. (2) Compared with traditional perforated panel absorber, the acoustic resistance of the sound-absorbing structure can be increased considerably due to the incorporation of the flexible tube bundle into the resonator absorber, so that higher sound absorption coefficients can be expected. (3) By reasonably utilizing the tube and cavity coupling resonances, the frequency bandwidth of sound absorption of the perforated panel resonator with flexible tube bundle can be expected to broaden. The perforated panel resonator with flexible tube bundle may be used in some cases with strict space limitations and with expectations of increasing the sound absorption at lower frequencies.

8:30

2aAAa2. Measurements and calculations on microperforated sound absorbers. Christian Nocke and Catja Hilge (Akustikbuero Oldenburg, Katharinenstrasse 10, 26121 Oldenburg, Germany)

The sound absorption of microperforated absorbers can be predicted with a high degree of precision. Microperforated absorbers have been made of various materials in many applications. In recent years different manufacturers have offered a variety of microperforated materials. This paper reviews the first microperforated sound absorbers made of foils introduced some 10 years ago. This paper compares measured sound absorption coefficients of the microperforated foil absorber with the theoretical simulations. Theoretical approaches towards layered sound absorbers with microperforated components will be suggested. Finally, future developments in the field of microperforation in room acoustics will be discussed.

8:55

2aAAa3. Practical applications of transparent microperforated and microslit absorbers. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

It has been over 100 years since absorption was introduced as a quantifiable tool in architectural acoustics. Currently, the architectural acoustic community faces ever-changing challenges that require innovative approaches to absorb sound. Due to the importance of day lighting in sustainable design certification, such as LEED, and the concern about particulate material from fiberglass and mineral wool, there has developed a need for transparent and translucent absorptive materials that do not require the use of fibrous porous absorption. One approach has been the development of microperforated and microslit foils, sheets, and panels with submillimeter openings that provide significant viscous losses when spaced from a boundary. This presentation will review the design theory and absorption mechanisms, describe the various commercial materials now available, and illustrate current projects and potential applications.

9:20

2aAAa4. Effect of honeycomb structure in the back cavity on the absorption characteristics of microperforated panel absorbers. Motoki Yairi, Atsuo Minemura (Kajima Tech. Res. Inst., 2-19-1 Tobitakyu Chofu, Tokyo 182-0036 Japan), Kimihiro Sakagami, and Masayuki Morimoto (Kobe Univ., 657-8501 Kobe, Japan)

Since the principle of microperforated panel absorbers (MPPs) was established, many studies have been made on their practical applications. MPPs are typically made of very thin metal or plastic panels to obtain high absorption. However, such a thin, limp panel is in many cases not suitable for an interior finish of room walls because of its insufficient strength. The authors have proposed to use a honeycomb structure attached behind an MPP to stiffen it without deteriorating its acoustical performance. The honeycomb behind an MPP can also be expected to enhance its absorption performance, considering previous studies on porous sound-absorbing layers with the back cavity completely partitioned by a honeycomb structure. In this paper, the effect of honeycomb structure in the back cavity on the absorption coefficient of MPP was experimentally studied. The results suggest that the reverberation sound absorption characteristics become close to those of an MPP without a honeycomb structure for normal incidence. This means that the characteristics are improved with the honeycomb structure to show a higher and broader absorption peak. MPPs can be used more efficiently in more various situations with applying elaborated configurations of backing structures. [Work supported by the Kajima Foundationfs Research Grant.]

Contributed Paper

9:45

2aAAa5. An ideal absorbing materials in architectural acoustics: Micro-perforated panel absorbers. Ke Liu, Jing Tian, Fenglei Jiao, and Xiaodong Li (Inst. of Acoust., Chinese Acad. of Sci., P.O. Box 2712, Beijing 100080, P.R. China)

Micro-perforated panel absorbers (MPA) can be used to reduce or better the noise in architecture. This absorber was developed in the 1960s, when a robust sound absorber was needed for severe environments, without porous materials. It has a simple structure and the absorption properties can be exactly calculated. The panel can be made of any materials from cardboard to plastic to plywood to sheet metal, for different purposes. Theoretical and experimental investigations on MPA are reviewed in this paper. By reviewing recent research work, this paper reveals a relationship between the maximum absorption coefficient and the limit of the absorption frequency bandwidth. It has been demonstrated that the absorption frequency bandwidth can be extended up to 3 or 4 octaves as the diameters of the micro-holes decrease. This has become possible with the development of the technologies for manufacturing MPA, such as laser drilling, powder metallurgy, welded meshing, and electro-etching to form micrometer-order holes. In this paper, absorption characteristics of such absorbers in random fields are presented and discussed both theoretically and experimentally. This review shows that the MPA has potentials to be an ideal absorbing materials in architectural acoustics in the 21st century.

TUESDAY MORNING, 6 JUNE 2006 ROOM 555AB, 10:25 TO 11:50 A.M.

Session 2aAAb

Architectural Acoustics, Musical Acoustics, and Engineering Acoustics: Surround Sound Essentials I

Alexander U. Case, Cochair
Fermata Audio + Acoustics, P.O. Box 1161, Portsmouth, NH 03802

K. Anthony Hoover, Cochair
Cavanaugh Tocci Associates, 327F Boston Post Rd., Sudbury, MA 01776

Chair's Introduction—10:25

Invited Papers

10:30

2aAAb1. Loudspeakers and listening rooms, Part I: The physical sound field. Floyd E. Toole (Harman International Industries, Inc., 8500 Balboa Blvd., Northridge, CA 91329, ftoole@harman.com)

In contrast to the considerable effort invested in concert halls and auditoria for live performances, the design of domestic listening spaces and recording control rooms has evolved with little guidance from science. In this presentation, an examination of the physical aspects of sound in small rooms reveals that traditional acoustical measures, such as reverberation time, are of little value. The reasons have to do with the differentiating facts that loudspeakers are relatively directional, and the rooms are relatively small, relatively absorptive, and furnished with large sound-scattering objects. As a result, at frequencies above about 300 Hz, listeners are in a transitional sound field consisting of direct and early reflected sounds, and a much-diminished late-reflected sound field. Loudspeaker directional properties and the acoustical behavior of the surfaces involved with early reflections are the factors relevant to perceptions. At low frequencies, room modes dominate. Notions of optimum room-dimensional ratios are of value only when loudspeaker and listener locations are known. New solutions, employing multiple subwoofers in configurations that result in selective modal excitation, can yield excellent low-frequency sound for several listeners in both rectangular and nonrectangular rooms.

10:50

2aAAb2. Loudspeakers and listening rooms, Part II: The perception of reflections. Floyd E. Toole (Harman International Industries, Inc., 8500 Balboa Blvd., Northridge, CA 91329, ftoole@harman.com)

A review of the effects of reflections on perceptions of sound quality, direction, image size, spaciousness, speech intelligibility, etc., lending perspective to the matter of which elements might be perceptually important in multichannel sound reproduction systems. In general they appear to be less problematic than is commonly believed. This, combined with the considerable ability of listeners to adapt to certain auditory influences of small rooms, brings into question several popular practices in the design of loudspeakers and listening rooms. Anechoic measurements on loudspeakers that have been done in a manner that anticipates sound reproduction in rooms can be used to calculate a figure of merit. The correlation between the calculated ratings, and those resulting from double-blind listening tests, is high: 0.86. This has implications for the acoustical treatment of rooms, and to the necessity for controlling the sound field at low frequencies. Only then can there be any assurance that the art in recordings will be heard by consumers as it was created in control rooms.

11:10

2aAAb3. Surround monitoring challenges in acoustically small spaces. Russ Berger (Russ Berger Design Group, 4006 Belt Line Rd., Ste. 160, Addison, TX 75001)

Critical listening spaces used to monitor and produce surround sound product are usually required to provide an accurate listening environment for more than one listener. Decisions are made not only about source material content and veracity, but also about how best to process and present the multichannel surround event in a manner that translates the aural experience to a wider audience. Oftentimes this audience is listening in other, less-than-ideal acoustical environments. Practical criteria and solutions are presented here to help overcome the limitations of monitoring surround sound in acoustically small spaces, as well as a discussion regarding first reflection control through the application of space coupling diffusors.

11:30

2aAAb4. Design considerations for an idealized domestic surround-sound listening space. David Moulton (Moulton Labs., 39 Ames Rd., Groton, MA 01450, dmoulton@moultonlabs.com)

I have been actively involved in the production of music for surround-sound systems since the late 1960s, including composition, production, recording (classical, multichannel pop, and experimental), loudspeaker design, and acoustical design. Based on my experiences, I am going to present some design principles (and constraints) for domestic (and public) surround-sound playback spaces, possible loudspeaker configurations, and topologies. I will also suggest a possible ideal domestic music playback space.

TUESDAY MORNING, 6 JUNE 2006 ROOM 552A, 8:25 A.M. TO 12:00 NOON

Session 2aAO

Acoustical Oceanography, Underwater Acoustics, and Engineering Acoustics: Ultra-Low and Low Frequency Marine Seismo-Acoustics

Ralph A. Stephen, Cochair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543-1542

Peter M. Mikhalevsky, Cochair
Science Applications International Corp., 4001 N. Fairfax Dr., Arlington, VA 22203

Invited Papers

Chair's Introduction—8:25

8:30

2aAO1. Hydroacoustic observations of the 2004 megathrust earthquake and other events. Catherine de Groot-Hedlin (Scripps Inst. of Oceanogr., 9500 Gilman Dr., La Jolla, CA 92093-0225)

Hydroacoustic signals (T-waves) generated by the 2004 Great Sumatra earthquake were recorded by a network of 5 small hydroacoustic arrays located in the Indian Ocean at distances of 2800 to 7000 km from the epicenter. The array configurations allow for accurate determination of the receiver-to-source azimuth given coherent arrivals. Analysis of a series of 10-s windows within the T-wave coda shows that the receiver-to-source azimuth varies smoothly as a function of time, suggesting a nonstationary T-wave source. Apparent motion in the T-wave source has also been observed for many small-scale earthquakes and, for those events, reflects T-wave excitation over a broad expanse of the seafloor. For small-scale earthquakes, it is shown that the variations in the source

azimuth are predictable using a seafloor scattering model. The azimuthal variations for the great Sumatra event are shown to be inconsistent with a small-scale source and are thus indicative of the rupture velocity. The data indicate that the rupture proceeded in two distinct phases; initially it progressed northwest at approximately 2.4 km/s along the Sunda trench. At 600 km from the epicenter, the rupture slowed to approximately 1.5 km/s.

8:50

2aAO2. Use of low-frequency marine seismo-acoustics in understanding earthquake processes: Applications to the Great Sumatra-Andaman earthquake. Maya Tolstoy and DelWayne R. Bohnenstiehl (Lamont-Doherty Earth Observatory of Columbia Univ., 61 Rte. 9W, Palisades, NY 10964-8000, tolstoy@ldeo.columbia.edu)

Low-frequency (1–100 Hz) hydroacoustic recordings can provide a wealth of information on earthquake characteristics in the shallow marine environment. This frequency band, considered low frequency for acousticians, is actually high frequency for regional or global seismologists who commonly work with frequencies of a few Hz and below. In some cases this higher frequency data may provide higher resolution of some earthquake characteristics than traditional seismic methods. The 2004 Great Sumatra-Andaman earthquake provides an example where detailed analysis of the T-wave arrival has allowed detailed characterization of changes in rupture velocity that have not proven easy to resolve in traditional seismic data. Our single-station interpretation of the T-wave data shows that the fault ruptured northward from the epicenter in at least two phases with lower rupture velocity in the north. The methods behind this single station approach will be discussed along with possible future uses of this type of data for improving our understanding of earthquake rupture processes, as well as providing a rapid assessment of the size, location, and geographic extent of large submarine earthquakes, and hence their tsunamigenic potential.

9:10

2aAO3. Fault dynamics of Indian Ocean earthquakes derived from the array analysis of T-waves. Jay Pulli and Zachary Upton (BBN Technologies, 1300 North 17th St., Ste. 400, Arlington, VA 22209)

The Great Sumatran earthquake of 26 December 2004 generated some of the largest and longest T-waves ever recorded, and certainly the largest T-waves recorded on modern high-dynamic range hydrophones. For example, the recordings at the Diego Garcia hydroacoustic array indicate a T-wave duration of approximately 15 min. Array processing of these signals by many investigators indicates that the T-wave generation started near the earthquake epicenter and propagated northwest along the trench for nearly 600 km. This corresponds to the faulting zone defined by the aftershock distribution. A dataset of T-wave recordings of Indian Ocean earthquakes has been examined, covering the period 2001 to the present, including events on the oceans ridges and trenches, as well as some interplate events. The recordings were made at Diego Garcia, Cape Leeuwin, and Crozet Island. Event sizes range from magnitude 4 to 8. Bearing versus time is derived via intersensor cross correlations with a time resolution of 0.010 s, corresponding to a bearing resolution of 0.4 deg. Examples of events along the Sumatran Trench are shown with extended T-wave source areas, from which estimates of fault length and rupture velocity are made. These studies further demonstrate the utility of T-phases for understanding seismic source dynamics.

9:30

2aAO4. Using T-phase spectra to estimate rupture depth and static offset of the seafloor. David Salzberg (SAIC/OSD, 1710 SAIC Dr, M/S 1-11-15, McLean, VA 22102, david.h.salzberg@saic.com)

Analysis of *T*-phases recorded from recent tsunamogenic earthquakes indicates that they contain significantly more high-frequency ($>$50 Hz) relative to low-frequency ($>$10 Hz) *T*-phase energy with longer durations ($>$30 s) than nontsunamogenic events. Modeling *T*-phase spectrum explains this observation, where the spectrum of the recorded *T*-phase is the product of the source spectrum, propagation loss in the solid earth, conversion from seismic to acoustic *T*-phase energy, propagation loss in the ocean, and the instrument response. The propagation loss in the ocean can be accurately modeled. Assuming the seismic-to-acoustic conversion is constant over frequency, the remaining unknown term is loss in the solid earth which is primarily due to anelastic attenuation. A refined depth estimate can be obtained by modeling the solid-earth attenuation. That, coupled with the earthquake source mechanism, allows for estimation of the static ocean floor deformation and thus tsunami height. This approach has been applied to two tsunamogenic earthquakes (26, Dec. 2004, 28, Mar. 2005), and one other event. Both tsunamogenic events had significant ($>$1 m) seafloor uplift. In addition, the peak static offset estimated for the 26 Dec. event is consistent with estimates both from other studies, and from observations of the seafloor deformation.

9:50

2aAO5. Sumatra tsunamis recorded on hydophones and seismic stations in the Indian Ocean. Jeffrey A. Hanson and J. Roger Bowman (Science Applications International Corp., 10260 Campus Point Dr., San Diego, CA 92121)

Tsunamis generated by the great Indonesian earthquakes of 26 December 2004 (Mw 9.3) and 28 March 2005 (Mw 8.6) produced high-frequency dispersed signals (5 to 60 MHz) recorded by hydrophone and seismic stations in and around the Indian Ocean. For the first and greater earthquake, the dispersed energy is frequently observed to 30 MHz and in one case to 60 MHz. The high-frequency signals are consistent with pointlike sources. The earliest-arriving tsunami signal is located to a source at 4.3N, 93.8E using event-to-station distances estimated by measuring dispersion and azimuth estimates using an array of hydrophones. Fine structure in the tsunami signal indicates a second high-frequency source just south of Great Nicobar Island. The dispersion of later-arriving energy matches predictions for reflections of the tsunami from bathymetric features in the Indian Ocean basin. For the 28 March 2005 tsunami, the high-frequency dispersion is observed at the Diego Garcia hydrophone station and the AIS seismic station, and tsunami signals without apparent dispersion are seen at four other seismic stations. Phase velocities are estimated at hydroacoustic arrays and agree with linear dispersion theory at frequencies above 12 MHz.

10:10–10:20 Break

10:20

2aAO6. Surface and body waves from hurricane Katrina observed in California. Peter Gerstoft (Marine Physical Lab., Scripps Inst. of Oceanogr., 9500 Gilman, 92093 La Jolla, CA), Michael C Fehler (Los Alamos Natl. Labs., Los Alamos, NM), and Karim Sabra (Scripps Inst. of Oceanogr., La Jolla, CA)

Hurricane Katrina struck land on 29 August 2005 as one of the strongest storms in the United States. The recent availability of continuous seismic data from a large number of stations enables us to characterize features of hurricane-generated noise in detail. By beamforming noise recorded on a distributed seismic array in southern California, we observe and track both the surface and body *P*-waves generated by Katrina in the 4–20-s period (0.05–0.25 Hz) microseism band. As the hurricane made landfall, the longer-period surface waves weakened, indicating that air/ocean/land coupling was a major factor in their generation. We observed *P*-waves that have propagated deep (1100 km) inside the Earth. These *P*-waves can be back-propagated to the hurricane. These findings demonstrate that ocean microseisms can propagate quite far and open the possibility of further use of seismic noise, even at very low signal level.

10:40

2aAO7. Source areas of midocean microseismic noise. Peter D. Bromirski (Integrative Oceanogr. Div., Scripps Inst. of Oceanogr., 9500 Gillman Dr., La Jolla, CA 92093-0209, pbromirski@ucsd.edu)

Comparison of microseismic noise data collected at the ocean-bottom Hawaii-2 Observatory (H2O), located midway between Hawaii and California, with wind and wave data shows three frequency bands having different source locations. At frequencies from about 0.3 to 0.5 Hz, the noise spectrum shows a strong correlation with overhead wind speed and direction, implying that this energy is generated locally. Correlation of swell height above H2O with microseism energy at frequencies below 0.3 Hz is generally poor, implying that these signals originate at distant locations. The highest noise levels observed occur in the 0.2- to 0.3-Hz band, apparently resulting from the interaction of waves forced by pre- and postfrontal winds from relatively large coincident storms in the same vicinity. Correlation of the H2O microseism levels with NOAA buoy data, with hindcast wave height data over the North Pacific, and with seismic data from mainland and island stations, defines likely source areas of the 0.1- to 0.2-Hz signals. Most of the microseism energy at H2O between 0.08 and 0.2 Hz appears to be generated by high-amplitude storm waves impacting long stretches of coastline nearly simultaneously. [Work supported by the California Department of Boating and Waterways.]

Contributed Papers

11:00

2aAO8. Sound of hurricane. Konstantin Naugolnykh (ESRL/Zeltech, 325 Broadway, Boulder, CO)

The strong perturbation of atmosphere such as hurricane is accompanied by infrasound radiation. This radiation can be observed at a distance of thousands of miles from the hurricane. The effect of sound radiation is connected apparently to the interaction of the counterpropagating sea-surface waves that produces a sound radiation of the doubled frequency of the surface wave oscillation. [M. S. Longuet-Higgins, A theory of the origin of microseisms]. Theoretical model of the sound generation by the sea-surface waves is developed on the basis of the advanced theoretical approach [Wright *et al.*, ''Hurricane directional spectrum spatial variation in the open ocean,'' J. Phys. Ocean. **31**, 2472–2488 (2000)]. The spectrum of sound radiation is presented as a function of the sea-surface wave features. Some estimation of the parameters of infrasound produced by hurricane is made using the data of Wright *et al.* on the hurricane-produced sea-surface waves spectrum. The diffraction effects are taken into account to evaluate the trapping of the radiation by a horizontal atmospheric waveguide, important for the long-range propagation of the noise. [K. A. Naugolnykh, and S. A. Rybak, ''Sound generation due to the interaction of surface waves,'' Acoustical Physics **49**, 88–90 (2003)].

11:15

2aAO9. Anomalous transparency of the water/air interface for low-frequency sound. Oleg A. Godin (CIRES, Univ. of Colorado and NOAA/Earth System Res. Lab., 325 Broadway, Boulder, CO 80305, oleg.godin@noaa.gov)

Sound transmission through water/air interface is normally weak because of a stark density contrast between the two media. For a monopole sound source, at depth larger than a wavelength, the ratio R of the acoustic power transmitted into the air to the total emitted acoustic power is a small quantity of the order of the ratio of impedances of air and water. Acoustic transparency of the interface is shown to rise dramatically at low frequencies, where depth of the monopole source is small compared to the wavelength, with R increasing by up to 36 dB. Under certain conditions, almost all of the acoustic energy emitted by a low-frequency monopole source underwater is predicted to be radiated into the air rather than into the water. Physically, higher transparency at lower frequencies is due to increasing role of inhomogeneous waves emitted by the source and a destructive interference of fields in water due to the actual and image sources. The anomalous transparency of the water/air interface may have significant implications for generation of low-frequency acoustic noise in the air by bubbles collapsing underwater and for acoustic monitoring and detection of powerful underwater explosions for the purposes of the Comprehensive Nuclear Test Ban Treaty.

11:30

2aAO10. Low-frequency sound reflection and conversion to Darcy-type type diffusion waves at bottom interfaces with marine sediments. Jason D. Holmes, Allan D. Pierce, William M. Carey (Boston Univ., Boston, MA 02215), and James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

The upper medium is sea water, and the bottom medium is a marine sediment whose physical description is consistent with the theory advanced by Frenkel in 1944 and with the model proposed by Biot in 1961 for a porous medium at low frequencies. Within the sediment, three generic types (modes) of wavelike disturbances are possible. These three modes are here termed the acoustic, shear, and Darcy modes. The first two are governed by propagating wave equations and the third by a diffusion equation. The first and third are dilatational. All three are nominally uncoupled, although coupling occurs at interfaces. When an obliquely incident acoustic wave impinges from the water on the interface, all three types of disturbances are in principle excited in the sediment, and the details of the excitation are governed by interface conditions, whose derivation is reported in the present paper. Each mode has its own characteristic parameters, and the sediment reflection coefficient, considered as a function of angle of incidence and of frequency, implicitly depends on all of these parameters. The question is addressed as to how, given such reflection coefficient data, one might extract the parameters for the Darcy mode for representative realistic experimental circumstances.

11:45

2aAO11. The relation between received microbarom signals and the sea state under a hurricane. Roger Waxler, Kenneth E. Gilbert, and Carrick L. Talmadge (NCPA, University, MS 38677)

It is known that the peak at about 0.2 Hz observed in both seismic (the microseism peak) and infrasound (the microbarom peak) spectra have their origin in the standing wave component of the ocean surface wave spectrum. The largest source of microseism and microbarom radiation is known to be from the sea states built up under hurricanes. In 1963 Hasselmann published a relation between the ocean wave spectrum and the microseism signal produced by the ocean waves. We have recently derived an analogous expression for the microbarom signal. This expression is presented and possible implications for remote monitoring of hurricane intensity are discussed.

TUESDAY MORNING, 6 JUNE 2006 BALLROOM A, 7:45 A.M. TO 12:10 P.M.

Session 2aBB

Biomedical Ultrasound/Bioresponse to Vibration: Memorial Session for Frederic Lizzi

Jeffrey A. Ketterling, Cochair
Riverside Research Inst., Lizzi Ctr. for Biomedical Engineering Research, 156 William St., New York, NY 10038-2609

Ernest J. Feleppa, Cochair
Riverside Research Inst., Lizzi Ctr. for Biomedical Engineering Research, 156 William St., New York, NY 10038-2609

Chair's Introduction—7:45

Invited Papers

7:50

2aBB1. Spectral parameter imaging of the corneal stroma. Ronald H. Silverman, Omer Gal, Monica Patel, and D. Jackson Coleman (Weill Medical College of Cornell Univ., 1300 York Ave., Rm. K808, New York, NY 10021)

Frederic L. Lizzi was a pioneer in the fields of high-frequency ultrasound and tissue characterization. This report details use of spectral parameter imaging of very-high-frequency radio frequency data for characterization of the cornea. The corneal stroma consists of broad 1–2-micron-thick belts (lamellae) embedded in a proteoglycan matrix. The lamellae consist of collagen fibrils 30 nm in diameter with center-to-center spacing of about 60 nm. Any disturbance to this orderly arrangement will result in loss of optical transparency. Parameter images representing the slope, intercept, and midband fit to the calibrated power spectrum were generated by post-processing of radio frequency data of 35- and 70-MHz scans of the cornea. Stromal backscatter was significantly higher at 70 MHz than at 35 MHz. In an uncomplicated LASIK-treated cornea examined 6 years postsurgery, significant differences in backscatter between the stroma anterior versus posterior to the LASIK interface were observed. This may relate to alteration in lamellar organization, even in the absence of loss of optical transparency. Frequencies beyond 50 MHz offer a probe of tissue microstructure that could be employed in evaluation of stromal organization. [Work supported by NIH Grant EB000238 and Research to Prevent Blindness.]

8:10

2aBB2. Inverse scattering bioacoustics: Under the influence of Fred Lizzi. Michael Insana (Univ. of Illinois, MC 278, 1304 Springfield Ave., Urbana, IL 61801)

A long-standing problem in bioacoustics is the assessment of soft tissue health from an analysis of backscattered echo signals—an inverse problem known as tissue characterization. Pathologists view magnified images of tissue morphology to assess tissue function and grade disease states. Ultrasonic tissue characterization methods are attempts to accomplish the same assessment noninvasively. The goal is not to image the microstructure directly, but rather to diagnose tissues from measurements of statistical moments of the echo spectrum properly normalized to isolate tissue features. Fred Lizzi pioneered this approach and provided an analytical framework used widely. It will be shown how his analysis evolved from classic studies in atmospheric physics to give a clear picture of what is becoming a material science approach to diagnosis using ultrasonics. Dr. Lizzi's methods have been used by many to diagnosis and monitor treatment of cancer from a very small feature space. Others have expanded these methods to discover the sources of ultrasonic scattering in tissue systems and observe dynamic features of physiology. Coupled with elasticity imaging, inverse scattering approaches provide *in vivo* tools for studying the influence of molecular signaling on tissue structure during the progression of cancer and atherosclerotic diseases.

8:30

2aBB3. Stimulation of collagen production with low-intensity acoustics. James Greenleaf, Heather Argedine, and Mark Bolander (Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55901)

Low-intensity pulsed ultrasound is a commonly prescribed therapy for delayed unions and nonunions after fractures. Several prospective double-blind clinical trials have shown that a 1.5-MHz ultrasound signal at a 200-s tone burst repeating at 1 kHz shortens the time to clinically evaluated healing by 30%. We have shown *in vitro* that pulsed ultrasound increases aggrecan gene expression and aggrecan production in chondrocytes, a key cell in the fracture healing process. Because the square wave pulsed 1.5-MHz signal produces radiation force vibration at 1 kHz, we tested the hypothesis that dynamic radiation force, not ultrasound, is responsible for the biological effect of the signal. Experiments showed that a 1-kHz acoustic square wave, simulating the radiation force caused by the clinically used signal, induced chondrogenesis similarly to pulsed ultrasound treatment in ATDC5 cells. These results may have implications for remote ultrasound or acoustic stimulation of different types of strain-sensitive cells in addition to those responsible for fracture healing.

8:50

2aBB4. Noninvasive prostate surgery using high-intensity focused ultrasound. Narendra T. Sanghvi (Focus Surgery, Inc., 3940 Pendleton Way, Indianapolis, IN 46226)

The salient features of ultrasound are to provide both medical diagnostic and therapeutic applications due to its short wavelength in soft tissue. These combinations have motivated us to apply ultrasound for the noninvasive surgery of the prostate diseases for both benign and malignant tumors. Over a decade of engineering and clinical research a successful transrectal ultrasound device, the Sonablate, has been developed and clinically practiced for the treatment of localized prostate cancer as well as benign prostatic hyperplasia. The main advantages of using the Sonablate device are on-line imaging of the prostate in 2D and 3D, accurate therapy planning, and real-time monitoring of treatment. The long-term, over 6 years follow-up clinical results from two multicenter clinical studies indicate no severe long-lasting side effects and acceptable efficacy of the treatment. The primary benefits of this treatment are fewer complications, outpatient procedure, and very quick recovery. The presentation will include design of the device, characterization of therapy and imaging transducer, and treatment parameters for the effective clinical results.

9:10

2aBB5. Quantitative ultrasound techniques. Timothy Hall, James Zagzebski, and Ernest Madsen (Univ. of Wisconsin, 1530 MSC, 1300 University Ave., Madison, WI 53706)

The ultrasound research group at UW-Madison has been pursuing quantitative measurement of the acoustic properties of tissues for more than three decades. Much of that effort parallels the efforts pioneered by Fred Lizzi and colleagues at Riverside Research Institute (RRI). While the RRI group was blazing trails in data acquisition, analysis, and interpretation techniques, the UW group was investigating the limits of measurement accuracy using similar approaches. Wise selection of the tissues to investigate, great intuition regarding fruitful paths, and good hard work resulted in substantial progress and success in the RRI quantitative ultrasound efforts. Careful measurements with high-quality phantoms, in parallel with tissue measurements, resulted in substantial progress in determining limits on quantitative ultrasound measurement accuracy under various experimental conditions. This presentation will highlight some of the differences and similarities in the approaches to quantitative ultrasound by these two groups and how we can use that information in future investigation.

9:30

2aBB6. Therapeutic applications of ultrasound in intracellular drug delivery and tissue ablation. Cheri X. Deng (Dept. of Biomed. Eng., Case Western Reserve Univ., 10900 Euclid Ave., Cleveland, OH 44106-7207)

Biomedical application of ultrasound has been advancing beyond diagnostic imaging. Innovative strategies of therapeutic applications of ultrasound have been exploited that utilize both thermal and mechanical effects generated by ultrasound. For ultrasound mediated intracellular drug and gene delivery, ultrasound has been used to increase cell permeability transiently. We have developed novel techniques to investigate the mechanisms of ultrasound-induced cell porosity or sonoporation in real time at the single cell level. We studied the dynamics of sonoporation and correlation with delivery outcome in mammalian cells, bacteria, and tissues. Applications of high-intensity focused ultrasound (HIFU) include tissue thermal ablation and other applications. We have developed interdisciplinary techniques using real-time fluorescent imaging, acoustic radiation elastography, and optical coherent tomography (OCT) techniques to study the changes of electrophysiology, mechanical, and optical properties of cardiac tissues during focal HIFU cardiac ablation. [Work supported by NIH, ACS, and NSF.]

9:50

2aBB7. Enhanced drug delivery with ultrasound and engineered delivery vehicles. Katherine W. Ferrara, Mark A. Borden, Paul A. Dayton, Azi Kheirolomoom, Dustin E. Kruse, Michaelann Tartis, and Aaron F. Lum (Dept. of Biomed. Eng., UC Davis 451 East Health Sci. Dr., Davis, CA 95616)

We demonstrate that local drug delivery can be achieved by ultrasound, combined with engineered delivery vehicles, where the vehicles have a diameter on the order of nanometers to microns. Delivery vehicles can be created from microbubbles with a thickened shell or a lipid shell decorated with drugs, genes, or nanoparticles. Alternatively, liquid-filled nanoparticles can be employed to carry the desired compound. Ultrasound can deflect these vehicles from the center of the flowstream, can fragment the vehicle releasing its

contents, and may enhance the uptake of the particle or its contents by cells in the desired region. The ultrasonic mechanisms behind these changes are summarized. The addition of targeting ligands to the shell to improve target specificity is also explored. Methods to measure the effectiveness of local drug delivery, including correlative imaging modalities, binding assays, and cytotoxicity assays, will be described. [The support of NIH CA 103828 is gratefully acknowledged.]

10:10–10:25 Break

10:25

2aBB8. Tissue characterization using novel high-frequency models. Michael Oelze (Dept. of Elec. & Comput. Eng., Univ. of Illinois at Urbana-Champaign, 405 N. Mathews, Urbana, IL 61801, oelze@uiuc.edu)

Conventional models for ultrasound tissue scattering at conventional ultrasound frequencies (less than 15 MHz) were insufficient for differentiating between two kinds of solid tumors in mice that had clear morphological differences. F-tests on backscattered power spectra from the two kinds of tumors revealed significant differences between the two curves for frequencies above 16 MHz. New models were constructed to better reflect the ultrasonic interaction with underlying tissue structure. One model considered only the cell nucleus as a dominant source of scattering. A second model, the whole cell model, considered the nucleus and cytoskeletal structure. These models were used at high ultrasonic frequencies (16 to 70 MHz) to characterize the two kinds of tumors and compared with scatterer estimates using a conventional model, the spherical Gaussian model. The nucleus model revealed significant differences between the two kinds of tumors but scatterer estimates did not match the underlying structure as observed from optical photomicrographs. The whole cell model did not reveal significant differences but scatterer estimates matched underlying structure. [Work supported by NIH Grant F32 CA96419 and start-up funds from the Department of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign.]

10:45

2aBB9. Scattering of high-frequency ultrasound by cells and cell ensembles: In search of the dominant scattering source. Michael Kolios (Dept. of Phys., Ryerson Univ., Toronto, Canada)

High-frequency ultrasound (HFU) imaging (in the range of 20–60 MHz) has been recently enabled by advances in transducer technology and electronics. These high frequencies afford greater image resolution, on the order of 40 microns, at the expense of penetration depth, limited to a few centimeters. The associated ultrasound wavelengths of HFU are 25–75 microns, on the order of the size of cells. Therefore, it is hypothesized that at these frequencies ultrasound tissue characterization will provide more meaningful information about the physical characteristics of cells and any changes that occur in their structure during cancer treatment. The late Dr. Frederic Lizzi pioneered ultrasound tissue characterization and the seminal papers he published provided the impetus for this work. Here, backscatter from individual cells and tightly packed cell aggregates of the same cells (emulating tumors) are examined and compared to theoretical models of ultrasound backscatter. Recent work with a series of cells lines with different sizes and physical characteristics will be presented. It will be shown that for *in vitro* and *in vivo* experiments, the backscatter signal characteristics best correlate with nuclear size.

11:05

2aBB10. Automated coronary plaque characterization with intravascular ultrasound backscatter: *In vivo* and *ex vivo* validation. D. Geoffrey Vince, Anuja Nair, M. Pauliina Margolis, Stacy L. Kirby, Jennifer J. E. Kuznicki (Volcano Corp., Advanced Technology Lab., 9500 Euclid Ave., Cleveland, OH 44195 gvince@volcanocorp.com), Xiaofeng Wang, Barry D. Kuban (The Cleveland Clinic Foundation, Cleveland, OH 44195), and James R. Margolis (Miami International Cardiology Consultants, Miami, FL 33137)

Atherosclerosis is considered both a systemic and local disease. Current diagnostic tools do not allow adequate *in vivo* characterization of lesions. Intravascular ultrasound (IVUS) backscatter analysis has displayed the potential for real-time plaque characterization by advanced spectral analysis. One hundred-fifteen plaques from 51 coronary arteries were imaged *ex vivo* at physiologic pressure, using 20-MHz IVUS transducers. After imaging, the arteries were pressure fixed and matching histology was collected. Regions were selected from histology and corresponding IVUS data were used to build the ''VH'' plaque classification system using spectral analysis and classification trees. VH images were validated *ex vivo* by comparison with histology via 407 selected regions that comprised 162 fibrous, 84 fibro-fatty, 69 necrotic-core, and 92 dense-calcium regions. The overall predictive accuracies were 92.9%, 93.4%, 94.3%, and 99.3% respectively, with sensitivities and specificities ranging from 84% to 99%. Reproducibility was assessed *ex-vivo* and *in vivo* with two pullbacks of IVUS catheters in each. A semiparametric regression model and the *t*-test were used to quantify reproducibility, which was high in each case. VH is a robust technique that provides accurate and reproducible information on plaque composition, with considerable potential for assessment of plaque vulnerability in real time.

Contributed Papers

11:25

2aBB11. Discovery of the ring resonant frequency for noninvasively estimating local elastic modulus of arterial wall. Xiaoming Zhang and James F. Greenleaf (Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905, Zhang.xiaoming@mayo.edu)

Increased stiffness of arteries has recently gained acceptance as an independent risk factor for cardiovascular and many other diseases. Pulse wave velocity (PWV) is widely used for estimating the stiffness of an artery. However, PWV is an average indicator of artery stiffness between the two measuring points, and therefore does not identify local stiffness. In addition, the thickness of the artery is needed to calculate artery elastic modulus. We have discovered that there is a series of ring resonant frequency in arteries that can be used for estimating local arterial elastic modulus. These resonances can be excited by remote palpation with selectively placed ultrasonic radiation force. Experiments on ring resonance were carried out on an excised artery as well as a rubber tube. This family of resonances occurs at relatively low frequency, around 356 Hz for the artery. Estimation of the Young's modulus is 135 kPa by the ring frequency 356 Hz without the requirement of the arterial thickness. The

estimated modulus is very consistent with the other ring resonant frequency measurements. The theory of ring resonance is developed and verified with the rubber tube experiments.

11:40

2aBB12. Imaging with high-frequency annular arrays. Jeffrey A. Ketterling, Sarayu Ramachandran, Ronald H. Silverman (Frederic L. Lizzi Ctr. for Biomed. Eng., Riverside Res. Inst., 156 William St., New York, NY 10038), Orlando Aristizábal, and Daniel H. Turnbull (New York Univ. School of Medicine, New York, NY 10016)

Frederic Lizzi helped pioneer the use of high-frequency ultrasound (HFU) for ophthalmic imaging and disease diagnosis. A limitation of HFU single-element transducers is their limited depth-of-field ($\approx$1 mm). One approach to improve the DOF for HFU applications is to use annular arrays. A five-element, 40-MHz annular array was developed for ophthalmic and small-animal imaging. The array consisted of a 9-μm PVDF film bonded to a patterned, copper-clad polyimide film. An experimental system was assembled to permit the digitization of radio-frequency (rf) data from the array. The rf data were collected from all transmit/receive ring pairs and the data were later postprocessed. The performance of the array was first validated with a wire phantom and then *in vitro* data were acquired from excised bovine eyes and a cadaver eye. *In vivo* data were acquired from an anesthetized rabbit and from mouse embryos. Image comparisons were made between rf data that had been summed with no delay corrections and rf data that had been processed with a synthetic-focusing algorithm. Synthetic-focus images showed a clear improvement in DOF in terms of image brightness and enhanced membrane and cavity-boundary definition. [This research was supported in part by grants from the NIH (EY014371 and NS038461).]

11:55

2aBB13. Ultrasonic and magnetic-resonance spectra in tissue-type imaging for planning and monitoring prostate cancer treatment. Shreedevi Dasgupta, Ernest Feleppa, Jeffrey Ketterling, Sarayu Ramachandran (Riverside Res. Inst., 156 William St., New York, NY 10038), Christopher Porter (Virginia Mason Medical Ctr., Seattle, WA), and Fernando Arias-Mendoza (Columbia Univ., New York, NY)

No reliable method of imaging prostate cancer currently exists. These prostate tissue-typing studies aim to improve the effectiveness of biopsy guidance and treatment monitoring by developing better imaging methods for identifying cancerous prostate tissue. Success will reduce the false-negative rate of biopsies and treatment side-effects. Ultrasonic (US) radio-frequency (rf) echo-signal data acquired during biopsy examinations were used to compute spectral parameters. These spectral parameters along with clinical parameters, e.g., PSA, were used to train a neural network classifier using biopsy results as the gold standard. Cancer-likelihood scores from a look-up table were used to generate tissue-type images (TTIs). The ROC-curve area for US neural-network-based classification was 0.844 $\pm$0.018 vs 0.638$\pm$0.031 for B-mode-based classification, and the sensitivity of neural-network based classification was superior to that of B-mode-guided biopsies. These classification methods are being extended to include magnetic-resonance spectral (MRS) techniques that use the choline to citrate ratio-to-distinguish cancerous from noncancerous prostate tissue. 3D renderings of prostatectomy histology, US, and MR images show encouraging correlations, and combining MRS parameters with US spectral parameters appears to have potential to further improve prostate-cancer imaging. [Work supported in part by NIH Grant CA053561.]

Session 2aED

Education in Acoustics: Hands-On Experiments for High School Students

Uwe J. Hansen, Chair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Students and senior scientists will staff experiment stations for local high school students who will perform the experiments. These students will thus be exposed to a variety of acoustics principles and get some practical hands-on experience with a number of research tools. Regular ASA meeting participants are welcome to the session as long as their participation does not interfere with student hands-on activities.

Session 2aMU

Musical Acoustics: Scaling of Musical Intstrument Families

George A. Bissinger, Chair
East Carolina Univ., Dept. of Physics, Greenville, NC 27858

Chair's Introduction—8:25

Invited Papers

8:30

2aMU1. Violin octet scaling revisited. George Bissinger (Phys. Dept., East Carolina Univ., Greenville, NC 27858, bissingerg@mail.ecu.edu)

Schelleng employed only two major resonances to scale the violin to other pitch regions, creating the violin octet: the main air (now labeled A0) to be placed at $1.5x$ and the main wood (B1, actually two modes) to be placed at $2.25x$, the lowest string frequency. B1 employed flat-plate scaling and an empirical relationship between plate mode frequencies and B1. Experimental modal analysis of a complete octet showed flat-plate scaling for the corpus generally worked well, but A0 scaling was unreliable, even with postassembly modifications. Failure of the rigid-cavity Rayleigh relationship for A0 originated from an obvious major omission, cavity wall compliance, and one unobvious source, A0 coupling with the next higher cavity mode A1, which emerged much later via Shaw's network model. Incorporating wall compliance into Shaw's model greatly improved A0 and A1 predictions to within 10% over the octet. Acoustical analysis combined with patch near-field acoustical holography showed that A1, via induced corpus motion, was the most important radiator for the largest instrument in the main wood region and that the B1 modes, via net volume changes, can radiate significantly through the f-holes, two radiation mechanisms never considered in Schelleng's original scaling.

9:00

2aMU2. Violin octet scaling: The practical side. Robert J. Spear (The New Violin Family Assoc., 42 Taylor Dr., Wolfeboro, NH 03894)

Schelleng' scaling procedure attempts to place A0 at $\sim 1.5x$ and the first corpus bending modes at $\sim 2.25x$, the lowest string pitch. An experiment to move A0 into this position (also influencing A1), using the Stradivari model G as reference, helps produce the characteristic sound of the violin in cases where B1 placement was not optimum. The experiments suggest that the model G worked well for the signature modes of a classic violin, successfully scaling at least four instruments (mezzo, alto, tenor, and a small baritone), but that instruments at the extremes—treble and soprano at one end of the scale and both basses at the other—will likely require a refinement of the model G and placement of modes different from those of the reference violin and from each other. Scaling for acoustics and scaling for size are not the entire answer since ergonomic demands are often at odds with both. Geometrical scaling is essential for aesthetics and the uniform shape of the instruments, but it does not hold up well for rib heights, rib thicknesses, and cavity air volumes. This experiment suggests that significant latitude of air and wood mode placement exists within which good results can occur.

9:30

2aMU3. Practical issues in designing strings for different-sized bowed instruments. Fan-Chia Tao (J. D'Addario & Co., P.O. Box 290, Farmingdale, NY 11735, fan.tao@daddario.com)

The playing length and frequency of musical strings are fixed by the design of the instrument. The playing tensions of commercially available strings for standard bowed instruments follow an approximately linear scaling law in relation to string length. However, the string designer can adjust the playing tension to yield different results, and playing tensions have changed throughout history as string technology and playing techniques have evolved. Another string property that can be varied by the string designer is damping. Empirical data suggest that over a wide variety of bowed instruments and frequency ranges, strings with similar damping give similar playing characteristics.

10:00–10:15 Break

10:15

2aMU4. Scaling of reed instruments: The case of the saxophone family. Jean-Pierre Dalmont, Bruno Gazengel (Laboratoire d'Acoustique de l'Universit'e du Maine, 72085, Le Mans cedex 9, France, jean-pierre.dalmont@univlemans.fr), and Jean Kergomard (CNRS, Marseille Cedex 20, France)

The saxophone family is one of the most extended musical instruments family from the contrabass to the sopranino. The timbre of these instruments is unmistakable: there is a small probability to mix up a saxophone with an oboe or a bassoon. Now a question is: What is so specific in the timbre of the saxophones? Notes of various saxophones are recorded, analyzed, and compared with those of other instruments such as the oboe. More specifically, the transient duration and the formant frequency are determined. Then, the question which physical parameters influence the transient duration and the formants is studied. The influence of geometrical parameters such as the cone angle and the length of the cone truncation will be discussed.

10:45

2aMU5. Scaling of flutelike instruments : An analysis from the point of view of the hydrodynamic instability of the jet. Benoit Fabre, Nicolas Montgermont, and Johan Gabriels (Lab. for Musical Acoust., LAM, Paris 6 Univ., Paris, France)

The sound produced by musical instruments of the flute family is the result of the efforts of the instrument maker as well as those of the player. For instance, different instruments of a recorder consort show, on one hand, dimensions optimized by the maker for their individual compass as well as for a global identity and coherence of the whole consort. On the other hand, the players have to adapt their blowing to each instrument of the consort. Therefore, the study of the different instruments has to include the geometry given by the maker as well as the required blowing conditions. The study reported here presents data gathered on different flute families: the geometries and the instrument passive resonances are studied together with the behavior of the instruments under artificial blowing and, finally, under the control of a musician. The analysis of the data in terms of the jet hydrodynamic instability offers an efficient framework to describe the choices of the maker and the adaptation of the player.

Contributed Papers

11:15

2aMU6. A new family of stringed musical instruments. Sophie Léger and Samuel Gaudet (Dept. of Mathematics, Univ. de Moncton, Moncton, NB, E1A 3E9 Canada)

A new family of stringed musical instruments will be introduced. The Tritare™ is the first member of this new family and has evolved from the mathematical analysis of the vibrations of networks of strings [Gaudet *et al.*, J. Sound & Vib. **281**, 219–234 (2005)]. It is a guitarlike instrument that uses six networks of three string sections (Y-shaped networks) instead of the classical single strings. The sounds that it produces can be more or less harmonic depending on the string attack conditions and the physical configurations of the string networks. A method of customizing the frequency spectrum to contain desired harmonic components will be demonstrated. A numerical model, which uses finite differences, used to simulate the 3D nonlinear vibrations of more complex string networks involving several string sections and junction points will also be described. The harmonic analysis of the frequency spectra resulting from the numerical simulations and experimental measurements will be presented, as well as network configurations which produce very unique and interesting tone colors. Some of these configurations form the basis for a new family of stringed musical instruments. The presentation will include a demonstration of the Tritare™.

11:30

2aMU7. Scaling of brasswind instruments. Robert W. Pyle, Jr. (11 Holworthy Pl., Cambridge, MA 02138, rpyle@post.harvard.edu) and Arnold Myers (Univ. of Edinburgh, Edinburgh EH8 9AG, UK)

Consider two musically similar brasswind instruments pitched an octave apart, like trumpet and trombone. The steady-state tones of the trumpet, shifted downwards in pitch an octave by playing a recording at half speed, sound very similar to those of the trombone. The bore shape of the trumpet, however, is not a half-size dimensional replica of that of the trombone. The trumpet bell is about half the size of the trombone bell, but the narrower parts of the trumpet are considerably more than half the diameter of the corresponding parts of the trombone. Acoustic impedance measurements show the Q of air-column resonances to be similar in both trumpet and trombone. This is consistent with the dimensions of the instruments, combined with the dependence of viscous and thermal damping on tubing diameter and frequency. (Viscous and thermal damping dominate radiation damping throughout the normal playing range.) These observations suggest that a plausible scaling law for a family of brasswinds would preserve the pitch-shifted spectrum and the Q of the resonances across members of the family. How well (or poorly) brass makers follow this in practice will be shown from measurements of dimensions and acoustic input impedance on a variety of instruments.

Session 2aNSa

Noise and Architectural Acoustics: Audio-Visual Design in Soundscapes I

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Technical Acoustics, Einsteinufer 25, 10587 Berlin, Germany

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

Chair's Introduction—9:30

Invited Papers

9:35

2aNSa1. Soundscape measurement and analysis. Klaus Genuit (HEAD Acoustics GmbH, Herzogenrath, Germany) and Wade Bray (HEAD Acoustics, Inc., Brighton, MI 48116)

Soundscapes typically include a variety of sources, with individual and collective attributes (level, spectral structure, time structure, and perceived characteristics such as loudness, sharpness, fluctuation, and roughness). Relevant sound sources impinge on listeners from multiple directions and distances. Examples will be given of large differences between conventional and psychoacoustic values as a function of distance, and of multi-source summations. As an example, the subjective loudness difference for sound incidence over a small versus large solid angle for a constant sound pressure level and spectrum is known. Individual-source-relevant information, important in context, may be unrecognized unless spatial hearing is included in soundscape assessment. A demonstration of binaural versus monaural soundscape data acquisition, and subsequent analysis by both conventional and psychoacoustic measures considering human perception including time dependency, will be given. Evidence will be presented for considering human hearing not only in the analysis, but also in the acquisition.

9:55

2aNSa2. Soundscape measurement in an urban area. Brigitte Schulte-Fortkamp (TU-Berlin, Einsteinufer 25, 10587 Berlin, Germany)

This presentation is about a demonstration of soundscape evaluation of an urban area with respect to measurements and interviews. In a recent field study in Berlin, indoor and outdoor measurements were carried out in a residential area in which inhabitants were complaining about noise and vibration due to heavy traffic over day and until late at night. In this context, the question of the appropriate positioning and calibration of the measuring devices in indoor and outdoor measurements was very important. The measurement procedure as well as procedures which determine how to conduct interviews will be described and discussed by taking into account how to make progress in measuring soundscapes. Moreover, the focus will be on necessary contributions to a measurement which usually deals with dB(A) and lacks psychoacoustical measuring factors.

10:15

2aNSa3. Traditional measurement methods for characterizing soundscapes. Bennett M. Brooks (Brooks Acoustics Corp., 27 Hartford Turnpike, Vernon, CT 06066, bbrooks@brooksacoustics.com)

A soundscape may be defined as the acoustical environment at an outdoor location. The physical sound occurring at the location may be from natural or man-made sources, or both. Perception of the soundscape can provide comfort and needed information to the listener, or may be annoying. The existing soundscape may be subject to design modifications for a variety of purposes. The goal could be to protect or enhance some sounds, such as the laughter of children and chirps of wildlife in a park, or the speech in an amphitheater. Conversely, the goal may be to reduce or eliminate unwanted sounds, such as transportation or industrial noise. Engineering and aesthetic soundscape design may proceed based on the characterization of the acoustical environment. The soundscape can be characterized by a number of measured acoustical parameters. Traditionally, these parameters have included A-weighted sound level, spectral content, sound-level statistics and averages, alone or in combination, and sound-level time history. Listener perceptions and nonacoustical parameters, such as time of day or year, community relevance, and listener attitudes may also be applied to soundscape characterization and design. Examples of several different soundscapes, their characterizations, and proposed design modifications will be discussed.

10:35

2aNSa4. Documenting visual context for soundscapes. Tim Lavallee (LPES, Inc., 14053 Lawnes Creek Rd., Smithfield, VA 23430), Vincent Catania (38 Temple St., Boston, MA 02114), and Robert Kull (Parsons, Norfolk, VA 23502)

There are several considerations to make when documenting a given soundscape. Not only is the technical sound information important, but the space, receptor, and visual context are important as well. This is primarily important in furthering the ability of sound and soundscape professionals to communicate with architects, planners, and other engineers—and vice versa. A portable digital

recording system was assembled and used along a high-end digital camera to document a range of soundscapes. These tests were preformed in urban, suburban, and rural areas at different times and in different contextual settings. Both visual and audio information of a locale were documented. Although this method can be simple or complex, the combination of these two components highlights the human factors that are present for a given soundscape. This will be a discussion of the portable equipment and a presentation of sound recordings, spectra, and photographs of a range of soundscapes. The variety and complexity of soundscapes and the benefits of visual clues to add depth to a given documentation effort will be highlighted.

Contributed Papers

10:55

2aNSa5. Soundscape analysis and acoustical design strategies for an urban village development. Youngmin Kwon, Pattra Smitthakorn, and Gary W. Siebein (Architecture Technol. Res. Ctr., School of Architecture, Univ. of Florida, 134 ARCH, P.O. Box 115702, Gainesville, FL 32611, ymkwon@ufl.edu)

Overall establishment of an acoustical landscape (soundscape) would be the fundamental source of sonic information that is useful in initiating urban design and planning strategies. A soundscape analysis was conducted in conjunction with an urban transportation and community design project targeting the western section of Gainesville, FL. The community was acoustically characterized and analyzed by means of quantitative measurements and qualitative assessments by observation. The measurements included instantaneous, 1-min short-term and 24-h long-term measurements. The parameters involved in these measurements are A-weighted average Lp, Leq, Lmax, Lmin, and LDN. The overall sonic environment of the community as well as the types of sounds that are desirable and noises that are not desireble in the existing and future contexts were identified. The results showed that the community was dominated mostly by traffic noise from several major thoroughfares and by occasional fire alarm noise. Finally, methods or strategies to reduce, buffer, and mitigate the undesirable sounds and to preserve the desirable sounds were suggested. The addition of new characteristics to the soundscape, such as people activity sounds and ''natural'' sounds, were also proposed. The application of acoustical urban design strategies to the specifically proposed urban transportation design plans was further analyzed.

11:10

2aNSa6. Territoriality and sound delimitation of urban space. Cobis Rabah Derbal and Hamza Zeghlache (BP 12 SMK, Constantine 25003, Algeria, cobis.derbal@free.fr)

Whether the urban space is physically materialized or abstracted it is often delimited by perceptive and sociocultural relations. In the present methodological approach, urban space is supposed to be modulated in its delimitation by qualitative and quantitative criteria of its noise environments. This introduces clearly the concept of sound delimitation and of noise environment territoriality. One can notice that sound phenomena are very present and are marking the social practice and behavior in urban spaces and give them a very specific sonority. Just like the visual or sociocultural aspects, the sonority of the public place or the urban noise environment marks their territoriality and spatiality by creating a whole sound environment characteristic of dimensions and forms of the space. The urban noise can thus be an element of marking, connecting, breaking, or structuring spaces which are morphologically and visually irreconcilable. One can speak about sound delimitation of urban space beyond its morphological limits. That delimitation can well articulate the social practice of urban space to its dimensional aspect. The study put out the concept of acoustic space in opposition to morphological space which it is inevitably necessary to reconcile in order to better practice the urban space.

11:25

2aNSa7. Recomposition of the urban sonic environment. Cobis Rabah Derbal and Hamza Zeghlache (BP 12 SMK, Constantine 25003, Algeria, cobis.derbal@free.fr)

Acoustics-related considerations are taken to be secondary in most of the design processes of public places. This fact always leads to an unpredictable noise and urban sonic environment around such spaces. Studies have shown the importance of the noise and the quality of the urban sonic environment in the social practice and in the daily use of public places. These places strongly mark the degree of dynamism of the urban structure of our cities and are often subjected to a very complex situation of noise nuisance. The main aim of the study is to introduce a source of noise that might have an effect of recomposing positively a deficient sonic environment of such places. This approach is also to try to restore the sound in its utilitarian and practical function and to articulate the perception of the spatial and temporal aspects of the public place. It could be an element of transforming urban noise into a peaceful soundscape. A fountain turned out to be that sonic source and it can be easily introduced in the urban environment around the public places. It is a sort of urban instrumentation used in order to recompose or requalify an urban noise environment.

11:40–12:00
Panel Discussion

12:30
Session participants meet in room 554AB for experimental field measurements

1:00–3:00
Experimental field measurements conducted

Note: Panel discussion will be held in Session 2pNSb at 3:00 p.m.

Session 2aNSb

Noise, Education in Acoustics, Engineering Acoustics, and Architectural Acoustics: Essential Acoustical Curriculum for Noise Control Techniques for Engineers, Industrial Hygienists, and Architects

William J. Murphy, Cochair
NIOSH, 4676 Columbia Pkwy., Cincinnati, OH 45226-1998

Ralph T. Muehleisen, Cochair
Illinois Inst. of Technology, Civil and Architectural Engineering, 3201 S. Dearborn, Chicago, IL 60616

Chair's Introduction—8:50

Invited Papers

8:55

2aNSb1. Essential elements of acoustical curricula for noise control engineers. Robert J. Bernhard (Ray W. Herrick Labs., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907, bernhard@purdue.edu)

Because the supply of noise control and acoustical engineers is small, many of the noise control practitioners practicing in general mechanical engineering fields are retrained on the job. These practitioners are often capable in the area in which they are mentored but lack sufficient background to move to new areas. In order to address a broad range of noise control issues, an engineer must be comfortable to work in waves, or modes, or dBese (engineering controls). The essential curriculum for engineering practitioners would include a foundation in each of these approaches as well as the analytical, computational, and experimental background upon which to build engineering solutions. This presentation will include discussion of how these objectives can be met at the undergraduate level, at the graduate level, and through continuing education.

9:25

2aNSb2. The need for an essential acoustics curriculum and its delivery via distance education. Victor W. Sparrow, Anthony A. Atchley, Karen P. Brooks, and Gerald C. Lauchle (Grad. Program in Acoust., Penn State, 214 Appl. Sci. Bldg., University Park, PA 16802, vws1@psu.edu)

Many individuals have identified a need to better educate engineers, industrial hygienists, and architects with an essential acoustical curriculum including noise control techniques. There are less than a dozen major programs training these individuals, and demand far exceeds the supply of education suppliers. One effective method of providing education to busy professionals who otherwise would not be exposed to an essential curriculum is distance education. Approximately 10 years ago Penn State developed a series of noise control engineering courses delivered through CD-ROM. Now with DSL and cable modems more common, full motion video streaming is a preferred delivery method. Students currently take graduate level courses in acoustics at Penn State via resident education, continuing education, and distance education. There is no reason why an essential acoustical curriculum could not be delivered similarly. One benefit of video streaming is an opportunity for both live and archived lecture material. Students can attend live recitation sections while retaining the ability to see archived material on demand. Technology such as high-bandwidth connections, inexpensive webcams, and microphones also allows two-way audio and video for direct student-instructor interaction.

9:55

2aNSb3. The acoustical education of an architect. Gary Siebein (School of Architecture, Univ. of Florida, 625 NW 60th St., Ste. C, Gainesville, FL 32607)

Sound can provide valuable contributions to the education and practice of architecture. First, architecture students should develop a theory of how sound and other environmental systems can become part of a larger theory of the making of architecture. Second, architects are trained to have highly developed sensibilities in the aesthetic design of space. Expanding the aesthetic component of buildings and environments to include the aural environment has the potential to enrich the fabric of communities and buildings. Third, architecture students must become acquainted with the materials, analytic techniques, and design approaches that acoustical consultants use. Furthermore, for acoustical information to become an integral part of architectural design and education, it must be organized in a way that allows expression or poesis to occur because these aesthetic principles are the building blocks of architectural design. Therefore, the science and engineering principles of noise control and architectural acoustics must be transformed into material that can become spatial, visual, and manipulable in creative ways by a design team. This requires that the basics of acoustics be reformulated in an architectural manner. This paper will address several of the transformations that have been attempted in the architectural curriculum at the University of Florida.

10:25–10:40 Break

10:40

2aNSb4. Acoustics curriculum for architectural engineers. Ralph T. Muehleisen (Civil and Architectural Eng., Illinois Inst. of Technol., 3201 S. Dearborn, Chicago, IL, 60616, muehleisen@iit.edu)

As building design and technology increase in complexity, the need to consider acoustics in the design also increases. For decades, architects have relied on engineers for the design of structural, mechanical, electrical, and lighting systems of buildings, and are increasingly relying on engineers for the design of high-efficiency building enclosures and natural lighting systems. At the same time, building systems have become increasingly integrated, and good design of the building systems requires a very interdisciplinary education. These needs have given rise to the formal discipline known as architectural engineering (AE). With education in architecture, structures, building systems, and construction management, the architectural engineer is uniquely qualified to begin a career as an acoustical consultant. There are now over a dozen ABET-accredited AE programs throughout the country and most offer at least some education in architectural acoustics and noise control. In this talk, the requirements of an acoustics education for architectural engineers are outlined and the curriculum of several AE programs is presented.

11:10

2aNSb5. Industrial noise control engineering principles and standard practices applicable to engineers and industrial hygienists. Robert R. Anderson (James, Anderson & Associates, Inc., 2123 University Park Dr., Ste. #130, Okemos, MI 48864)

Industrial organizations have accepted the notion that noise control is no longer 100% of one person's responsibility, but a collaborative effort among engineers and industrial hygiene professionals. Proven noise control measures exist and are commercially available for many industrial noise sources. There are also noise control concepts and techniques that can be adopted by maintenance and other personnel as standard practices. Awareness of the elements of noise control engineering can offer benefit in areas such as equipment procurement, workplace risk assessment, and manufacturing program planning. The focus of this presentation is to discuss the elements of noise control engineering that manufacturing engineers and industrial hygiene professionals would need to apply out of their university education into practice, and the benefits that would be derived from this information.

11:40–12:00
Panel Discussion

TUESDAY MORNING, 6 JUNE 2006 ROOM 550AB, 9:00 TO 11:45 A.M.

Session 2aPA

Physical Acoustics: Outdoor Sound Propagation

Richard Raspet, Cochair
Univ. of Mississippi, National Ctr. for Physical Acoustics, Coliseum Dr., University, MS 38677

James P. Chambers, Cochair
Univ. of Mississippi, National Ctr. for Physical Acoustics, Coliseum Dr., University, MS 38677

Contributed Papers

9:00

2aPA1. Ground-to-ground sound propagation in cloud-topped boundary layers over land. Roger Waxler, Carrick L. Talmadge, and Kenneth E. Gilbert (NCPA, University, MS 38677)

Sound propagation under cloud-covered skies is discussed. Observations are presented which suggest that under dense cloud cover ground-to-ground sound propagation is dominated by a single mode similar to the classic surface mode for outdoor sound propagation. A theory of sound propagation under cloud covered skies has been developed. It is shown that there is a sound duct centered in the cloud layer. Of the modes in this duct a small number (typically one or two for frequencies up to a few hundred Hz) of ground interacting modes, modes whose amplitudes on the ground are significant, are identified. Over land, one of the ground-interacting modes can be identified with the surface mode.

9:15

2aPA2. A continuously cascaded conformal mapping for sound propagation over irregular terrain. Santosh Parakkal, Kenneth E. Gilbert, and Di Xiao (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, Coliseum Dr., Universtiy, MS 38677)

The so-called earth flattening conformal mapping is a common approach for radio wave propagation over a spherical earth. Applied sequentially (cascaded), it can be used to represent a series of hills, each of which has a constant radius of curvature. We show that the mapping can also be applied to general terrain that has a continuously varying radius of curvature. In the mapped coordinates, the terrain is flat, but the mapped sound speed is $c/\sqrt{J}$, where c is the actual sound speed and J is the Jacobian of the transformation. The continuously cascaded mapping method is outlined, and numerical comparisons are made with a well-tested nonconformal mapping method.

9:30

2aPA3. The absorption of sound on Mars using the direct simulation Monte Carlo. Amanda D. Hanford and Lyle N. Long (Grad. Program in Acoust., The Penn State Univ., University Park, PA 16802, ald227@psu.edu)

The physical properties that govern the absorption of sound on Mars are very similar to those on Earth: classical losses associated with the transfer of acoustic energy into heat, and relaxation losses associated with the redistribution of internal energy of molecules. The difference in molecular composition between Earth and Mars as well as the lower atmospheric pressure on Mars results in larger values for the absorption coefficient on Mars. The direct simulation Monte Carlo (DSMC) method is the simulation tool used for modeling sound propagation in the Martian atmosphere. DSMC describes gas dynamics through direct physical modeling of particle motions and collisions. DSMC is based on the kinetic theory of gas dynamics, where representative particles are followed as they move and collide with other particles. The validity of DSMC for the entire range of Knudsen numbers (Kn), where Kn is defined as the mean free path divided by the wavelength, allows for the exploration of sound propagation in the Martian atmosphere for all values of Kn. Successful application of the DSMC method to acoustic waves in the Martian atmosphere, including the details of molecular relaxation in gas mixtures, will be shown. [Work supported by NASA.]

9:45

2aPA4. Correlation functions of temperature and velocity fluctuations in a turbulent atmosphere. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305 and Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003), Sergey N. Vecherin (New Mexico State Univ., Las Cruces, NM 88003), D. Keith Wilson (U.S. Army Engineer. Res. and Development Ctr., Hanover, NH 03755), and Sandra L. Collier (U.S. Army Res. Lab., Adelphi, MD 20783-1197)

von Kármán spectra of temperature and velocity fluctuations are widely used in theories of turbulence and wave propagation in random media, including sound propagation through a turbulent atmosphere. They are probably the simplest generalization of the Kolmogorov spectrum of homogeneous and isotropic turbulence which accounts for the outer scale of turbulence. However, the correlation functions of temperature and velocity fluctuations corresponding to the von Kármán spectra are proportional to the modified Bessel functions and are rather involved. In this paper, we propose to use much simpler correlation functions of temperature and longitudinal velocity fluctuations whose dependence on the distance R between observation points is given by $\exp(-R^{2/3}/L^{2/3})$, where L is the outer length scale. It is shown that the spectra of the proposed correlation functions coincide with the Kolmogorov spectrum in the inertial subrange and are bounded in the energy subrange. For some problems, the proposed correlation functions can simplify analysis of sound propagation through a turbulent atmosphere. Examples of the use of these correlation functions in acoustic tomography of the atmosphere and calculations of the coherence function of a sound field are presented. [Work supported by ARO, Grant DAAD19-03-1-0104.]

10:00

2aPA5. Upward propagation of nonlinear sound and magnetic sound in the atmosphere. Lev Ostrovsky (Zel Tech/NOAA ESRL, R/PSD4, 325 Broadway, Boulder, CO 80305)

As is known, even relatively weak ground-level perturbations (earthquakes, tsunami, explosions) may cause very significant displacements in upper atmospheric layers where the gas density is small. It results in shock formation and subsequent rapid dissipation of wave energy. This presentation deals with the upward propagation of a short, finite-amplitude wave in atmosphere with exponentially decreasing density, including areas where the magnetic pressure prevails, and the acoustic energy is transformed to that of fast and/or slow magnetic sound. The height of shock formation is determined, and the wave attenuation due to shocks is described. Propagation from a localized source is also discussed. Some estimates are given.

10:15–10:30 Break

10:30

2aPA6. Effect of longitudinal coherence on angle-of-arrival estimation in atmospheric turbulence. Sandra Collier (U.S. Army Res. Lab., AMSRD-ARL-CI-ES, 2800 Powder Mill Rd., Adelphi, MD 20783-1197, scollier@arl.army.mil), Vladimir Ostashev (NOAA/Earth System Res. Labs., Boulder, CO 80305), and Keith Wilson (U.S. Army Engineer Res. and Development Ctr., Hanover, NH 03755)

A statistical model that accounts for the effects of atmospheric turbulence on a received acoustic signal has been previously developed [Collier and Wilson, J. Acoust. Soc. Am. **113**, 2704–2718 (2003)]. It was used in calculations of the performance bounds, and of a maximum likelihood estimator (MLE), for the angle of arrival (AOA) of an acoustic signal received at an array operating in atmospheric turbulence with fluctuations described by a von Karman spectrum. This statistical model is based on the moments of a sound field as derived in the theory of wave propagation in a random medium (WPRM) [Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN Spon 1997)]. In this model, an approximation to the second moment was used based on the transverse coherence from WPRM. Recently, the longitudinal coherence has been derived for WPRM [Ostashev *et. al*, 151st, Meeting of the Acoustical Society of America (2006)]. We examine limiting cases to determine the effects of the longitudinal coherence on the MLE and Cramer-Rao lower bound of the AOA.

10:45

2aPA7. The coherence function of a sound field propagating in a turbulent atmosphere. Vladimir E. Ostashev (NOAA/Earth System Res. Lab., Boulder, CO 80305 and Dept. of Phys., New Mexico State Univ., Las Cruces, NM 88003), Sandra L. Collier (U.S. Army Res. Lab., Adelphi, MD 20783), and D. Keith Wilson (U.S. Army Engineer. Res. and Development Ctr., Hanover, NH 03755)

The coherence function of a sound field propagating through an atmosphere with temperature and wind velocity fluctuations is important for many practical applications involving microphone arrays. In the literature, the transverse coherence function has been studied in detail, when acoustic sensors are located in a plane perpendicular to the sound propagation path. In this paper, for the case of line-of-sight sound propagation, a closed-form expression for the coherence function of a plane wave is derived for arbitrary location of acoustic sensors. Using this expression, for some limiting cases simple formulas for the coherence function are obtained which are valid for arbitrary spectra of turbulence. Furthermore, the coherence function is calculated and analyzed in detail for the Gaussian and von Kármán spectra of temperature and wind velocity fluctuations. In particular, the longitudinal coherence radius, when acoustic sensors are located along the sound propagation path, is calculated. It is also shown that contours of the coherence radius are elongated perpendicular to this path for relatively small propagation distances, and along it for large distances.

11:00

2aPA8. Measurements of acoustic and seismic pulses from outdoor explosions. Donald G. Albert (ERDC-CRREL, 72 Lyme Rd., Hanover, NH 03755), Keith Attenborough, and Patrice Boulanger (The Univ. of Hull, Hull HU6 7RX, UK)

Measurements of the ground vibrations produced by airborne detonations of C4 were conducted at locations with a variety of ground types, including concrete, soil, forest, tropical vegetation, and snow cover. The measurements show that, although two separate seismic (ground vibrational) arrivals can be detected in all cases, the early seismic arrival from an underground path is always much smaller than the vibration induced by the air blast arrival. The acoustic-to-seismic coupling ratio for the atmospheric wave is a constant with respect to distance and peak pressure at a given location, but varies from site to site, usually between 1 and 14 μm/s/Pa. A conservative empirical equation to predict ground vibration from explosions is derived. This equation predicts that the commonly used

vibrational damage criteria of 12 and 25 mm/s will be exceeded when the peak positive pressure exceeds 480 Pa (147.6 dB) or 1 kPa (154.0 dB), respectively. Either of these levels is much higher than the Army overpressure damage criterion of 159 Pa (138 dB). Thus in most situations damage from blast overpressure will occur long before damaging levels of ground vibration are reached. [This research supported by the U.S. Army Corps of Engineers and SERDP Seed Project SI-1410.]

11:15

2aPA9. Comparison of high-frequency measurements of blast wave propagation to computational model predictions. Alexandra Loubeau and Victor W. Sparrow (Grad. Prog. Acoust., Penn State, 202 Appl. Sci. Bldg., University Park, PA 16802, aloubeau@psu.edu)

In November 2005, the latest in a series of blast wave propagation experiments was conducted with the U.S. Army. Blast waveforms measured with a wideband microphone [W. M. Wright, ONR Report NR-384-321 (1971)] had shorter rise times than those captured with conventional $\frac{1}{8}$ -in. microphones. This indicates that the $\frac{1}{8}$-in. microphone setup did not have the high-frequency bandwidth needed to capture such short rise times. The trend observed is that nonlinear effects steepen the waveform, thereby decreasing the shock rise time, up to a range of 50 m. At 100 m, the rise times had increased slightly, but they were still shorter than they were at 25 m. The experimental results are compared to predictions from a computational model. The model, based on the Pestorius/Anderson hybrid time-frequency domain algorithm, accounts for nonlinearity, geometrical spreading, thermoviscous absorption, and absorption and dispersion due to molecular relaxation. It is found that the model predicts shorter rise times than what was measured with the wideband microphone, but the rise-time variation with range is similar. Atmospheric turbulence present in the experiments, but not accounted for in the model, is a likely cause for the difference in rise times. [Work supported by U.S. Army ERDC CERL.]

11:30

2aPA10. The use of Lorentz transformations for simplifying aeroacoustic boundary value problems. M. H. Dunn (Natl. Inst. of Aerosp., 100 Exploration Way, Hampton, VA 23666, mhd314@aol.com)

The prediction of community noise from aircraft flyovers is computationally complex when arbitrary aircraft configurations are subjected to realistic flows. Mathematically, the problem is described by a set of scattering surfaces (wings, nacelles, and fuselage) moving uniformly with known noise sources in the presence of an inviscid, irrotational, background flow. A four-dimensional (space-time) boundary value problem (BVP) for noise propagation and scattering is defined by the mass and momentum conservation equations along with appropriate hard- or soft-wall boundary conditions on the scattering surfaces and the Sommerfeld radiation condition. This system is augmented by the appropriate Kutta conditions for geometry with edges to produce a uniquely solvable BVP. The governing differential equations are complicated by factors strongly dependent upon the flow field. However, Lorentz transformations can be used to define new independent and dependent variables such that the differential equations depend on the local sound speed plus terms that involve flow gradients. Neglecting flow gradient terms produces the classical wave equation BVP, which can be solved using contemporary numerical methods. The original acoustic variables are easily recovered via the inverse Lorentz transformation. In addition, the formulation can be used to analyze physical conditions under which the uniform flow approximation is valid.

TUESDAY MORNING, 6 JUNE 2006 ROOM 551AB, 8:15 TO 9:45 A.M.

Session 2aPPa

Psychological and Physiological Acoustics: Loudness, Outer Hair Cells, and Otoacoustic Emissions

Walt Jesteadt, Chair
Boys Town National Research Hospital, 555 N. 30th St., Omaha, NE 68131

Contributed Papers

8:15

2aPPa1. A comparison of intensity discrimination and increment detection as measures of intensity resolution. Walt Jesteadt, Jessica Messersmith, Lori Leibold, Samar Khaddam, and Hongyang Tan (Boys Town Natl. Res. Hospital, 555 N. 30th St., Omaha, NE 68131)

Intensity resolution can be measured by determining the threshold for discrimination of two stimuli differing only in intensity or by determining the threshold for detection of an increment in a longer duration stimulus. Better resolution has been reported for the latter task, but the results are inconsistent across conditions and subjects. In the present study, data were obtained for the two tasks in a range of conditions for 12 adult subjects with normal hearing. Thresholds for intensity discrimination and increment detection were measured for a 4000-Hz tone presented at 20, 40, 60, or 80 dB SPL, using a two-track adaptive procedure that estimated levels required for 71% and 89% correct. The increment was 20, 40, 80, or 160 ms in duration. The pedestal was either the same duration as the increment or was 200 ms longer, centered on the increment. Thresholds expressed in units of delta L were more uniform across subjects and conditions for increment detection than for intensity discrimination. Thresholds were lower on average for increment detection than for intensity discrimination at pedestal levels above 20 dB, but only at the shorter increment durations, and thresholds were not consistently lower for all subjects. [Work supported by R01 DC006648.]

8:30

2aPPa2. Primary speech frequencies: Adaptation inside versus outside—Inhibition or excitation? Ernest M. Weiler, David E. Sandman (Dept. Of Commun. Sci., Univ. Cincinnati, Cincinnati OH 45267), and Joel S. Warm (Univ. Cincinnati, Cincinnati, OH 45267)

Previously, loudness adaptation from the ipsilateral comparison paradigm (ICP) and the simple adaptation (SA) procedure were compared at primary speech- related frequencies (250 to 4000 Hz), with only the ICP showing significant loudness decline/adaptation [Tannen *et al.*, J. Gen Psychol. **128**, 385–399 (2001)]. Eighty naive undergraduates participated in any one of three subsequent studies of ICP versus SA adaptation. In agreement with Tannen *et al.*, only the ICP showed significant loudness decline/adaptation from 250 to 6000 Hz. However, at 8000 Hz both the ICP and SA show significant adaptation, and correlate significantly. However, despite the lack of correlation between individual scores under 8000 Hz, a further analysis of *group mean* values from 250 to 8000 Hz showed a significant correlation ($r=0.97$, $p<0.001$). Perhaps the SA procedure shows a lack of adaptation at speech frequencies because the method is insensitive. Alternatively, it may be considered that a ''stable platform'' [i.e., no adaptation for *constant* stimuli] favors the perception of the constantly varying speech signals. Support for the ''stable platform'' versus the ''insensitivity'' hypothesis will be discussed.

8:45

2aPPa3. Induced loudness reduction as a function of frequency separation. Jeremy P. Marozeau, Michael Epstein, and Mary Florentine (Commun. Res. Lab., Dept. of Speech-Lang. Pathol. and Audiol., Northeastern Univ., Boston, MA 02155, marozeau@neu.edu)

When a loud tone (inducer tone) precedes a softer tone (test tone), the loud tone can reduce the loudness of the softer tone. This phenomenon is called induced loudness reduction (ILR). The amount of ILR is greatest when the inducer tone and the test tone are presented at the same frequency. As the difference in frequency between the two increases, the amount of ILR decreases [Marks L. E., J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396 (1994)]. However, the exact course of this decrease with increasing frequency separation has not been thoroughly tested. In this experiment, the amount of ILR produced by a 2.5-kHz 80-dB-SPL inducer was measured from 800 Hz to 6 kHz using a frequency-sweeping test tone. Nine normal-hearing listeners matched the loudness of the test tone to a 70-dB-SPL 500-Hz tone using a tracking procedure. Average results show a maximum amount of ILR between 2.5 and 3 kHz, with a loudness reduction equivalent to 12 dB. This amount decreases with frequency separation, but is still significant when the frequency of the test tone is between 1.5 and 4.5 kHz, corresponding to more than nine critical bands. [Work supported by NIH/NIDCD Grant R01DC02241.]

9:00

2aPPa4. A multicompartmental cochlear model with piezoelectric outer hair cells. Shan Lu, David C. Mountain, and Allyn E. Hubbard (College of Eng., Boston Univ., 8 Saint Mary's St., Boston, MA 01720, slu@bu.edu)

A hydromechanical, multicompartment model of the cochlea, which employed a phenomenological outer hair cell (OHC) cell-body electromotility, was able to mimic the physiologically measured response of the basilar membrane (BM) [A. E. Hubbard *et al.*, "Time-domain responses from a nonlinear sandwich model of the cochlea," in *Biophysics of the Cochlea: From Molecule to Model*, edited by A. W. Gummer (World Scientific, Singapore)]. An improved model that included OHC electroanatomic parameters and scalar electrical parameters was able to mimic cochlear microphonic (CM) data at low frequencies, but not BM responses at high-frequencies, because the OHC membrane capacitance severely reduces OHC transmembrane potential. To improve the high-frequency performance, we used a piezoelectrical model of the OHC. Consequently, the mechanical loading of the OHC translates bidirectionally with the electrical impedance of the OHC. The new model compares well with physiological data from Gerbil at 40 dB SPL [T. Ren and A. Nuttall, "Basilar membrane vibration in the basal turn of the sensitive gerbil cochlea," Hearing Res. **151**, 48–60 (2001)]. [This work was supported by NIH.]

9:15

2aPPa5. Application of force to the cochlear wall: Effect on auditory thresholds, outer hair cell transduction, and distortion-product otoacoustic emissions. Greg A. O'Beirne[a] and Robert B. Patuzzi (The Auditory Lab., Physiol., School of Biomed. and Chemical Sci., Univ. of Western Australia, Crawley WA 6009, Australia)

Described are the changes in cochlear sensitivity and mechanoelectrical transduction during a novel cochlear perturbation: the application of force to the cochlear wall. While numerous methods exist to create *transient* shifts in the operating point of the outer hair cell (OHC) transducer, including low-frequency acoustic bias [G. Frank and M. Kössl, Hear. Res. **113**, 57–68 (1997)] and hydrostatic bias [A. N. Salt and J. E. DeMott, Hear. Res. **123**, 137–147 (1998)], attempts to create *prolonged* operating point shifts are largely thwarted by the numerous sources of ac coupling in the auditory system which prevent transmission of dc stimuli to the hair cells. The application of force sufficient to deform the otic capsule produced a consistent drop in neural thresholds and a sustained bias of the OHC operating point that did not rapidly adapt back to normal. Near-simultaneous measurements of compound action potential thresholds, distortion-product otoacoustic emissions, the OHC transfer curves derived from low-frequency cochlear microphonic waveforms, and the endocochlear potential were performed. The data provide ample evidence of the resistance of the cochlea to dc mechanical stimuli, particularly those which do not cause a large pressure differential across the basilar membrane. [The authors gratefully acknowledge the surgical assistance of Dr. Peter Sellick.] [a]Currently at Department of Communication Disorders, University of Canterbury, Christchurch, New Zealand.

9:30

2aPPa6. Influence of the sweeping direction of the primaries f_1 and f_2 in the assessment of distortion product otoacoustic emissions (DPOAEs) fine structure. Miguel Angel Aranda de Toro, Rodrigo Ordoñez, and Dorte Hammershøi (Dept. of Acoust., Aalborg Univ., Fredrik Bajers Vej 7 B5, DK-9220, Aalborg, Denmark, maat@acoustics.aau.dk)

It was studied if the sweeping direction of the primaries f_1 and f_2 influences fine structure measurements of the $2f_1-f_2$ distortion product otoacoustic emission (DPOAE). Two different methodologies named DPOAE5$_{ASC}$ (ascending sweep) and DPOAE5$_{DES}$ (descending sweep) were implemented with the commercial system ILO96. Both methods perform fine structure measurements in the same frequency range (1.4–6 kHz) with $L1/L2=65/45$ dB, $f_2/f_1=1.22$ and 0.7 s averaging time per primary presented. DPOAE fine structures were measured in the right ear of 14 normal-hearing subjects with both methods in a balanced experiment and without refitting the sound probe. Results showed that the two methods are highly repeatable and able to detect fine structures even in subjects with low S/N. However, when DPOAE5$_{ASC}$ measurements are compared with DPOAE5$_{DES}$, fine structures appear shifted and the level contour is altered. A minor level difference was inadvertently induced due to differences in the level calibration procedure of the two methods employed. The significance of this is currently being studied. [Work supported by the Danish Research Council for Technology and Production.]

Session 2aPPb

Psychological and Physiological Acoustics, ASA Committee on Standards, and Noise: Individual Susceptibility to Noise-Induced Hearing Loss

Sharon G. Kujawa, Cochair
Massachusetts Eye and Ear Infirmary, Audiology Dept., 243 Charles St., Boston, MA 02114

Lynne Marshall, Cochair
Naval Submarine Medical Research Lab., Subase New London, Box 900, Groton, CT 06349

Invited Papers

10:15

2aPPb1. Individual susceptibility to noise-induced hearing loss: A review. Donald Henderson (Ctr. for Hearing and Deafness, State Univ. of New York at Buffalo, 137 Cary Hall, Buffalo, NY 14214)

After 40+ years of study, we can conclude that there is a great range of individual differences in susceptibility to noise-induced hearing loss (NIHL). The presentation will review and evaluate the major sources of variances: *acoustic*, where all exposures are evaluated with dBA and some energy-related metric; *exogenous*, where the effects of noise can be mitigated by external factors such as temperature, coexposure to chemicals, smoking, etc.; *endogenous*, where the effects of noise are partially related to an individual's sex, skin color, acoustic reflex, distortion product suppression, degree of fatigue, and genetic factors.

10:45

2aPPb2. Searching the whole genome to identify genes contributing to noise resistance in inbred mouse strains. Bruce L. Tempel, Valerie A. Street (Dept. of Otolaryngol.-HNS and the V.M. Bloedel Hearing Res. Ctr., Univ. of Washington, Seattle, WA 98195-7923, bltempel@u.washington.edu), and Sharon G. Kujawa (Harvard Med. School, Boston MA)

Toward understanding the biological basis of noise resistance and eventual formulation of therapies to prevent noise-induced hearing loss (NIHL), the identification of genes contributing to noise resistance is a critic step. To simplify our whole genome search, we chose to study isogenic, inbred strains of mice. When exposed to noise (octave band 8–16 kHz, 103 dB SPL, 2 h) the good-hearing CBA/CaJ strain showed ~50-dB thresho shift measured 2 weeks postexposure; in contrast 129S6 mice showed ~10-dB shift. Through directed breeding schemes we showed that the noise resistance trait in 129S6 is recessive to CBA. Screening N2 backcross animal for noise resistance suggested a multigenic pattern of inheritance with 4 to 6 genes contributing. Quantitative trait locus (QTL) mapping of 234 N2 animals reveals 4 different chromosomal regions that contribute significantly to the resistance. Congenic strains carrying only one or two of these QTL regions are currently being bred, with the expectation that these strains will show partial noise resistance. Further studies including ancestral haplotype analysis and differential gene expression array analysis will allow us to narrow the contributing genomic regions and identify specific genes contributing to noise resistance. [Work supported by NIH.]

11:05

2aPPb3. Can otoacoustic emissions indicate susceptibility to noise-induced hearing loss in individual ears? Lynne Marshall, Judi A. Lapsley Miller (Naval Submarine Medical Res. Lab., Groton, CT 06349-5900, marshall@nsmrl.navy.mil), and Laurie M. Heller (Brown Univ., RI 02912)

Otoacoustic emissions (OAEs) can potentially be used to identify normal-hearing individuals who are susceptible to imminent noise-induced hearing loss. Until now, there has been no way to monitor an individual's susceptibility dynamically as it varies due to physiological and environmental factors. Although it is known that groups of normal-hearing noise-exposed people have lower than average levels of OAEs, such data have never been used to predict future hearing loss in the same individuals. Here, we discuss two studies where the OAEs and hearing thresholds of individuals were measured before and after hazardous noise exposure. Individuals with normal hearing and low-level or absent OAEs at baseline were at increased risk for acquiring noise-induced hearing loss after the noise exposure. This supports the theory that OAEs can reflect incipient inner-ear damage undetected by standard behavioral hearing tests. Furthermore, OAE efferent strength measures may be predictors of NIHL in humans. The animal data are promising, but developing an efficient OAE test that reliably distinguishes a strong efferent reflex from a weaker one in humans is challenging. [Work supported by ONR & USAMRMC].

11:25

2aPPb4. Noise-induced and age-related hearing loss interactions. Sharon G. Kujawa (Dept. of Otology and Laryngology, Harvard Med. School, and Dept. of Audiol., Massachusetts Eye and Ear Infirmary, 243 Charles St., Boston, MA 02114) and M. Charles Liberman (Massachusetts Eye and Ear Infirmary, Boston, MA 02114)

Noise-induced hearing loss (NIHL) and age-related hearing loss (AHL, or presbycusis) are widespread health problems that will continue to increase as our society ages. NIHL and AHL often coexist in the same ear; however, the conditions under which they interact and the mechanisms by which they do so remain poorly understood. Inspired by epidemiological studies suggesting that noise-exposed ears age *differently* from nonexposed ears [Gates *et al.*, ''Logitudinal threshhold changes in people with audiometric notches,'' Hear. Res. **141**, 220–228 (2000)], we studied interactions between NIHL and AHL in mouse; an animal with a short life span, with intrastrain genetic homogeneity to minimize variability and with interstrain differences in vulnerability which can be exploited to probe mechanisms. Using such models, we have uncovered evidence that early noise exposure can have an ongoing influence on the nature and progression of an age-related hearing loss. The nature of this age-related hearing loss exacerbation is special; it can occur even in ears without permanent threshold shifts from the noise and leads to massive loss of spiral ganglion neurons despite intact hair cell populations. Findings have practical importance for investigations of NIHL in animal models, and may have implications for clinical practices allocating noise-induced and age-related components of hearing loss as strictly additive. [Work supported by NIH.]

Contributed Paper

11:45

2aPPb5. Hearing levels in US adults aged 20–69 Years—National Health and Nutrition Examination Survey (NHANES) 1999–2004. William J. Murphy, Christa L. Themann, and Mark R. Stephenson (Hearing Loss Prevention Team, Natl. Inst. for Occupational Safety and Health, 4676 Columbia Pkwy., MS C-27, Cincinnati, OH 45226-1998)

The National Health and Nutrition Examination Survey (NHANES) is a nationally representative, population-based survey designed to assess the health and nutritional status of the civilian, noninstitutionalized U.S. population. Data were collected through a personal interview regarding health history and through physical examination. Earlier NHANES surveys were conducted on a periodic basis; however, in 1999, NHANES began collecting data on a continuing, annual basis. During NHANES I, which ran from 1971 to 1975, audiometric testing was conducted on adults aged 25–74 years. No subsequent testing of adults was conducted in the NHANES program until 1999, when NHANES began audiometric testing of adults aged 20–69 years. In 2004, the sampling from this age group was completed. This report examines the hearing levels for adults in the United States and compares them with the hearing data from NHANES I. Hearing levels are grouped by age and by ethnicity and gender.

TUESDAY MORNING, 6 JUNE 2006 BALLROOM D, 8:15 A.M. TO 12:00 NOON

Session 2aSC

Speech Communication: Enhancement and Multiplicity of Cues in Speech (Lecture/Poster Session)

Yi Xu, Cochair
Wolfson House, 4 Stephenson Way, London, NW1 2HE, UK

Carlos Gussenhoven, Cochair
Radboud Univ. Nijmegen, Ctr. for Language Studies, Postbus 9103, 6500 HT Nijmegen, The Netherlands

Chair's Introduction—8:15

Invited Papers

8:20

2aSC1. Defining and enhancing attributes for features. Kenneth N. Stevens (Res. Lab of Elec., Dept. EECS, & Div. Health Sci. & Tech, MIT, Cambridge, MA 02139, stevens@speech.mit.edu)

We take the view that words are represented in memory as sequences of bundles of binary distinctive features. It is further assumed that each feature is defined by a quantal relation between an articulatory attribute and a corresponding acoustic property that provides acoustic cues to the feature. In the speaking process, other acoustic properties are introduced in addition to the defining acoustic attributes, and these contribute additional cues that enhance the perceptual saliency of the feature. For example, the defining attributes for the voicing feature is the presence or absence of glottal vibration during an obstruent consonant, but there are several possible enhancing properties, such as the presence of aspiration noise and a modified fundamental frequency in the following vowel. In running speech, the saliency of a feature may be in jeopardy, especially when there is interfering noise or when gestural overlap weakens some of the cues, particularly those derived from the defining gesture for the feature. An example is the casual production of batman where the closure for /t/ may be eliminated by gestural overlap, but two enhancing gestures ($F2$ transition for the alveolar and the glottalization for /t/) are preserved. [Work supported in part by NIH Grant DC00075.]

8:40

2aSC2. Is some information perceptually primary? John Kingston (Ling. Dept., Univ. of Massachusetts, Amherst, MA 01003-9274, jkingston@linguist.umass.edu) and Randy Diehl (Univ. of Texas, Austin, TX 78712)

The existence of multiple correlates for phonological contrasts invites the hypothesis that one is perceptually primary and that the others enhance the primary correlate. (Stevens Acoustics Phonetics, MIT Press, 1998) particularly emphasizes the perceptual primacy of acoustic landmarks. Our experiments manipulate the quality or variability of the acoustic information in a landmark versus a neighboring nonlandmark to determine whether listeners rely on the landmark or the most reliable information source. Stimuli are synthetic stop-vowel sequences where the stop's place of articulation is conveyed by its burst spectrum—landmark—and the formant transitions to the following vowel—nonlandmark. Parameter values in the two intervals fall on opposite sides of, but close to, a category boundary and convey competing place information. Information is degraded in one interval by broadening the bandwidths of the spectral peaks or formants, or it is made more variable by varying the peaks frequencies or the following vowel quality. If landmark information is perceptually primary, then degrading or varying that information should shift place judgments more toward the category conveyed by the nonlandmark interval than vice versa. However, if the quality or consistency of the information matters, then manipulating either interval should produce comparable shifts in place judgments. [Work supported by NIH.]

9:00

2aSC3. Weighted multiple cues and phonological features. Patrice S. Beddor (Dept. of Linguist., Univ. of Michigan, 1190 Undergraduate Sci. Bldg., Ann Arbor, MI 48109, beddor@umich.edu)

Do speaker-listeners determine a primary phonological feature from among multiple cues for a phonological contrast? Linguistic analyses traditionally postulate a primary feature, yet experimental evidence indicates that speaker-listeners assign relative weights to multiple cues, in context-dependent ways, and that these weighted cues are part of the phonological grammar. In this paper, evidence for a noncategorical, cue-weighted perspective is drawn in part from perception of nasal consonants (Ns). English Ns are shorter before voiceless than before voiced obstruents (in VNC contexts), and anticipatory vowel nasalization is temporally more extensive before the shorter, prevoiceless Ns. Unsurprisingly, listeners' judgments of stimuli in which N duration and temporal extent of vowel nasalization were independently manipulated in CV(N)C sequences (coda C = [t] or [d]) broadly parallel the production data. But individual listeners differed in their cue weightings. Some listeners heard vowel nasalization and the nasal murmur as perceptually equivalent information about the presence of N, other listeners were highly sensitive to vowel nasalization, and still others' cue sensitivity depended on coda voicing. The patterns of responses are difficult to interpret in terms of a primary feature, and the phonological implications of these patterns will be discussed. [Work supported by NSF.]

9:20

2aSC4. Enhancing the durational enhancement of a tone contrast. Carlos Gussenhoven (Radboud Univ. Nijmegen, Letteren, Taalwetenschap, Postbus 9103, 6500 HT Nijmegen, The Netherlands and Queen Mary, Univ. of London, Mile End Rd., London E14NS)

In central and low Franconian dialects of German and Dutch, a lexical tone contrast (accent 1 versus accent 2) exists that is enhanced by a durational difference. While pitch differences vary according to intonation and sentence context, syllables are generally shorter when they have accent 1. In addition, vowels in those syllables are more open, in comparison with vowels in syllables with accent 2. Perception experiments with natural vowels whose acoustic durations had been manipulated revealed that high vowels are perceived as longer by Dutch listeners than midvowels when acoustic durations are equal. It is argued that vowel raising of longer vowels, or vowel lowering of shorter vowels, serves as an enhancement of duration differences. The tendency is widely attested to in the vowel systems of Germanic languages, where short (''lax'') vowels are opener than long (''tense'') vowels, and vowel height is used to enhance a *phonological* duration contrast. In the dialects investigated, vowel height is manipulated to enhance a durational difference, which in turn enhances a tonal contrast. The explanation of the durational illusion of vowel height is the listener's compensation of the universally shorter duration of higher vowels.

9:40

2aSC5. Enhancement between voice onset time (VOT) and *F*0: Evidence from Kera voicing perception. Mary D. Pearce (Univ. College London & SIL. Phonet. and Linguist., UCL, Gower St., London, WC1E 6BT, UK, mary_pearce@sil.org)

Previous research on production data from Kera (a Chadic tone language) has shown that tone plays a major contrastive role, whereas VOT has a lesser role which could indicate an ''enhancing'' relationship. This paper investigates the roles of VOT and *F*0 in Kera voicing perception, and compares this to the nontonal language English. Fifteen Kera and 15 English subjects were given voicing judgment tasks over 200 tokens that had been manipulated in PRAAT to cover a range of VOT and *F*0 values. The results show a marked difference in English and Kera perception reflecting the phonological differences between the two languages. For Kera, *F*0 is of major importance, while the VOT plays a lesser role in the judgment. The present findings are consistent with a view that whereas in English, the VOT is contrastive and *F*0 arguably acts as an enhancing cue, in Kera, the tone is the main contrastive cue to voicing perception while the VOT values enhance the tone cues. A bonus finding from the perception tests is that the relationship between VOT and *F*0 in Kera voicing perception tasks depends on their exposure to French during childhood.

10:00–10:15 Break

10:15

2aSC6. Commentary on invited papers for "Enhancement and multiplicity of cues in speech." D. H. Whalen (Haskins Labs., 300 George St., New Haven, CT 06511)

Enhancement has been shown to occur in a variety of contexts in the papers presented here. This commentary will examine the similarities of these phenomena and look for their proper level of description within linguistic theory.

10:35

2aSC7. Possible mechanisms of enhancement and multiplicity of cues. Yi Xu (Dept. of Phonet. and Linguist., Univ. College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, UK, yi@phon.ucl.ac.uk)

It has been proposed that phonological contrasts are phonetically realized not only by their primary articulations, but often also by enhancing gestures [Kingston & Diehl "Phonetic knowledge," Language, **70**, 419–454 (1994); Stevens & Keyser, "Primary features and their enhancement in consonants," Language, **65**, 81–106 (1989)]. Meanwhile, enhancement appears to parallel the well-known phenomenon of multiplicity of cues in speech perception [Lisker "Voicing in English: Acatologue of acoustic features signaling /b/ versus /p/ in trochees." Lang. Speech, **29**, 3–11 (1986)]. An interesting question is therefore whether the two are one and the same. By a narrow definition, only articulatory maneuvers not obligated by the articulatory gestures of the phonetic units in question play an enhancing role. Thus not all cues that benefit perception are generated for the sake of enhancement. On the other hand, there may indeed be cases where genuine enhancement occurs, for which a further question is why is it needed in the first place? One possibility is that enhancement is not only for resisting noise in the speech environment, but also for handling the competition among communicative functions that are transmitted in parallel. That is, multiple communicative functions often share the same acoustic/articulatory dimension and such functional crowding may create the need to exploit additional means of encoding. If so, both enhancement and multiplicity of cues could be part of a live process that maintains the effectiveness of speech communication.

10:55–11:00 Break

Contributed Papers

All posters will be on display and all contributors will be at their posters from 11:00 a.m. to 12:00 noon.

2aSC8. Enhancement of the vowel quantity contrast in Japanese and German by F0 and vowel quality. Heike Lehnert-LeHouillier (Dept. of Linguist., Univ. at Buffalo, 602 Baldy Hall, Buffalo, NY 14260, hlehnert@buffalo.edu) and D.H. Whalen (Haskins Labs., New Haven, CT 06511)

Previous studies on the perception of vowel quantity in Japanese and German have shown an influence of F0 for Japanese and of vowel quality for German. The current study investigates F0 and vowel quality enhancement of vowel length perception in both languages. Native speakers of Japanese and German were presented with three series of stimuli: (*i*) vowel continua for the vowels /a/, /e/, and /i/ which differed in duration only; (*ii*) vowel continua for the same vowels which differed in duration and the steepness of the F0 fall over the vowel; and (*iii*) vowel continua for the same vowels containing both durational and quality cues. While vowel quantity in the first two series was perceived categorically by Japanese listeners, vowel quantity in the stimuli testing the influence of vowel quality was perceived as gradual rather than categorical by the same listeners. German listeners perceived all series as categorical. Thus F0 cues generally enhance the distinction of the vowel length contrast whereas the presence of quality cues facilitates the distinction in German but impedes the perception of the same distinction in Japanese. This suggests that enhancement of vowel quantity is governed by perceptual cues that are in part language specific.

2aSC9. The role of acoustic familiarity in the perception of L2 contrasts. Gautam K. Vallabha (Ctr. for the Neural Basis of Cognition, Carnegie Mellon Univ., 4400 Fifth Ave., Mellon Inst. 110, Pittsburgh, PA 15213)

Difficulties in L2 perception are often attributed to assimilation of the L2 contrast by an L1 category or to interference by L1-relevant acoustic cues [Iverson *et al.* "Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/p/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278 (2005)]. These approaches assume that cue weights change slowly over developmental time and are relatively constant for short time scales. An alternative view is that cue extraction and weighting are dependent on the overall familiarity of the current sound, with more familiar sounds facilitating more robust cue extraction and better discrimination [e.g., McFadden and Callaway, JEP:HPP, **25**, 543 (1999)]. This hypothesis was tested with the perception of English sounds by Japanese adults. The Japanese listeners are insensitive to a cue (F3 onset) when it is part of (putatively unfamiliar) English /ra/ and /la/ sounds. The effect of familiarity was evaluated by fixing the range of F3 onsets and manipulating the rest of each sound in two ways: (1) by changing the transition durations to be more suitable for a moraic rhythm and (2) by changing the F1, F2, and amplitude trajectory to be more like the Japanese /r/. Discrimination improved markedly with the latter manipulation, suggesting that cue sensitivity is conditioned upon overall acoustic familiarity. [Work supported by NIH MH64445.]

2aSC10. Learning to ignore a perceptual cue: Nonnative listeners outperform native listeners. Mirjam Broersma (Max Planck Inst. for Psycholinguisti., P.O. Box 310, 6500 AH Nijmegen, The Netherlands, mirjam.broersma@mpi.nl)

Native and nonnative listeners categorized /v/ and /f/ at the end of English nonwords. For each participant, the duration of the previous vowel was kept constant, so that it was not informative and sometimes mismatched other information in the signal. Vowel duration was varied between participants. Previously presented results [M. Broersma, J. Acoust. Soc. Am. **117**, 3809–3901 (2005)] showed that native English listeners relied strongly on the misleading vowel duration cue. For Dutch listeners, no effect of vowel duration was found. Due to the redundancy of information in the signal, Dutch listeners categorized the contrast more categorically than English listeners. New analyses investigated whether Dutch listeners did not attempt to use vowel duration at all, or whether they learned to ignore the misleading cue more easily than the English listeners did. The results showed that Dutch listeners did use vowel duration initially, but stopped using this cue after very few trials. By the end of the practice part (33 trials) the effect of vowel duration had fully disap-

peared. The English listeners used vowel duration as a voicing cue throughout the experiment. This suggests that it may be easier to learn to ignore uninformative perceptual cues in a nonnative language than in one's native language.

2aSC11. Effects of task complexity on perceptual organization of speech cues by older listeners. Jessica A. Rossi-Katz and Kathryn H. Arehart (Dept. Speech Lang. and Hearing Sci., Univ. of Colorado, 409 UCB, Boulder, CO 80309)

Older adults with and without clinically significant hearing loss often report difficulty communicating in multitalker environments. Past work in our laboratory has shown that older listeners with hearing loss benefit less from differences in talkers' vocal characteristics (e.g., fundamental-frequency differences) relative to younger listeners with comparable hearing losses. The present study examined the extent to which age-related changes in the perceptual organization of speech cues are due to task demand. Two experiments used the Coordinate Response Measure to examine how the benefit listeners derive from talker characteristics interacts with the complexity of the response task as a function of both age and hearing status. Stimuli were amplified to assure audibility for listeners with hearing loss. Task complexity depended on whether the distracting signal was linguistically meaningful as well as the level of auditory processing required (discrimination versus identification). Older listeners generally benefited more from talker characteristics when task complexity was reduced (e.g., nonmeaningful distracter). The results are interpreted in the context of an information-processing model in which the increased difficulty of older listeners in multitalker environments is a consequence of reallocating limited resources towards the initial stages of processing at the expense of higher-level stages. [Work supported by Centers for Disease Control.]

TUESDAY MORNING, 6 JUNE 2006 ROOM 552B, 8:00 A.M. TO 12:00 NOON

Session 2aSP

Signal Processing in Acoustics: General Topics in Signal Processing in Acoustics

David J. Moretti, Chair
Naval Undersea Warfare Ctr., Code 71, Newport, RI 02841

Contributed Papers

8:00

2aSP1. A biomimetic robotic system for localizing gunfire. Socrates Deligeorges, Aleksandrs Zosuls, David Mountain, and Allyn Hubbard (Dept. Biomed. Eng, Boston Univ., 44 Cummington St., Boston, MA 02215, sgd@bu.edu)

Using a biomimetic approach, a new method of acoustic signal processing was created which has tremendous advantages in complex acoustic environments. The biomimetic approach was used as the basis for a system to localize and identify sound sources in noisy and reverberant conditions. The algorithms are based on mammalian hearing and mimic the acoustic processing of the auditory periphery and midbrain. The system uses spectro-temporal cues exploited by the auditory system including such features as interaural time difference (ITD), interaural level difference (ILD), spectral profile, and periodicity content. The initial system of algorithms were designed and tested using the EARLAB [earlab.bu.edu] software modeling environment. The system of algorithms was then adapted to a mixed-signal real-time hardware solution and mounted on an iRobot PackBot robotic platform to perform simple behavioral tasks. The integrated system can detect and localize gunfire in a complex, reverberant acoustic environment and orient a camera towards the shooter. Accuracy in field tests with live fire was 1.5 deg in azimuth. The performance of the prototype platform demonstrates the potential of the biomimetic approach and its applications to practical problems for commercial, civilian, and military acoustic processing. [Work funded by ARL:DAAD19-00-2-0004.]

8:15

2aSP2. The multiple coherence function for sound field prediction. Torsten Leth Elmkjær (Terma AS, Hovmarken 4, DK-8520 Lystrup, Denmark, tln@terma.com)

From the *uniqueness theorem* it is well known that the sound field in a region exterior to the source region can be determined from knowledge of the pressure or the normal fluid velocity *everywhere* on the enclosing surface. In practice the acoustical field quantities are obtained from space-time discretization. The space-time samples on the enclosing surface and arbitrary field points in the interior to the volume will be considered as input and output processes to a multiple-input–multiple-output (MIMO) system, respectively. The *multiple coherence function* $\gamma^2_{y:x}$ provides a measure of linearity between a set of inputs and an output *independent* of the correlation among the inputs. Various system imperfections, however, will lead to a nonunity $\gamma^2_{y:x}$ In particular, spatial sampling, similar to time-domain sampling, requires limited bandwidth to avoid aliasing. Therefore, a sufficient number of sensors must be allocated to ensure $\gamma^2_{y:x} \approx 1$. In this paper a spherical enclosure is investigated. The objective is to express $\gamma^2_{y:x}$ as a function of internal field point position for different space-time sampling strategies. Simulation results for single plane wave, diffuse fields, and multiple-source fields is presented. Measurement results obtained in a reverberant room are discussed. The work is also useful in active control system design.

8:30

2aSP3. Adaptive filtering algoritm for time-variant symmetric α-stable signals. Torsten Leth Elmkjær (Terma AS, Hovmarken 4, DK-8520 Lystrup, Denmark, tln@terma.com)

The existence of non-Gaussian distributed impulsive signals has roots in the *generalized central theorem*. In particular, the symmetric α-stable ($S\alpha S$) distribution has been used to model heavy-tailed phenomena encountered in communication, underwater acoustics, and radar. Most previous work in the field of active control uses some of the numerous variants of LMS algorithm extended with the *filtered-x* scheme. The stable processes, however, do not possess finite second-order moments. Hence, the filter should be based on l_p-norm optimization. The p-norm enters the expression for the tap-weight update $[p(t) < \hat{\alpha}(t)]$. In this work a new adaptive algorithm referred to as regularized normalized least−mean p−norm algorithm [ϵNLM$p(t)$] is presented. A running estimate $\hat{\alpha}(t)$ of the time−variant characteristic exponent of the stable signal is obtained. Hence, by adaptively tuning the filter p−norm a more optimal filter performance results in time−variant situations. Simulation analysis provides insight into the adaptive filter performance for various time−invariant and time−variant noisy signals for $1 < \alpha(t) \leqslant 2$. Moreover, it is demonstrated

that the filter performance is only marginally degraded compared with LMS algorithm for ordinary time−invariant Gaussian noise signals. Performance of an active control system exposed to time−varying heavy−tailed noise signals is presented.

8:45

2aSP4. Loudspeaker defect analysis using ultrasonic harmonic characterization. W. Scott Galway, Matthew A. LaBruzzo, Robert D. Celmer (Acoust. Program & Lab., Univ. of Hartford, 200 Bloomfield Ave., W. Hartford, CT 06117), and Daniel Foley (Independent Consultant, Marlborough, MA 01752)

This study presents the results of an ongoing investigation of defect detection in transducers. Loudspeaker assembly faults, such as a rubbing voice coil, bent frame, loose spider, etc., have traditionally been detected using experienced human listeners at the end of a production line, or through the use of low-order harmonics for total harmonic distortion measurement. The findings of Thompson *et al.* ["Higher-order harmonic signature analysis for loudspeaker defect detection," J. Acoust. Soc. Am. **114**, 2400 (2003)] described a new method wherein the total energy of high-order harmonics groups (10th through the 20th, or 31st through the 40th), were measured and analyzed. This approach demonstrated the ability to delineate distinct signatures that correlated to the root cause of audible rub and buzz distortions. This paper presents the results of a follow-up study using equipment that collected harmonic data up to 100 kHz. Results are presented for ten 5-in. midrange loudspeakers, in which trends for specific acoustic signatures were correlated to known defects. The use of a computer-based electroacoustic measurement and analysis system using frequencies that are inaudible but beneficial for analysis of defects are discussed. [Work supported by Listen Inc., Boston, MA.]

9:00

2aSP5. Adaptive collaborative array trajectories for optimum passive detection. Granger W. Hickman and Jeffrey L. Krolik (Dept. of Elec. and Comput. Eng., Duke Univ., Hudson Hall, Durham, NC 27708)

Two major challenges faced by sensor arrays used for passive sonar detection in littoral ocean channels are the suppression of highly dynamic interferers and signal wave-front mismatch resulting from complex multipath propagation. Large static arrays require accurate signal wave-front modeling and a highly stationary noise field to achieve their maximum potential array gain. In this paper, the detection performance of collaborative unmanned underwater vehicles (UUVs) towing short-line arrays is evaluated. In particular, vehicle trajectories which maximize overall probability of detection are computed as a function of the noise-field directionality and target location probability distribution. The UUV trajectories are constrained to share common endpoints connected by a series of line segments for each vehicle. The probability of at least one far-field detection by at least one UUV over the course of their transit is then maximized with respect to the sequence of the vehicles' headings. The results for a 2D horizontally isotropic noise field indicate that trajectories with a diversity of endfire (or equivalently broadside) bearings maximize detection performance. Scenarios with multiple interferers lead to solutions which minimize backlobe masking of the target. [Work supported by ONR.]

9:15

2aSP6. Data error covariance matrix for vertical array data in an ocean waveguide. Chen-Fen Huang, Peter Gerstoft, and William S. Hodgkiss (Marine Physical Lab., Scripps Inst. of Oceanogr., La Jolla, CA 92093-0238)

Information about the data errors is essential for solving any inverse problem. The likelihood function plays a critical role in describing the data uncertainties in geoacoustic inversion. The choice of likelihood function depends on the statistics of the errors (the difference between observed and estimated fields). In all work to date, the likelihood function has been derived based on an assumption of Gaussian errors. Typically, the errors are assumed to be independent, identically distributed with equal variance (referred to as the error variance), and the error variance is estimated as part of the optimization. Recently, there has been interest in estimating a more full data error covariance matrix. To estimate a truly full data error covariance matrix, we adopt a maximum-likelihood approach based on ensemble averages using the observed errors over many inversions. The approach is illustrated using data obtained during the ASIAEX 2001 East China Sea experiment. The parameter uncertainties resulting from incorporating the full data error covariance matrix are compared with those obtained from the simplified data error covariance matrix characterized by the error variance alone.

9:30

2aSP7. Increasing cross-talk immunity in simultaneous multiple acoustic sources measurements using coded signals. Shu Li and Ning Xiang (Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst., 110 8th St., Troy, NY, 12180)

Maximal-length sequences (MLS) and related signals demonstrate excellent correlation properties including a pulselike autocorrelation function and a relatively smaller-valued cross-correlation function. They have been increasingly applied in acoustical measurements [Xiang *et al.*, J. Acoust. Soc. Am. **117**, 1889–1894 (2005)]. Recent work employs specialized FFT-based algorithm [Daigle and Xiang, J. Acoust. Soc. Am. **119**, 330–335 (2006)] using MLS-related signals into simultaneous multiple acoustic sources measurements (SMASM). The resulting impulse responses of acoustic channels, however, suffer from cross-talk noise. To enhance cross-talk immunity, this paper proposes an algorithm using pseudoinverse algorithm to improve signal-to-noise ratio in the SMASM. This proposed algorithm applies specialized correlation method and pseudoinverse processing to implement acoustic measurements for the SMASM technique. This paper discusses both simulation and experimental results of the SMASM with the pseudoinverse algorithm to demonstrate considerable improvements in signal-to-noise ratio. This paper also evaluates the efficiency of pseudoinverse algorithm for the increased cross-talk immunity.

9:45

2aSP8. Multi-static target tracking using a bistatic invariance constraint. Jorge Quijano, Manish Velankar, and Lisa Zurk (Portland State Univ., 1900 SW Fourth Ave., Ste. 160-17, Portland OR 97207, zurkl@cecs.pdx.edu)

The ability to track moving targets in shallow water channels is critically effected by bottom reverberation which can generate false alarms and substantially increase the domain of possible tracks. Even with application of advanced tracking algorithms, false alarms degrade performance and increase computational burden. The invariance principle, which predicts structure in the frequency-time domain, has been applied to passive sonar systems but has not been considered for active systems. In this work, a mathematical expression (called the bistatic invariance principle, see companion paper) is proposed as a constraint to discriminate target detections from false alarms and thus reduce the search domain for multi-static tracking algorithms. The constraint does not require knowledge of the underwater environment and is thus robust to environmental uncertainty. The bistatic invariance principle provides a relationship between changes in the bistatic range of a moving target to changes in the source frequency, and this relationship can be used to track the target position over time in broadband active sonar systems. Simulations of the bistatic invariance principle for multi-static active systems are presented and discussed and multi-static tracking performance is characterized for several choices of target scattering matrices and different environmental conditions.

10:00−10:15 Break

10:15

2aSP9. A time-frequency approximation with applications in target recognition. Patrick Loughlin (Univ. of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261)

In active sonar, the received signal can often be modeled as a convolution of the transmitted signal with the channel impulse response and the target impulse response. Because the received signal may have a time-varying spectrum, due, for example, to target motion or to changes in the

channel impulse response, time-frequency methods have been used to characterize propagation effects and target effects, and to extract features for classification. In this presentation, we consider the Wigner time-frequency representation of a received signal modeled as the convolution of the transmitted signal with the channel and target responses. We derive a simple but very insightful approximation that shows the effects of the magnitude and phase of the frequency response of the target and of the channel on the Wigner representation of the transmitted signal. The results suggest possible features for classification, which we discuss. We also consider time-varying effects on the Wigner representation, such as changes in reflected energy due to changes in aspect angle, which we model by amplitude modulation. [Research supported by ONR Grant No. N000140610009.]

10:30

2aSP10. Model-based blind deconvolution for noisy dynamic channels. Michael Roan (Virginia Polytechnic Inst. and State Univ., Blacksburg, VA 24060) and Mark Gramann (Aerosp. Corp., Chantilly, VA)

It is common in acoustics to measure a signal that has been degraded by propagation through an unknown, noisy channel prior to measurement. While only the degraded measured signal is available, the data of interest are the original signal and the channel parameters. Often, it is desirable to reverse the filtering process by application of an inverse filter to recover the original signal. When neither the input signal properties nor the channel properties are known, this process is known as blind deconvolution (BDC). Typically, BDC algorithms assume noiseless, stationary propagation channels and input sources. These assumptions are usually violated in practical applications (e.g., noisy multipath propagation environments with moving source and receiver). To model these effects, predictive techniques are applied to incorporate *a priori* information about the system into the existing blind processing framework. The original contributions of this work follow. First, a novel formulation of the extended Kalman filter (EKF) is proposed. This allows incorporation of *a priori* information into gradient-based blind processing algorithms. This formulation is then applied to the natural gradient (NG) BDC algorithm. Finally, results are presented that suggest significant improvement in signal recovery performance over the NG BDC algorithm for dynamic noisy channels.

10:45

2aSP11. Detection of phase-coupled noises in the ocean. M. H. Supriya and P. R. Saseendran Pillai (Dept. of Electron., Cochin Univ. of Sci. and Technol., Kochi—682022, India)

Estimates of power spectral density are being used in a variety of signal processing applications. For the precise characterization of certain noise sources in the ocean, the phase blind power spectral estimation does not yield the required result. Estimates of the bispectrum and bicoherence have been found useful in detecting non-Gaussianity and nonlinearity, in system identification, as well as transient signals. The power spectrum, in general, fails to quantify any nonlinear interactions between the component frequencies. Such interactions induced by second-order nonlinearities give rise to certain phase relations called quadratic phase coupling (QPC). Bispectral analysis, which is based on the third-order cumulant sequence of the signal, can reveal the presence of as well as provide a measure to quantify such phase couplings. Bispectral analyses of various noise data records collected from different ocean noise sources have been carried out and it is seen that the bispectrum is not identically zero, implying the non Gaussian nature of the noise data. This paper presents the bispectral estimation approach for identifying the nonlinearities in the underwater noise-generating mechanisms as well as detecting the presence of quadratic phase couplings, thus leading to the identification of the noise sources.

11:00

2aSP12. Exploiting temporal coherence models to extend detection range. Ronald A. Wagstaff (Natl. Ctr. for Physical Acoust., University, MS 38677)

Turbulent flow over acoustic sensors and other sources of noise can severely corrupt signals and reduce detection range. The ranges can be much greater when based on the temporal coherences of signals and noise instead of their magnitudes. The signals need not be perfectly coherent to be detected at longer ranges, just more coherent than noise. This is illustrated using a fluctuation-based temporal coherence signal-processing model. This model uses measured data to produce temporal coherence patterns that are random for noise, deterministic for signal, and a blend of the two for small, e.g., negative, signal-to-noise ratio (SNR). Fluctuation-based processing techniques are presented for temporal coherence pattern generation, recognition, and automatic identification of signal presence or only noise in a frequency bin. It is shown for both underwater and atmosphere acoustic data that a signal's influence in a frequency bin alters the usual random ''noiselike'' temporal coherence patterns by causing them to cluster. Hence, signal-altered patterns can be recognized and the signals preserved. Corresponding ''noise-only'' patterns, being random, can also be recognized and attenuated, or eliminated. Using these models results in increased SNR and automatic signal detection at longer ranges with greater confidence. [Work supported by ARDEC and SMDC.]

11:15

2aSP13. Generating simulated reverberation using noiselets. Edmund J. Sullivan, Robert P. Goddard (Prometheus Inc., 103 Mansfield St., Sharon, MA 02067), Hyman A. Greenbaum (Anteon Corp., Middletown, RI 02842), and Kevin P. Bongiovanni (Naval Undersea Warfare Ctr., Newport, RI 02841)

Of the approaches to generating stochastic realizations of reverberation time series, the simple point scatterer model (PSM) is appealing because the statistics evolve naturally as the number of scatterers increases. Unfortunately, the number of eigenrays required is often overwhelming, especially for broadband signals. The more common ''variance factorization model'' (VFM) uses far fewer eigenrays to compute a time-dependent estimate of the variance of the reverberation, and applies the spectral factorization theorem to generate a sequence of short-time spectra, which is inverse transformed and concatenated to form the desired time series. However, the variance and factorization steps scale with the square and the cube, respectively, of the number of receiver channels, so the advantage is lost for element-level simulations of complex receivers. The ''noiselet model'' is conceptually similar to the PSM, but the transmit pulse is replaced by an ensemble of precomputed ''noiselets,'' which are sums of randomly weighted and delayed copies of the original transmit pulse. The number of eigenrays is thus comparable to the VFM, but scaling is linear in the number of channels, rendering this method tractable for real-time broadband element-level simulation. Examples are shown for a linear FM pulse at 10 kHz and 400-Hz bandwidth.

11:30

2aSP14. A phenomenological model for sonic crystals based on artificial neural networks. Elies Fuster-Garcia, Vicent Romero-Garcia, Juan V. Sanchez-Perez (ACARMA and Dep. de Fisica Aplicada, ETSICCP, UPV, Cno. de Vera s/n, 46022 Valencia, Spain, jusanc@fis.upv.es), Luis M. Garcia-Raffi, and Enrique A. Sanchez-Perez (ETSICCP, UPV, 46020 Valencia, Spain)

In this paper we develop a phenomenological model that simulates the acoustic attenuation behavior of sonic crystals. The input of the model is a

set of parameters that characterizes each experimental setup and the output is the associated simulated attenuation spectrum. The model consists of a combination of multiresolution analysis based on wavelet functions and a set of artificial neural networks. An optimized coupling of this tool allows us to reduce drastically the experimental data needed and also to have a fast computational model that can be used for technological purposes. [This research has been partially supported by the Generalitat Valenciana (Spain), under Grant GV04B-371, by Comision Interministerial de Ciencia y Tecnologia of Spain, under Grant BFM2003-02302, and by Comision Interministerial de Ciencia y Tecnologia of Spain, Contract Ns. MAT 2003-04993-03.]

11:45

2aSP15. Nonlinear behavior of interacting harmonics in audio spectra. V. Vijayakumar and C. Eswaran (Multimedia Univ., FOSEE, Jalan AIr Keroh Lama, 75450, Melaka, Malaysia)

In this paper we will show how individual harmonics of a frequency-modulated dynamic audio spectrum undergo interactions. This results in nonlinear behavior of the spectrum. We have illustrated the nonlinear behavior with respect to the number of harmonics and the modulation parameter. It is shown that an audio spectrum containing higher-order harmonics is more susceptible to interation among itself. Musical instrument spectra are used to illustrate this effect and audio samples are provided.

TUESDAY MORNING, 6 JUNE 2006 ROOM 553AB, 8:00 A.M. TO 12:00 NOON

Session 2aUW

Underwater Acoustics, Acoustical Oceanography, and Animal Bioacoustics: Joe Blue Memorial Session I: Shallow Water and Marine Animal Acoustics

Edmund R. Gerstein, Cochair
Leviathan Legacy Inc., 1318 S.W. 14th St., Boca Raton, FL 33486

George V. Frisk, Cochair
Florida Atlantic Univ., Dept. of Ocean Engineering, 101 North Beach Rd., Dania Beach, FL 33004-3023

Chair's Introduction—8:00

Invited Papers

8:05

2aUW1. The direct path impulse response in bubbly shallow water. Ralph Goodman, Jerald Caruthers, and Timothy Ruppel (Stennis Space Ctr., Univ. of Southern Mississippi, Bay St. Louis, MS 39529)

It is well known that bubbles, even in small numbers, cause significant changes in the acoustic properties of seawater. In shallow water bubbles enter the sea, either by wind-driven breaking waves or by surf-induced rip currents. Both have been observed in acoustic transmission experiments [IEEE, JOE **25**, 507–515 (2000); J. Acoust. Soc. Am. **106**, 617–625 (1999)]. Wind-driven bubbles were observed, during a period of high winds, to reduce sound speed from the undisturbed sea condition of 1520 m/s to about 1494 m/s. From bubble theory, it is estimated that the average bubble void fraction was approximately 2.6×10^{-6}. For a distribution of bubble radius size, $n(a)$, there will be a frequency-dependent sound speed and absorption related by the Kramers-Kronig relation, from which the change in a pulsed signal can be determined. By using a bubble size distribution of $n(a)=Na\times10^{-4}$, for radii between values of 200 and 10 microns and the void fraction estimated above, the dispersion and absorption are calculated. The impulse response can be calculated as a function of range. [Work supported by the Office of Naval Research.]

8:25

2aUW2. Modeling nonlinear acoustics in shallow water. B. Edward McDonald (Naval Res. Lab, Washington, DC 20375, mcdonald@sonar.nrl.navy.mil) and William A. Kuperman (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

Lessons learned in modeling nonlinear acoustics and weak shock propagation in deep water are helpful in application to shallow water (mine countermeasures in particular), but the medium is much more complicated. Numerous improvements are needed; Joe Blue's work touched on several of them. First is the need to treat higher propagation angles. Second is the presence of bubbles near the surface and possibly in the bottom. Finally, there are problems introduced by nonlinear propagation in granular sediments. These include greatly increased nonlinearity parameter and time-domain formulation of frequency—linear attenuation. Recent progress in the application of the NPE model [McDonald and Kuperman, "Time domain formulation for pulse propagation including nonlinear behavior at acoustic," J. Acoust. Soc. Am. **81**, 1406–1417 (1987)] to mine countermeasures will be outlined. [Work supported by the Office of Naval Research.]

8:45

2aUW3. Sound waves and shear waves in sediments. Michael J. Buckingham (MPL, SIO, UCSD, 9500 Gilman Dr., La Jolla, CA 92093-0238, mjb@mpl.ucsd.edu)

In a long and illustrious career, Joe Blue made many significant contributions to the discipline of underwater acoustics, although he is probably best known for his work on the design and calibration of hydrophones and hydrophone arrays. Such instrumentation has many applications, including measurements of the geoacoustic properties of marine sediments. In this paper, recent developments on the properties of sound waves and shear waves in sediments will be summarized and certain difficulties that are encountered when attempts are made to match theory to data will be highlighted. Central to the discussion are the compressional-wave and shear-wave dispersion relations, that is, phase speed and attenuation as functions of frequency. In a saturated sediment, the phase speeds and attenuations also depend on various physical properties, for example, the porosity, density, grain size, and overburden pressure (depth). These dependencies will be illustrated using extensive data sets taken from the literature. Any satisfactory theory of waves in sediments must yield correctly not only the dispersion relations but also all the dependencies on the physical properties. A recently developed theory that satisfies these criteria will be discussed in the context of the available data. [Research supported by ONR.]

9:05

2aUW4. Joe Blue's influence on low-frequency, shallow-water acoustics. George V. Frisk (Florida Atlantic Univ., 101 N. Beach Rd., Dania Beach, FL 33004)

As Superintendent of the Naval Research Laboratory's Underwater Sound Reference Division, Joe Blue played a key role in the advancement of low-frequency, shallow-water acoustics. Specifically, his leadership and oversight of this unique facility provided an engineering science environment in which low-frequency transducers were developed, calibrated, and maintained. These assets were made available to the research community in ocean acoustics at reasonable cost and have offered dependable, reliable service for more than 25 years. This talk focuses on advances made using the J15-3 transducer, which emerged as a workhorse in numerous experiments in low-frequency (50–500 Hz), shallow-water acoustics. A series of experiments is described in which this transducer was used as the principal source of sound. These measurements led to an increased understanding of acoustic propagation, geoacoustic inversion, and physical oceanography in shallow water. [Work supported by ONR.]

9:25

2aUW5. A long-range, high-resolution Doppler sonar for fishery stock assessment. Harry A. DeFerrari (RSMAS—Univ. of Miami, 4900 Rickenbacker Cswy., Miami, FL 33149)

A new approach is developed for horizontal-looking active sonar that uses continuous bi-/multistatic transmission of very long period m-sequences. Compared to conventional monostatic ping rather than listen sonar, this approach improves the Doppler resolution by a factor of 20 and increases FOM by perhaps 13 dB by increasing the temporal coherent processing gain. The improved time/Doppler resolution helps to discriminate fish school from bottom reverberation, and improves the resolution of taxonomic parameters of fish assemblages. The improved FOM allows coverage of large areas of shallow oceans with source levels low enough to avoid environmental concerns for marine mammals. The principal obstacle to continuous transmission sonar operation is the masking of the target return by the time and frequency leakage of the direct path arrivals. Unique time-correlation property of m-sequences elimination time leakage and a process calls coordinate zeroing (CZ) eliminates all the Doppler leakage. A linear temporal Doppler search (LTDS) algorithm is used to shift the detection problem to a high-resolution pulse arrival-time/Doppler plane with a noise-limited background. Ultrafast Hadamard transforms lighten the computational burden. The combination of these ideas is presented including the theory, numerical experiments that demonstrate the concept, and verification with real ocean signals.

9:45

2aUW6. Development and application of high-frequency sonars for imaging naturally occurring plumes in the ocean. David R. Palmer (NOAA/Atlantic Oceanogr. and Meteorological Lab., 4301 Rickenbacker Cswy., Miami, FL 33149)

Hydrothermal plumes are formed when hydrothermal solutions are discharged from chimneylike vents on the ocean floor where adjacent tectonic platesare separating. Their study provides information on the transfer of heatand chemicals from the solid earth to the ocean. When high-frequency acousticswas proposed for imaging and studying hydrothermal plumes, Joe Blue contributedin a general way to the development of the effort. In this presentation the history of this effort is reviewed. After an introduction to hydrothermal venting, some of the problems addressed in the development and application of the sonar are discussed. These include sonar calibration, identification of the dominant scattering mechanism, estimating cross sections, and the intrinsic statistical variability associated with incoherent scattering. The results obtained from four sea trials are recalled. These trials involved either three-person, deep-diving submersibles or the JASON/MEDEA ROV system. The opportunities provided by the somewhat surprising progress that has occurredare discussed. Finally, an application of the developed technology is mentioned, acoustical imaging of the bubble plume emitted from the underwater volcano Kick em Jenny.

10:05–10:25 Break

10:25

2aUW7. The art of merging physics and math with biology and psychology to study marine animal acoustics. Mardi C. Hastings (Office of Naval Res., Life Sci. Res. Div., 875 N. Randolph St., Arlington, VA 22203-1995, mardi_hastings@onr.navy.mil)

Joe Blue's most recent contributions to marine animal acoustics resulted from successful collaborations with biologists and psychologists, each bringing their individual expertise to the table and working together to understand how the characteristics of the acoustic field generated by a moving vessel could provide critical insight into the observed behavior of right whales and manatees.

Joe's hands-on approach of going out and actually recording and measuring the sound fields provided data that not only verified his analyses, but also enhanced practical understanding of acoustics in the animal bioacoustics community. His work demonstrated the importance of physics and math in studying complex multidisciplinary acoustic problems and resolving perplexing dichotomies. Examples will be presented showing that Joe's collaborative style is a model for everyone to follow.

10:45

2aUW8. Simulation models of click production in dolphins. Nikolay A. Dubrovskiy and Igor A. Urusovskiy (Andreyev Acoust. Inst., Shvernika Ul, #4, Moscow 117036, Russia)

The simulation model (SM) of dolphin click source originally presented by [Dubrovskiy and Giro, ''Echolocation in Bats & Dolphins,'' Chap. 10, 59–63 (2003)] was further developed by Gladilin, Dubrovskiy, Mohl, and Walberg, Acoust. Phys. **50**, 463-468 (2004). The SM describes movements of a solid sphere frozen in a compliant ring. A jet of air in a duct connecting air cavities brings the sphere into self-oscillations. The SM allows calculating the waveform and sound pressure of the generated clicks. Here the mathematical and simulation model (MSM) of click production in dolphins is further developed. The MSM describes radiation of clicks by bodies comprising of waterlike tissue with different shapes (a ring, a blunted cone) pulsating and (or) oscillating under action of muscles tension. Calculations are made for the energy efficient case when acting force is consecutively applied not to the entire body but to its thin layer, which resulted in action of the progressing wave array. Directivity patterns of such a waterlike compliant body in main directions are similar to that of a rigid body with the same shape and dimensions. Comparisons are made of MSM predictions with properties of an actual radiation field of dolphins (in clicks wave forms, sound pressure levels, and directivity patterns) and dimensions of biological structures supposedly responsible for click production in dolphins.

11:05

2aUW9. A dolphin acoustic tapetum? Sam Ridgway, Brad Blankenship, Dorian Houser, Don Carder (SPAWAR Systems Ctr. San Diego, Div. D235, 53560 Hull St., San Diego, CA 92152-5001), and Carl Hoh (Univ. of California, San Diego, CA 92093)

Many mammals including dogs, cats, deer, and dolphins, but not humans, have a reflective membrane at the back of the eye. This tapetum lucidum affords the light-sensitive retinal cells a second chance for photon-photoreceptor stimulation, enhancing low light visual sensitivity. Dissection of preserved specimens and computed tomography (CT) scans of live dolphins have shown prominent air cavities around the ear, especially medial and posterior. Histologic evaluation of peri-otic tissue revealed epithelial lined air sinuses and erectile tissue containing nerves, ganglia, smooth muscle bands, vascular spaces, and periarterial venous networks. Two dolphins were trained to periodically recognize and respond to a 1.5-s, 70-kHz target tone. In three experiments, 20 target tones were presented through a jaw phone after injection of 18 Fluro-deoxy-d-glucose (FDG). Subsequent positron emission tomography (PET) scans revealed that the tone, to which the dolphin had responded and been rewarded numerous times in the past, elicited a marked increase in metabolism in the erectile tissue medial to the stimulated ear but not the contralateral ear. Tissue around reflective air sinuses medial to the ear become active with acoustic stimulation, suggesting that sinus air may be shaped so as to provide an ''acoustic tapetum'' for dolphins.

11:25

2aUW10. Of manatees and men: A future: Thanks to you, Joe. Edmund Gerstein, Laura Gerstein, Joseph Blue (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), and Steve Forsythe (Naval Undersea Warfare Ctr., Newport, RI 02841)

A comprehensive series of underwater psychoacoustic tests was conducted to measure the hearing abilities of the West Indian manatee under varying acoustic conditions. Pure tones, complex noise and real sounds, boat noise and species-specific calls, were presented to captive manatees under various acoustical conditions. The results from more than 30 000 threshold trials define the first behavioral audiograms, directional hearing, and critical masking ratios for any sirenian species. Psychoacoustic tests demonstrate the noise produced by slow-moving boats is more difficult for manatees to hear and locate than the high-energy cavitation noise from faster moving vessels. The dominant frequency spectra of slow-moving boats fall near the fringe of their low-frequency hearing limits and below critical masking ratios. Complementing underwater ambient noise and vessel noise measurements revealed physical propagation factors that further render manatees vulnerable to watercraft collisions in their shallow-water habitats. Ironically, slow speed zones implemented to protect manatees may exacerbate the problem, making vessels difficult to hear, while increasing transect times and opportunities for collisions. By exploiting the manatee's best hearing capabilities, it may be possible to mitigate collision risks with a low-intensity, environmentally friendly acoustic warning device that alerts manatees of approaching vessels. [Research supported by DOD Navy Legacy Program, Inland Navigation District, FFWCC]

11:45

2aUW11. Listening to fish: Applications of passive acoustics to fisheries. Rodney Rountree, Francis Juanes (Dept. of Natural Resources Conservation, UMass Amherst, Holdsworth Hall, Amherst, MA 01003, rrountree@fishecology.org), and Cliff Goudey (MIT Sea Grant College Program, Cambridge, MA 02139)

Passive acoustics is a rapidly emerging field of marine science that until recently has received little attention from fishery scientists and managers. In its simplest form, it is the act of listening to the sounds made by fish and using that information as an aid in locating fish so that their habitat requirements and behaviors can be studied. We believe that, with the advent of new acoustic technologies, passive acoustics will become one of the most important and exciting areas of fisheries research in the next decade. However, widespread unfamiliarity with the technology, methodologies, and potential of passive acoustics has hampered the growth of the field and has limited funding opportunities. Herein, we provide an overview of important new developments in passive acoustics together with a summary of research, hardware, and software needs to advance the field.

Meeting of the Standards Committee Plenary Group

to be held jointly with the meetings of the

ANSI-Accredited U.S. Technical Advisory Groups (TAGs) for:
ISO/TC 43, Acoustics,
ISO/TC 43/SC 1, Noise,
ISO/TC 108, Mechanical vibration and shock,
ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures,
ISO/TC 108/SC 3, Use and calibration of vibration and shock measuring instruments,
ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock,
ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines,
ISO/TC 108/SC 6, Vibration and shock generating systems,
and
IEC/TC 29, Electroacoustics

P. D. Schomer, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 43, Acoustics and ISO/TC 43/SC 1 Noise
2117 Robert Drive, Champaign, IL 61821

D. J. Evans, Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108, Mechanical Vibration and Shock and Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 3 Use and calibration of vibration and shock measuring devices
National Institute of Standards and Technology (NIST), 100 Bureau Drive, Stop 8220, Gaithersburg, MD 20899

A. F. Kilcullen, Co-Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2, Measurement and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
734 Walden Road, Hedgesville, WV 25427

R. Taddeo, Co-Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 2, Measurment and evaluation of mechanical vibration and shock as applied to machines, vehicles and structures
1333 Isaac Hull Avenue, SE, Washington Navy Yard, Washington DC 20376

D. D. Reynolds, Chair, U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 4, Human exposure to mechanical vibration and shock
3939 Briar Crest Court, Las Vegas, NV 89120

D. J. Vendittis, Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines
701 Northeast Harbour Terrace, Boca Raton, FL 33431

R. Taddeo, Vice Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 5, Condition monitoring and diagnostics of machines
1333 Isaac Hull Avenue, SE, Washington Navy Yard, Washington DC 20376

G. Booth, Chair of the U.S. Technical Advisory Group (TAG) for ISO/TC 108/SC 6, Vibration and shock generating systems
44 Bristol Street, Branford, CT 06405-4842

V. Nedzelnitsky, U.S. Technical Advisor (TA) for IEC/TC 29, Electroacoustics
National Institute of Standards and Technology (NIST), Sound Building, Room A147, 100 Bureau Drive, Stop 8221, Gaithersburg, MD 20899-8221

The reports of the Chairs of these TAGs will not be presented at any other S Committee meeting.

The meeting of the Standards Committee Plenary Group will precede the meetings of the Accredited Standards Committees S1, S2, S3, and S12, which are scheduled to take place in the following sequence:

ASC S1 Acoustics	6 June 2006	1:45 p.m. to 3:15 p.m.
ASC S12 Noise	6 June 2006	3:30 p.m. to 5:00 p.m.
ASC S2 Mechanical Vibration and Shock	7 June 2006	8:30 a.m. to 10:00 a.m.
ASC S3 Bioacoustics	7 June 2006	10:30 a.m. to 12:00 noon

Discussion at the Standards Committee Plenary Group meeting will consist of national items relevant to all S Committees.

The U.S. Technical Advisory Group (TAG) Chairs for the various international Technical Committees and Subcommittees under ISO and IEC, which are parallel to S1, S2, S3, and S12 are as follows:

U.S. TAG Chair/Vice Chair	TC or SC	U.S. Parallel Committee
ISO		
P. D. Schomer, Chair	**ISO/TC 43** Acoustics	S1 and S3
P. D. Schomer, Chair	**ISO/TC 43/SC1** Noise	S12
D. J. Evans, Chair	**ISO/TC 108** Mechanical Vibration and Shock	S2
A. F. Kilcullen, Co-Chair R. Taddeo, Co-Chair	**ISO/TC 108/SC2** Measurement and Evaluation of Mechnical Vibration and Shock as Applied to Machines, Vehicles and Structures	S2
D. J. Evans, Chair	**ISO/TC 108/SC3** Use and Calibration of Vibration and Shock Measuring Instruments	S2
D. D. Reynolds, Chair	**ISO/TC 108/SC4** Human Exposure to Mechanical Vibration and Shock	S3
D. J. Vendittis, Chair R. F. Taddeo, Vice Chair	**ISO/TC 108/SC5** Condition Monitoring and Diagnostics of Machines	S2
G. Booth, Chair	**ISO/TC 108/SC6** Vibration and Shock Generating Systems	S2
IEC		
V. Nedzelnitsky, U.S. TA	**IEC/TC 29** Electroacoustics	S1 and S3

TUESDAY AFTERNOON, 6 JUNE 2006 ROOM 555AB, 1:00 TO 4:45 P.M.

Session 2pAAa

Architectural Acoustics, Musical Acoustics, and Engineering Acoustics: Surround Sound Essentials II

Alexander U. Case, Cochair
Fermata Audio + Acoustics, P.O. Box 1161, Portsmouth, NH 03802

K. Anthony Hoover, Cochair
Cavanaugh Tocci Associates, 327F Boston Post Rd., Sudbury, MA 01776

Invited Papers

1:00

2pAAa1. Mastering music in a 5.1 world. Adam Ayan (Gateway Mastering Studios, 428 Cumberland Ave., Portland, ME 04101, adam@adamayan.com)

The evolution of 5.1 surround sound in the world of popular music has changed the way artists, producers, and engineers conceptualize their recordings. Along with the artistic benefits of 5.1 surround production come the challenges of working in this relatively new format. Mastering engineers are responsible for making the final sonic enhancements to a recording, musically maximizing the recording, and making it compatible and competitive in the marketplace. 5.1 surround mastering shares many of the same techniques and philosophies as stereo mastering, but also presents a new set of concerns for the mastering engineer. This talk focuses on the specific production and engineering concerns that are unique to the world of 5.1 surround production, specifically as they relate to mixing and mastering techniques.

1:20

2pAAa2. The real reasons you should invest in a surround-sound system. Barbara G. Shinn-Cunningham (Boston Univ. Hearing Res. Ctr., 677 Beacon St., Boston, MA 02215)

Most consumers of surround-sound systems understand that increasing the number and locations of speakers provides better control of perceived source location. The first goal of this talk is to explore why surround-sound systems generally work well in a variety of common settings, even though the acoustic effects of the rooms in which they are used can differ dramatically. What acoustic and psychoacoustic principles come into play? A second goal is to discuss the different perceptual benefits provided by a typical surround-sound system. Specifically, a good system not only allows listeners to hear sources at different directions, but also improves a listener's ability to understand simultaneous sources and monitor a complex auditory scene. Finally, we will discuss why it is important to realize that surround-sound systems are entertainment systems, not simulators. If asked, most consumers would probably argue that surround sound is good because it makes it possible to recreate a realistic auditory experience. In fact, the aim of many recordings and games that use surround-sound technology is not realism, but vicarious thrill and vivid experience. In entertainment, bigger (and louder and more extreme) is better than everyday, ordinary sensations. [Work supported by NIDCD.]

1:40

2pAAa3. The influence of changes in diffusive acoustical treatment on spatial imagery associated with multichannel sound reproduction. William L. Martens (Schulich School of Music, McGill Univ., Montreal, QC, Canada)

Listening experience in the newly constructed, largely diffuse small room at Blackbird Studios has led to the design of a new research space for the Sound Recording program in McGill University's Schulich School of Music. The space was created to allow for experimental investigation of the influence of changes in diffusive acoustical treatment on variation in auditory imagery associated with multichannel sound selections while holding listener location and orientation constant relative to a fixed five-channel loudspeaker array. Preliminary tests suggest that a number of benefits result when low-amplitude indirect sound is provided by diffusive room acoustic treatment within the first 30 ms of the arrival of the loudspeaker signals. Both for conventional multitrack mixes, and for recordings made using multichannel microphone arrays, these benefits include greater Listener EnVelopment (LEV), improved segregation of simultaneously sounding musical sources, and more solid phantom images of virtual sources, especially those located to the listener's extreme sides between front and rear loudspeakers (between 30 and 110 deg of azimuth to the left and right). This paper will summarize those results, and describe the specially designed system of variable acoustic treatment that enables such direct experimental investigation. [Work supported by Canada Foundation for Innovation.]

2:00

2pAAa4. Low-frequency control options in surround-sound critical listening rooms. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

Acoustically small rooms, which are typically used for critical listening, be it mono or surround, can potentially overwhelm the spatial and spectral textures that contribute to an enjoyable listening experience. The options to control potential modal and speaker boundary interference in rooms include optimal dimensional room ratios, optimal speaker/listener placement, optimal multiple subplacement and dedicated low-frequency absorbers, including quarter-wave traps, anechoic wedges, Helmholtz and membrane resonators, and a relatively new and potentially significant addition, the metal plate resonator, which absorbs efficiently down to 50 Hz with a thickness of only 4 in. The design theory, proof-of-performance evaluation testing methods, and optimal placement guidelines for all of these tools will be reviewed.

2:20

2pAAa5. From mono to surround: A review of critical listening room design and a new immersive surround design proposal. Peter D'Antonio (RPG Diffusor Systems, Inc., 651-C Commerce Dr., Upper Marlboro, MD 20774)

As our critical listening rooms have evolved from mono to stereo to an ill-fated attempt at quad to full surround, many approaches have been taken in the design of these spaces. This presentation will review significant milestones in this evolution including the early compression ceilings, reflective front-absorptive rear, absorptive front-absorptive rear nonenvironment, absorptive front-diffusive rear, spatio-temporal reflection-free zone front-diffusive rear, and reflection-rich zone in front-diffusive rear. Recently there has been a renaissance in the evaluation of the importance of specular and diffuse reflections in auditory perception. The presentation will conclude with a proposal for a new immersive surround environment, which can be described as an ambient anechoic or ambechoic environment. Fully enveloping diffuse room reflections are 30 dB below the direct sound, as in an anechoic environment, yet the room feels ambient, provides precise localization, and offers a creative work environment for recording and monitoring. An example of a proof-of-concept room will be illustrated and described.

2:40–3:00 Break

3:00

2pAAa6. An innovative professional music recording/monitoring space. George Massenburg (2805 Bransford Ave., Nashville, TN 37204-3101)

A new kind of professional recording/monitoring space has been designed and constructed as part of the Blackbird Studio complex in Nashville, Tennessee. The room will be introduced and the early subjective impressions will be discussed. The objective of this new room was threefold: first, to provide a surround mixing space with much-improved virtual source localization over a broader footprint in the space, second, to provide a highly diffuse space for the recording of live music performances by one or more vocalists and/or

musicians, and third, to investigate a largely diffuse monitoring environment as it might benefit the quality and/or efficiency of professional music mixing in different formats. This presentation will discuss some of the results and some of the surprise outcomes, including practitioner-reported improvements in quickly attaining comprehensive, surgical equalization of sources, possibly attributable to an environment with somewhat more equal reverb times across the frequency spectrum.

3:20

2pAAa7. Surround-sound wish list. Bob Ludwig (Gateway Mastering Studios, Inc., 428 Cumberland Ave., Portland, ME 04101, staff@gatewaymastering.com)

This presentation describes this engineer's involvement with surround sound since the days of vinyl LP, 8-track, and four-channel Quadrophonic sound. It includes my experiences as one of the few mastering engineers hired by Digital Theater Systems in 1997 to master and help establish their new 5.1 surround compact disc. In 2005, the author mastered 2 of the 5 Grammy-nominated ''Best Surround Sound discs''—the SA-CD Dire Straits 20th Anniversary edition of ''Brothers in Arms'' and the Foo Fighters ''In Your Honor'' DualDisc. This presentation also includes a short wish list of issues with popular music surround-sound mixes with some suggestions for improvement.

Contributed Papers

3:40

2pAAa8. Simultaneous equalization of a loudspeaker and a room: Insights from a one-dimensional model. Xi Chen and Timothy Leishman (Acoust. Res. Group, Dept. of Phys. and Astron., Brigham Young Univ., Provo, UT 84602)

In recent years, many audio practitioners have concluded that a room excited by a loudspeaker cannot be effectively equalized. They argue that the free-field response of the loudspeaker may be equalized, but that the room response itself must be ignored. This paper presents a one-dimensional model to illustrate many of the difficulties encountered in past equalization methods that have led to this way of thinking. It explores the spatial-spectral variations of the reproduced sound field that make the methods problematic. It then uses the model to explain the benefits of using total energy density as a basis for equalization instead of acoustic pressure. Both the loudspeaker and the enclosed field can then be simultaneously equalized for nearly optimal results. The method is also implemented effectively using adaptive filtering techniques.

3:55

2pAAa9. A hybrid approach to rendering multichannel audio, using wave-field synthesis and stereophonic techniques, with virtual microphone control. Daniel Valente and Jonas Braasch (Rensselaer Polytechnic Inst., Troy NY 12188)

In the evolution of multichannel audio presentation, a number of formats have emerged including 5.1 surround and wave-field synthesis (WFS). Due to the complexity and number of components required to realize an immersive WFS system, the use in a home theater environment is currently impractical. This study aims to investigate the perceptual advantages of integrating both systems into a hybrid array capable of providing immersive sound for a number of listeners, at a scale that is attainable for the consumer in terms of budget and complexity. In this approach the three front speakers (left, center, and right) are replaced with a linear line array allowing for a better lateral and depth imaging, while increasing the sweet spot. The rendering will be created using Virtual Microphone Control (ViMiC) for both WFS, simulating a line of omnidirectional microphones with ViMiC, and surround sound, which will allows dynamic source movement, speaker positioning, and scalability. The advantage of using this hybrid approach is the application of WFS to home theater systems without facing the complexity of a complete two-dimensional WFS array, while still allowing for an immersive sound presentation. A psychoacoustic evaluation will be presented, as well as measurements of the composed sound field.

4:10–4:45
Panel Discussion

Session 2pAAb

Architectural Acoustics and National Council of Acoustical Consultants: Student Design Competition

Robert C. Coffeen, Cochair
Univ. of Kansas, School of Architecture and Urban Design, Lawrence, KS 66045

Norman H. Philipp, Cochair
Yantis Acoustical Design, 720 Olive Way, Ste. 1400, Seattle, WA 98191-1878

Byron W. Harrison, Cochair
The Talaske Group, Inc., 105 N. Oak Park Dr., Oak Park, IL 60301

The Technical Committee on Architectural Acoustics of the Acoustical Society of America and the National Council of Acoustical Consultants are sponsoring this Student Design Competition that will be professionally judged at this meeting. The purpose of this design competition is to encourage students enrolled in architecture, architectural engineering, and other university curriculums that involve building design and/or acoustics to express their knowledge of architectural acoustics in the design of a facility in which acoustical considerations are of significant importance. Submissions will be poster presentations. This competition is open to undergraduate and graduate students from all nations.

The submitted designs will be displayed in this session and they will be judged by a panel of professional architects and acoustical consultants. An award of $1,250 will be made to the submitters of the entry judged "First Honors." Up to four awards of $700 each will be made to submitters of entries judged "Commendation."

Session 2pAB

Animal Bioacoustics, Acoustical Oceanography, and Underwater Acoustics: Effects of Anthropogenic Sounds on Fishes

Arthur N. Popper, Cochair
Univ. of Maryland, Dept. of Biology, College Park, MD 20742

Richard R. Fay, Cochair
Loyola Univ., Parmly Hearing Inst., 6430 N. Kenmore, Chicago, IL 60626

Chair's Introduction—1:00

Invited Papers

1:10

2pAB1. Anthropogenic sound—Introduction and overview of the ambient and anthropogenic environment. Arthur N. Popper (Dept. of Biol., Univ. of Maryland, College Park, MD 20742) and Richard R. Fay (Loyola Univ., Chicago, IL 60626)

Interest in anthropogenic sound in the marine environment has been directed at concerns about the effects of sound on marine mammals. However, fish make up a far larger and more diverse portion of the oceans than do marine mammals, and they hold considerable economic importance internationally. This has led to a growing interest in the effect of anthropogenic sounds on fish in environments ranging from shallow waters near in-shore shipping lanes to the deep sea. Concerns for effects on fish are parallel to those for marine mammals, and include issues ranging from the death of individual animals to the potential for behavioral changes that could lead to impacts on the survival of populations or species. While the body of data on the effects of anthropogenic sounds on fish is still small, and much of the work is not in the peer-reviewed literature, the number of peer-reviewed studies is growing. The purpose of this special session is to provide an overview of what we currently know about the effects of anthropogenic sound on fish, help define the major outstanding questions on these effects, and to provide the basis for discussion of current and future research in order to help resolve these questions.

1:35

2pAB2. Attempting to define sound exposure metrics for fish. Mardi C. Hastings (Office of Naval Res., Life Sci. Res. Div., 875 N. Randolph St., Arlington, VA 22203-1995, mardi_hastings@onr.navy.mil)

One of the big issues in addressing the biological effects of sound on humans and animals is metrics. For fishes the issue is further complicated because their inner ear responds directly to acoustic particle motion and in some species, indirectly to pressure through multiple pathways. So, metrics based only on pressure can be misleading in developing sound exposure criteria. The decibel scale also confounds the problem because it is used to express not only levels of sound pressure, but also levels of sound power, intensity, and energy. Even for just sound pressure, decibels are used to express peak, peak-to-peak, and time-averaged values, which can hamper extrapolation of laboratory data to anthropogenic sources in the field. Add to that the changes in decibel reference values when going from air to water, more than 25 000 different fish species, and the fact that a significant amount of sound exposure research on fish has been done in small enclosures at low frequencies, and one quickly realizes the difficulties in developing meaningful relationships between sound exposure and observed effects. This presentation will focus on relevant sound exposure metrics, their practical definitions, and their application to quantifying the biological effects of sound on fish.

2:10

2pAB3. Effects of high-intensity sonar on fish. Michele B. Halvorsen, Lidia Eva Wysocki, and Arthur N. Popper (Dept. of Biol., Univ. of Maryland, College Park, MD 20742, apopper@umd.edu)

Studies investigated the effects of exposure to low-frequency active (LFA) sonar on rainbow trout, channel catfish, and pumpkinseed sunfish using an element of the SURTASS LFA. Animals were exposed to sounds that had a received peak level of 193 dB *re*: 1 uPa (rms) (188.5 dB *re*: 1 uPa^2 s SEL). Measurements were made of hearing sensitivity using auditory brainstem response and of behavioral responses. Additional studies were done on pathology of the ear and other organ systems. Trout and catfish both showed temporary threshold shifts immediately after sound exposure, while sunfish showed no threshold shift. However, the amount of threshold shift and frequencies concerned varied. In addition, trout and catfish showed immediate behavioral responses to the sound onsets and in relative distribution during sound presentation. Details of the responses differed between species. There was no mortality in any species attributable to sound exposure during or after experiments. SEM examination of hair cell ciliary bundles of the inner ears did not show any damage. Similarly, there were no effects on the swim bladder or nonauditory tissues. Overall, the effects of LFA on fish depend upon species and appear to be temporary and relatively modest. [Work supported by Chief of Naval Operations.]

2:35

2pAB4. Effects of riverine seismic air-gun exposure on fish hearing. David A. Mann (Univ. of South Florida, College of Marine Sci., 140 7th Ave. S., St. Petersburg, FL 33701, dmann@marine.usf.edu), Peter Cott, Bruce Hanna (Dept. of Fisheries and Oceans, Yellowknife, NT, Canada X1A 1E2), Alex MacGillivray, Melanie Austin (JASCO Res. Ltd., Victoria BC, Canada V8Z 7X8), Michael Smith, and Arthur N. Popper (Univ. of Maryland, College Park, MD 20742)

Seismic airguns produce considerable amounts of acoustic energy that have the potential to affect marine life. This study investigated the effects of exposure to an airgun array in the Mackenzie River Delta on the hearing of three fish species: northern pike (*Esox lucius*), broad whitefish (*Coregonus nasus*), and lake chub (*Couesius plumbeus*). Fish were placed in cages in 1.9 m of water and exposed to 5 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 auditory evoked potentials (AEPs). Threshold shifts were found for exposed fish as compared to controls in the northern pike and lake chub, with recovery within 24 h 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 a 2D airgun array used in a river seismic survey. Care must be taken in extrapolation to other species in other environments and to fishes exposed to airguns in 3D tests or where the animals are exposed to a larger number of airgun shots over a longer period of time.

3:00–3:15 Break

3:15

2pAB5. Effects on fish of pile driving, wind turbines, and other sources. Anthony Hawkins (Loughine Ltd., Kincraig, Blairs, Aberdeen AB12 5YT, UK)

The sounds associated with construction work in the sea are reviewed. The installation of offshore oil and gas platforms, the mounting of offshore wind turbines on the seabed, the construction and renovation of harbors and piers, and the laying and trenching of pipelines and cables can be exceptionally noisy activities. The sounds associated with pile driving, in particular, have the potential to injure fish, induce hearing loss, and evoke avoidance reactions. The sound-pressure levels generated in water by these activities are generally measured, but many fish respond to particle motion rather than sound pressure. Measurements of particle motion and intensity will be necessary in the future if the full effects of these sounds are to be assessed.

3:40

2pAB6. Effects of background sound on fish. Michael E. Smith (Dept. of Biol., Western Kentucky Univ., Bowling Green, KY 42101, michael.smith1@wku.edu), Lidia E. Wysocki, and Arthur N. Popper (Univ. of Maryland, College Park, MD 20742)

There is increasing awareness that underwater anthropogenic sounds may be detrimental to marine organisms. While most of this work has focused on marine mammals, recent studies have shown that sounds higher than background can have detrimental effects on fish as well. This presentation is a review of the current data available on the effects of increased background ambient sound on fish. Loud continuous sounds can cause temporary hearing loss in fish, with the extent of hearing loss increasing with noise exposure duration and intensity, and recovery of hearing occurring within a few days or weeks. The extent of hearing loss is also species

specific—fish with more sensitive hearing thresholds are more prone to hearing loss, whereas nonsensitive fish are barely affected. The latter was shown in a long-term study on the effects of aquaculture noise in rainbow trout. A model termed the linear threshold shift hypothesis (LINTS) has been proposed to predict the extent of noise-induced hearing loss in fish, and this model should be beneficial in mitigating the effects of anthropogenic sound on fish. Other effects of noise include acoustical masking and physiological and behavioral stress responses, which also depend on species and noise type.

4:05

2pAB7. Noise exposure criteria for fish and turtles; lessons learned from criteria for marine mammals. Roger L. Gentry (ProScience Consulting, P.O. Box 177, Dickerson, MD 20842-0177, Roger.gentry@comcast.net)

A committee that is just finishing noise exposure criteria for marine mammals has learned the following lessons that may apply to the writing of similar criteria for fish and turtles. Writing criteria will require scientists from highly diverse backgrounds meeting multiple times over several years. Scientific discourse should prevail throughout; special interests will hinder the writing of science-based criteria. Because empirical data will undoubtedly be inadequate to the task, extrapolation procedures will be needed. Especially lacking will be animal data reporting metrics like SEL and particle motion. The wide diversity of sound types and of animal species will necessitate some grouping within each category. The most efficient format for the criteria will be a matrix of the two groupings. However, existing data will not support a matrix having too many cells. Some criteria will need to be expressed in more than one metric to account for all the relevant aspects of complex signatures. Frequency weighting functions will be needed. Criteria for injury will be easier to write than for behavioral responses due to data availability. Finally, several iterations of the criteria will be needed as science improves. The first version should be written to accommodate change.

4:30

2pAB8. Working group on "The effects of sound on fish and turtles": An update. Richard R. Fay (Parmly Hearing Inst., Loyola Univ. Chicago, IL 60626) and Arthur N. Popper (Univ. of Maryland, College Park, MD)

In order to develop acoustic exposure criteria for fish and marine turtles based on scientific evidence, an ASA Standards Working Group was formed with a membership that includes basic scientists expert on fish and turtle bioacoustics, physical acousticians with special expertise in marine bioacoustics, and several individuals with expertise on effects of anthropogenic sound in the marine environment. This presentation is a summary of the group's progress to date. One of the Working Group's goals is to publish several papers, one on what is known about damage risks due to exposure to underwater noise (e.g., from pile driving, seismic exploration, sonar systems), and one assessing the efficacy of various measures to mitigate adverse effects of noise exposure. The working group has tentatively decided that the fish potentially affected should be categorized with respect to their hearing sensitivity (potential for auditory system damage) and whether the species is known to detect sound pressure or acoustic particle motion. Over the next year to 18 months, the Working Group will be evaluating the known effects of sound sources on fish and turtles, identifying areas in which further research is needed, and beginning to develop noise exposure criteria for both turtles and fish.

4:55–5:25
Panel Discussion

TUESDAY AFTERNOON, 6 JUNE 2006 BALLROOM D, 1:30 TO 4:35 P.M.

Session 2pBB

Biomedical Ultrasound/Bioresponse to Vibration, Physical Acoustics, and Signal Processing in Acoustics: Sensing and Imaging Using Light and Sound

Ronald A. Roy, Cochair
Boston Univ., Dept. of Aerospace and Mechanical Engineering, 110 Cummington St., Boston MA 02215

Todd W. Murray, Cochair
Boston Univ., Dept of Aerospace and Mechanical Engineering, 110 Cummington St., Boston, MA 02215

Invited Papers

1:30

2pBB1. Nanoacoustics: Towards imaging nanostructures using picosecond ultrasonics. T. B. Norris (Ctr. for Ultrafast Optical Sci., Univ. of Michigan, Flint, MI 48502, tnorris@eecs.umich.edu)

Coherent acoustic phonon pulses with a peak frequency and bandwidth of the order of 100 GHz can be generated by ultrafast optical excitation of a thin metal film on a crystalline substrate. We demonstrate submicron resolution imaging of surface-grown nanostructures using these picosecond ultrasonic pulses. Ultrafast optical pulses from a Ti:sapphire oscillator are used to thermoelas-

tically excite a patterned 15-nm-thick Al film on a Si substrate. The spatio-temporal diffracted acoustic strain field is measured by time-delayed optical probe pulses on the opposite side of the substrate, and this strain field is used in a time-reversal algorithm to reconstruct the Al pattern. The image resolution is characterized using lithographically defined 1-micron-period Al structures on Si. Straightforward technical improvements should lead to a resolution of at least 45 nm, which will significantly extend the resolution of acoustic microscopy and provide a novel contrast mechanism for nanoscale imaging. To assist the development of this imaging method, we have also made experimental and theoretical studies of the behavior of these coherent acoustic phonon pulses after propagation through millimeter-scale distances in crystalline Si.

1:55

2pBB2. Applying optics in surface wave laser ultrasonic microscopy. Michael Somekh, Stephen Sharples, and Matthew Clark (Univ. of Nottingham, NG7 2RD, UK, mike.somekh@nottingham.ac.uk)

Laser ultrasonics has many advantages compared to contacting ultrasonics, especially when both optical generation and detection are used. In addition to the obvious advantages of remote inspection there are several other benefits. In this paper we concentrate on the flexibility that can be conferred by employing both optical technology and optical concepts in our laser ultrasonic systems. The papers shows how spatial light modulators (SLMs) can be used to control the optical beam profile used to generate the ultrasonic waves. Initially, SLMs were introduced to improve the signal-to-noise ratio of the system. This is achieved by exciting the tone burst with a grating structure so that large signals can be generated without ablation. The SLMs can also be used to produce optical profiles that generate self-focusing surface waves; this also increases the ultrasonic amplitude. Our recent work has used SLM generation to overcome aberrations or distortions in the ultrasonic field during propagation. To this end an adaptive system using integrated electronics to overcome the phase aberrations during propagation has been developed. We also show how the SLM can be used to perform acoustic wavelength spectroscopy, which has proved invaluable in the characterization of texture in industrially important materials.

2:20

2pBB3. Laser generation and detection of zero-group velocity Lamb mode resonance in thin plates. Claire Prada, Dominique Clorennec, Daniel Royer (LOA, ESPCI, Universit Paris 7, CNRS UMR 7587, 10 rue Vauquelin, 75231 Paris, Cedex 05, France), Oluwaseyi Balogun, and Todd W. Murray (Boston Univ., Boston, MA 02215)

Some Lamb modes exhibit a resonant behavior at frequencies where the group velocity vanishes while the phase velocity remains finite. Such a zero group velocity point exists in most isotropic materials for the first-order symmetric mode. Laser sources couple efficiently into this resonance and a sharp peak is observed with source and receiver on epicenter. Moving the detection point away from the source allows us to establish the dispersion curves of the plate and observe the negative phase velocity. We used a modulated diode laser source and a Michelson interferometer to observe the resonance at high frequency. Measurements were made on different material of thickness down to 4μm. The resonance peak is sensitive to the thickness and mechanical properties of plates and may be suitable for the measurement and mapping of nanoscale thickness variations. We also observed this behavior with a 20-ns pulsed YAG laser source and a heterodyne interferometer. The low attenuation in duralumin allows the excitation of a very sharp resonance in a single shot. Thickness variations as small as 0.1μm have been measured in a 0.50-mm-thick duralumin plate. Furthermore, the time decay of the signal provides an estimation of the local attenuation.

2:45

2pBB4. Fabry Perot polymer film sensors for ultrasound field characterization and biomedical photoacoustic imaging. P. C. Beard (Dept. of Medical Phys. and Bioengineer., Univ. College London, Gower St., London WC1E 6BT, UK)

A contact ultrasound sensing technique based upon the detection of acoustically induced changes in the optical thickness of a Fabry Perot polymer film sensing interferometer has been developed for ultrasound measurement and imaging applications in medical and industrial fields. The technique provides a broadband ($>$50 MHz) response and excellent detection sensitivities ($<$1 kPa), comparable to those of piezoelectric PVDF transducers. A key feature is that the sensing geometry is defined by the area of the polymer sensing film that is optically addressed. As a consequence, very small element sizes (in principle down to the optical diffraction limit of a few μm) can be obtained without compromising detection sensitivity—a useful advantage over piezoelectric transducers. Additionally, by spatially sampling over a relatively large aperture, a high-density ultrasound receive array can readily be configured. These attributes lend the technique to a variety of applications such as mapping the output of broadband ultrasound sources, transmission ultrasound imaging, and biomedical photoacoustic imaging. The latter is particularly promising, offering the prospect of high-resolution imaging of optically absorbing soft tissue structures such as blood vessels for the study of tumors and other tissue abnormalities characterized by changes in the structure and oxygen status of the microvasculature.

3:10–3:25 Break

3:25

2pBB5. Playing with speckle in acousto-optic imaging. Albert-Claude Boccara, Benoit Forget, Francois Ramaz, Emmanuel Bossy (CNRS and Laboratoire Optique, ESPCI, 10 rue Vauquelin, F75005 Paris, France), Michel Gross (ENS Laboratoire, F75005 Paris, France), Pedro Santos, Florence Jean, and Michael Atlan (CNRS and Laboratoire Optique, F75005 Paris, France)

Using light to image in-depth structures and to quantify them in terms of optical (absorption, scattering) properties is a goal for physicists. We will recall the approaches which overcome the strong loss of the photons directional memory in scattering media and their limitations in terms of depth and resolution. Then we will consider hybrid methods which couple acoustics and optics to reach the optical contrast with the acoustic resolution: opto-acoustics, where a locally absorbed laser pulse generates an acoustic signal whose position and depth can be localized by an array of piezoelectric detectors, and acousto-optics (A.O.), where an ultrasonic beam

modulates the speckle. We will focus on the principle of A.O. signal generation for tagging spatially selected photons and describe various approaches based on numerical heterodyne holography or heterodyne holography using a photo-refractive beam splitter. To get enough signals through thick tissues (e.g., breast) in a short time we use these techniques at the shot noise limit. Finally we show images revealing optical contrasts in artificial structures or in tissues.

Contributed Papers

3:50

2pBB6. Characterization of nano-electro-mechanical systems (NEMS) using photoacoustic microscopy. A. SampathKumar, K. L. Ekinci, and T. W. Murray (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215, twmurray@bu.edu)

A photoacoustic microscopy system has been developed for the non-contact characterization of materials at the micro- and nanometer scale. The system uses an electroabsoption modulated laser source for material excitation and a path-stabilized interferometer to detect the resulting displacements. Using this system, we describe the optical excitation and detection of high-frequency nanoelectromechanical systems (NEMS) at ambient temperature and pressure. Using a tightly focused modulated laser source, we have actuated the out-of-plane flexural resonances of bi-layered doubly clamped nanomechanical beams. The optically detected displacement profiles in these beams are consistent with a model where the absorbed laser power results in a local temperature rise and a subsequent thermally induced bending moment. The excitation technique allows probing and actuation of NEMS with excellent spatial and temporal resolution. From a device perspective, the technique offers immense frequency tunability and may enable future NEMS that can be remotely accessed thus obviating the need for electronic coupling. [The authors gratefully acknowledge funding from National Science Foundation under Grants ECS-0304446 and ECS-0210752.]

4:05

2pBB7. Focusing by use of time reversal of photoacoustic waves. Emmanuel Bossy (Laboratoire d'Optique Physique, ESPCI-CNRS UPR5, 10 rue Vauquelin 75005 Paris, France), Gabriel Montaldo, Jean-Francois Aubry, Michael Tanter, Mathias Fink (ESPCI-CNRS Universite Paris 7, 75005 Paris, France), and Claude Boccara (ESPCI-CNRS UPR5, 75005 Paris, France)

Time reversal of ultrasonic waves allows focusing of ultrasound through complex media, such as highly aberrating media. On one hand, focusing with time reversal can be performed by backpropagating detected ultrasonic waves generated by a localized ultrasound source. This source classically consists of high acoustic contrasts echoing ultrasonic waves generated by an incident ultrasonic field, or directly by pointlike transducers inserted in the medium. On the other hand, the photoacoustic effect consists of the emission of ultrasound caused by the absorption of light by an optical absorber. In particular, this effect can be used to ultrasonically image a distribution of optical absorbers simply by recording photoacoustic signals on an imaging transducers array. In this work, we perform time reversal of ultrasound generated by a pointlike photoacoustic source in order to automatically focus ultrasound back to an optical contrast. A Q-switched pulsed Nd:YAG laser was used to illuminate tissue phantom with nanosecond laser pulses. Conventional ultrasonic transducers arrays (in the MHz frequency range) and a time-reversal electronics were used to detect, record, and time reverse photoacoustic signals. We show that this allows focusing of ultrasound through highly aberrating medium, and perform acoustical imaging with good resolution within the isoplanetic region surrounding the optical contrast.

4:20

2pBB8. Laser pulse impact characterization using ultrasonic transducers. Subash B. Jayaraman, Mark A. Bohenick, Vladimir Semak, Bernhard R. Tittmann (The Penn State Univ., 212 Earth & Eng. Sci. Bldg., University Park, PA 16802), Jill R. Welsch, and Nicholas C. Nicholas (The Penn State Univ., University Park, PA 16802)

Experiments were conducted with an excimer laser (wavelength 248 nm) to cause damage to gelatinous materials, which have optical properties similar to skin tissue. The laser pulses, when they impact the specimen, generate an elastic wave that propagates through the material. A 150-MHz broadband ultrasonic transducer was placed underneath the specimen to observe the material response and the pressure generated by the laser pulse. Different incident laser energies and laser beam spot sizes were utilized to obtain a range of input parameters, and the corresponding transducer responses were recorded. Fourier analysis of the signals was performed to identify the frequency response from the laser pulse. Additional tests were carried out to observe the effects due to a train of laser pulses. Earlier tests were performed with 200-kHz hydrophone-calibrated transducer but due to the shortness of the laser pulses ($\sim$20 ns), a higher frequency transducer was used to accurately characterize the laser impact. The initial results indicate that there is a dominant frequency response around the 25–30-MHz regime.

Session 2pEA

Engineering Acoustics and Physical Acoustics: Joe Blue Memorial Session II: Transduction, Linear and Nonlinear

Thomas G. Muir, Cochair
Univ. of Mississippi, National Ctr. for Physical Acoustics, Coliseum Drive, University, MS 38677

Joseph F. Zalesak, Cochair
2359 US Hwy. 70 SE, #325, Hickory, NC 28602-8300

Chair's Introduction—1:00

Invited Papers

1:05

2pEA1. Dr. Blue at the Underwater Sound Reference Division. Joseph F. Zalesak (2359 US Hwy. 70, SE #325, Hickory, NC 28602-8300)

At the Underwater Sound Reference Division Dr. Blue had many varied interests including acoustic metrology, acoustic materials research, enginering acoustics, underwater transducers, and bioacoustics. He performed the first thorough study on the reproducibility of calibrations of standard hydrophones. He supported the development of traveling-wave tubes and large pressure vessels for calibrating Navy sonar transducers and for the measurement of acoustic materials. The Acoustic Pressue Tank Facility is a 30 000-gal acoustic test vessel ideally suited for acoustic measurements at hydrostatic pressures up to 2700 psi and at frequencies as low as 1.0 kHz and can accommodate acoustic devices up to 6 ft in diameter. Dr. Blue supported research for improved measurement techniques for determining the properties of acoustic materials. Dr. Blue actively supported the development of new elastomeric materials used in the construction of underwater transducers. The Sonar Transducer Reliability Improvement Program was an extinsive program to properly specify a sonar transducer for Navy applications. He first suggested that coupler reciprocity might be used to perform ultra-precise calibrations of hydrophones. He also supported research in manatee vocalizations and dolphin echolocation clicks. Low-noise hydrophones with noise levels better than sea-state zero were developed for general use.

1:25

2pEA2. A perspective of Navy transducer calibration standards. Kirk Jenne (Office of Naval Res., Code 03D&I, 875 N. Randolph St., Arlington, VA 22203-1995)

Research and development efforts to establish underwater electroacoustic transducer standards for use in a variety of environmental conditions, performed concurrently with research on calibration methodologies, began nearly 60 years ago. This research and development effort successfully generated a series of standards now made available under the auspices of the U.S. Navy's Underwater Sound Reference Division in Newport, RI (a.k.a. USRD). Because of consistent and insightful investment of resources made over the years by Navy technical agents and advocates like Dr. Joseph Blue, scientists and engineers from government (U.S. and select foreign), academia, and industry can use Navy transducer calibration standards to perform accurate measurements of acoustic energy underwater. Definition and characteristics of a standard are considered. Performance characteristics of transducer standards currently used, along with some discussion of applications, are presented.

1:45

2pEA3. Development of 1-3 piezocomposites for U. S. Navy applications. Thomas R. Howarth and Kim C. Benjamin (Naval Sea Systems Command Div. Newport, 1176 Howell St., Newport, RI 02841, Howarthtr@npt.nuwc.navy.mil)

Efforts to develop 1-3 piezocomposite materials for U.S. Navy applications began in the early 1990s at the former Naval Research Laboratory's Underwater Sound Reference Detachment in Orlando, FL. These initial investigations concentrated on evaluating the materials properties and parameters from a number of different contractors and manufacturing techniques. This was followed by efforts to initially fabricate receiving devices to replace traditional piezoceramic sensor prototypes, followed by development of devices with both receiving and transmitting capabilities. The presentation will discuss the initial NRL-USRD efforts and follow the development of 1-3 piezocomposite materials from these humble beginnings to devices that are currently being used and further developed by the U.S. Navy. [Work supported by the NRL, ONR, DARPA, and NAVSEA.]

2:05

2pEA4. Use of complex source points to simplify numerical Gaussian beam synthesis. Stephen Forsythe (Naval Undersea Warfare Ctr., 1176 Howell Ave., Newport, RI 02841)

It is often desirable to generate the acoustic field due to a so-called Gaussian beam. One way to do this is to use the free-space Green's function for the acoustic field and to sum small area sources over a circular plate with the appropriate shading for the desired Gaussian beam. For very high frequencies and narrow beams, the computation time to give an accurate sum can be large when calculating the sum for many points in the acoustic field. An alternate approach comes from the use of a single point source with complex coordinates $R=(Xr+i^*Xi, Yr+i^*Yi, Zr+i^*Zi)$. When this complex source point is used in the free-space Green's function, the formal expressions for pressure and particle velocity can be used if careful attention is paid to the interpretation of the complex distance, r, that arises in the $\exp(ikr)/r$ term. The singularity is no longer a single point in the case of a complex source, but a circular disk. The far field of a complex source point is a good approximation to a Gaussian beam. Several computational uses of the technique will be demonstrated. Extension to the shear wave Green's function will be explored. [Work supported by ONR.]

2:25–2:45 Break

2:45–2:50 Introduction

2:50

2pEA5. Joe Blue and nonlinear acoustic transduction: Lessons learned and forgotten, some even ignored; and practical issues to revisit, anew. Thomas Muir (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, One Coliseum Dr., University, MS, 38677)

The 20th century was ripe with major advances in the understanding of nonlinear acoustics and its practical applications, in the American as well as the European and Soviet schools of thought. Some of us were privileged to study this art and contribute to its science, guided by the world's finest educators and mentors. The results originated in physics and were applied to aircraft acoustics and underwater sound, and later to biomedical acoustics, and other arenas, with great success. Joe Blue was a contributor and long-time advisor in many of these endeavors. Some early nonlinear acoustics topics will be reviewed, to refresh memories and introduce newcomers to the field. This includes parametric transmitters, the effects of shock formation and harmonic generation, parametric receivers, various theoretical models, classic experiments in air and water, as well as practical applications. In the early days, ideas for new nonlinear acoustic applications were often met with the comment, ''That's crazy.'' Perhaps many of these ideas were not crazy enough. At the present time, there appears to be some lack of boldness and innovation, coupled with abundant mistakes, in what is proposed and reported. Some comments on the state of the art are offered. [Work supported, throughout the years of the mid to late 20th century, by the U.S. Navy.]

3:10

2pEA6. Investigation of media with large, negative parameters of nonlinearity and their application to the enhancement of a compact, omnidirectional, parametric source. Peter H. Rogers, Etienne Dufour, and David H. Trivett (Woodruff School of Mech. Eng., Georgia Inst. of Technol., Atlanta, GA 30332)

This paper reports on two media whose acoustic properties are highly pressure dependent and, thus, have large nonlinear parameters ($B/A \sim -3000$ to -5000). One is Xanthan gelatin filled with, 10%, Expancel microspheres. The other is a polyurethane rubber, Polytek 74, with 20% Expancel microspheres. The microspheres have a high bulk modulus at ambient pressure due to the hoop stiffness of the spherical shell. As pressure is increased some of the microspheres buckle, leading to significantly decreased bulk modulus. The decreasing bulk modulus with increasing hydrostatic pressure (negative dc/dp) results in both negative B/A and low sound speed (150 m/s at the minimum), which further enhances nonlinear distortion. These media are used to enhance the efficiency of an underwater parametric source. Experiments on a spherical source with a thin layer (2 cm thick) of the Xanthan gelatin material resulted in a generated difference frequency source level 16 dB higher at 300 Hz, than would be obtained driving the transducer directly at the difference frequency. Experiments on the same source with a thin layer of the microvoided, polyurethane rubber are in progress. This material has a smaller coefficient of nonlinearity but operates over a broader range of hydrostatic pressures.

3:30

2pEA7. Nonlinear acoustic detection of buried landmines: A tribute to Joe Blue. Murray S. Korman (Phys. Dept., U. S. Naval Acad., Annapolis, MD 21402)

As a graduate student at Brown University ($\sim$1978), this researcher was very fortunate to have met Joe Blue at Acoustical Society meetings. As I had little knowledge of the members in the society and an extremely limited grasp of the practical technology behind nonlinear acoustics, Joe was extremely friendly and took the time to answer a long list of questions about nonlinear underwater transducer designs with applications to the parametric array [Muir, T.G. & Blue, J.E. ''Experiments on the acoustic modulation of large-amplitude waves,'' J. Acoust. Soc. Am. **46**, 227–232 (1969)]. This influenced this researcher's Ph.D. thesis involving experiments on the scattering of sound by sound in the presence of turbulence. Recently, experiments on the nonlinear acoustic detection of buried landmines [Korman, M.S. & Sabatier, J.M., ''Nonlinear acoustic techniques for landmine detection,'' J. Acoust. Soc. Am. **116**, 3354–3369 (2004)] has brought some attention to understanding the difference between classical and nonclassical nonlinearity. As a tribute to Joe Blue, the latest results will be presented involving the connections to mesoscopic nonlinear elasticity and slow dynamics in nonlinear buried landmine detection. Here, tuning curve behavior of the soil-surface vibration over a landmine exhibits a linear decrease in resonant frequency with increasing amplitude. Two-tone test excitation reveals a rich spectrum of combination frequency components useful in detection schemes.

3:50

2pEA8. The application of acoustic standards to sound sources and the subsequent levels in the ocean. William M. Carey (Aero and Mech. Eng., College of Eng., Boston Univ., Boston, MA 02215)

The American National Standards on Acoustics are applied to common sound sources used in both acoustics research and in naval sonar systems operation. Metrics are quantified for both continuous and transient sources of sound. The standard definitions are reviewed with theoretical sound source models and applied to examples of energy sources of sound such as transients from a small omni explosive, an air gun, a light bulb, and a dolphin. A qualitative model of a typical surface ship sonar system is discussed and active sonar transmissions are analyzed with the requisite quantitative metrics required to characterize these emissions. The relative role of peak pressure, time spread, intensity, and energy flux is discussed for deep- and shallow-water environments. These results should be useful in environmental assessments, biological experiments, and in the system design.

Contributed Papers

4:10

2pEA9. Calibrating parametric sonars for backscattering applications. Kenneth G. Foote (Woods Hole Oceanogr. Inst., 98 Water St., Woods Hole, MA 02543)

The status of parametric sonar calibration is reviewed. For backscattering applications, the standard-target method, which is widely used in calibrating ultrasonic scientific echo sounders and multibeam sonars that provide the water-column signal, can be adapted. Two problems generic to parametric sonars are addressed. (1) The virtual nature of the array may require controlled measurements at a number of ranges, spanning the near field in the general case. (2) Calibration at rather low difference frequencies requires a suitably powerful target. Both measurement protocols and target design are considered for a particular parametric sonar based on primary frequencies in the band 15–21 kHz, with difference frequencies in the band 0.5–6 kHz and sum frequencies in the band 30–42 kHz. Solid elastic spheres are specified for both frequency bands. For the lower band, this is an aluminum sphere of diameter 280 mm, with immersed weight of 190 N.

4:25

2pEA10. Evaluating the significance of interface nonlinearity for shallow-water parametric array sonar. Roger Waxler and Thomas G. Muir (NCPA, University, MS 38577)

Difference frequency generation in shallow water is considered for the case in which wave steepening in the primary beam is not significant. A general formula for the effective difference frequency source strength is presented. The Pekeris shallow water model, in which sound speeds and densities in both the water column and sediment are assumed to be constant and both interfaces are assumed to be planar, is considered. In this model, the propagating difference frequency pressure field is computed. Contributions from the interface nonlinearity are compared to those from nonlinearity in the water column.

4:40

2pEA11. High frequency acoustic wave propagation. Mohsen Badiey (Ocean Acoust. Lab., Univ. of Delaware, Newark, DE 19716)

A program studying high frequency acoustics was established in the mid-1990's by the Office of Naval Research that is still going on and has evolved into a substantial basic and applied research initiative to study many areas, including underwater communication. This abstract explains the genesis of that program, and Joe Blue's involvement in making it happen. During the course of the initial experiments under that program, a series of issues were raised dealing with the effects of acoustic transmissions on marine mammals. Again, Joe was a key person in resolving those issues, which could have brought the program to a halt. His insight into the technical problems, as well as, his calm nature helped to bring an excellent new program in acoustic research into being.

4:55

2pEA12. Ship strike acoustics: A paradox and parametric solution. Edmund Gerstein, Joseph Blue (Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486), and Steve Forsythe (Naval Undersea Warfare Ctr., Newport, RI 02841)

Marine mammals are vulnerable to ship collisions when are near the surface. Here, acoustical laws of reflection and propagation can limit their ability to hear and locate the noise from approaching vessels. Defining the physics of near-surface acoustical propagation as it relates to ship noise and hearing is central to understanding and mitigating ship strikes. Field data from controlled ship passages through vertical hydrophone arrays demonstrate a confluence of acoustical factors that poses detection challenges including (i) downward refraction; (ii) spreading loss; (iii) Lloyd's mirror effect; (iv) acoustical shadowing; and (v) masking of approaching ship noise by ambient noise and distant ships. A highly directional, dual-frequency parametric sonar has been developed to mitigate these challenges and alert marine mammals of approaching vessels. The system projector is a planar array, comprised of 45 elements, band centered to transmit a high carrier frequency along with a lower sideband signal. The nonlinearity of water is used to demodulate the mixed high-frequency carrier into a lower-frequency waveform audible to both manatees and whales. The bow-mounted arrays project a narrow beam directly ahead of vessels, and fill in acoustical shadows in an effort to alert marine mammals of the approaching danger. [Research supported by DOD Navy Legacy Program.]

Session 2pMU

Musical Acoustics and Structural Acoustics and Vibration: Finite Element and Finite Difference Methods in Musical Acoustics I

Rolf Bader, Cochair
Univ. of Hamburg, Inst. of Musicology, Neue Rabenstr. 13, 20354 Hamburg, Germany

Uwe J. Hansen, Cochair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Chair's Introduction—1:25

Invited Papers

1:30

2pMU1. Mode studies in a triangular bell plate. Uwe J. Hansen, James Garner (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809), Thomas D. Rossing (Northern Illinois Univ., DeKalb, IL 60115), and Suzanne Hogg (Univ. of Technol., Sydney, Australia)

A triangular bell plate serves as a poor man's hand bell. A clapper affixed to the handle strikes the plate, resulting in a sound somewhat resembling the tone of a hand bell. A similar arrangement has been used with slotted Eames tubes. This paper will present a comparison of modes of vibration of a triangular bell plate as observed with holographic interferometry, modal analysis, and as calculated using finite element methods. The holographic interferometry studies were done in the Acoustics laboratory of NIU, the modal analysis work was carried out at ISU, and the FE calculations were done at the UT in Sydney. The similarity of mode shapes and mode frequencies lends credence to the validity of the techniques.

2:00

2pMU2. Finite-element calculation of a bass drum. Rolf Bader (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstrasse 13, 20354 Hamburg, Germany)

The bass drum of a drum kit used with rock or jazz music is modeled as a 3D geometry in a transient finite-element formulation. The geometry was taken from a Gretch drum set called ''Vinnie Colaiuta'' with no cushion damping within the drum or at the resonance membrane. The 12 cross layers or wood used for the drum body was assumed to result in an isotropical Young's modulus all over the wood. The drum was modeled with air enclosed, with the beating and resonating membranes being coupled to the air and to the wooden rim of the drum body. Also, as a starting point of the transient calculation, the strain of the wooden body caused by the tension of the membranes was taken into consideration. The transient model started with a beating on the front membrane and the radiation of the system was integrated over the radiating area. The results are compared to recordings of a real bass drum of this Gretch drum set.

2:30

2pMU3. Finite-element transient calculation of a bell struck by a clapper. Bjrn-Andre Lau, Peter Wriggers (Institut fr Baumechanik und Numerische Mechanik, Univ. of Hannover, Appelstrae 9a, 30167 Hannover, Germany), Albrecht Schneider, and Rolf Bader (Univ. of Hamburg, 20354 Hamburg, Germany)

A finite-element calculation of a bell struck by the bell clapper was performed as a transient contact problem. The geometry of the bell was taken from a real bell which was carved especially in behalf of this research. The eigenmodes and eigenfrequencies of the bell were calculated in advance, being in good agreement with the measured frequencies. The clapper/bell contact time and impulse shape show the expected behavior. The impulse shape is a Gaussian impulse with a tendency of a sharper attack slope and a softer decay slope. The periodical shape of the spectrum of this impulse caused by its finite width with zero values for some frequencies is shown. Nevertheless, this spectrum also drives the frequencies of the bell within the low-amplitude region of the driving impulse with expected and measured strength. The reason for this mode energy transfer is discussed. Furthermore, the sound of the bell is calculated as radiation from its surface and analyzed with signal-processing tools like wavelet transforms or autoregressive models.

3:00–3:15 Break

3:15

2pMU4. Finite-difference modeling of the piano. N. Giordano (Dept. of Phys., Purdue Univ., 525 Northwestern Ave, West Lafayette, IN 47907, giordano@purdue.edu)

The construction of a complete computational model of the piano is described. This model treats the three major subsystems of the instrument: the hammers and strings, the soundboard, and the air (the sound radiation). Hammer nonlinearities and hysteresis are included, along with realistic stiff strings with frequency-dependent damping, following the work of Chaigne and Askenfelt. [Chaigne and A. Askenfelt, ''Numerical simulations of piano strings. I. A physical model for a struck string using finite difference methods,'' J. Acoust. Soc. Am. **95**, 1112–118 (1994); A. Chaigne, J. Acoust. **5**, 181 (1992)]. The strings are coupled to a soundboard model that includes the added thickness and stiffness of ribs and bridges. The vibrating soundboard drives the room air, with walls modeled through a complex acoustic impedance following Botteldooren. [D. Botteldooren, ''Acoustical finite-difference time-domain simulation in a quasi-Cartesian grid,'' J. Acoust. Soc. Am. **95** 2313–2319 (1994); **98**, 3302 (1995)]. All of the calculations take a finite difference approach, producing a time domain computation. Calculated tones are presented and future directions discussed. [Work supported in part by NSF grant PHY-9988562.]

3:45

2pMU5. Finite-element simulation of the acoustic guitar. Gregoire P. Derveaux, Eliane Becache, Patrick Joly (INRIA, domaine de Voluceau, BP105, 78153 Le Chesnay Cedex, France, gregoire.derveaux@inria.fr), and Antoine Chaigne (ENSTA, 91761 Palaiseau Cedex, France)

The purpose of this work is a 3-D time-domain numerical modeling of the acoustic guitar. The model involves the transverse displacement of the string excited by a force pulse, the flexural motion of the soundboard, and the sound radiation. A specific spectral method is used for solving the Kirchhoff-Love's dynamic plate model for a damped, heterogeneous orthotropic material. The airplate interaction is solved with a fictitious domain method, which allows for making a good approximation of the geometry of the instrument while using a regular mesh for the calculation of the 3-D sound radiation. A conservative scheme is used for the time discretization. Frequency analysis is performed on the simulated sound pressure and plate velocity waveforms in order to evaluate quantitatively the transfer of energy through the components of the coupled system. The effects of structural changes in soundboard thickness and cavity volume on the produced sounds are presented. Simulations of the same guitar in three different cases are also performed: ''*in vacuo*,'' in air with a perfectly rigid top plate, and in air with an lastic top plate. This allows comparisons between structural, acoustic, and structural acoustic modes of the instrument.

TUESDAY AFTERNOON, 6 JUNE 2006 ROOM 556AB, 1:00 TO 3:30 P.M.

Session 2pNSa

Noise, ASA Committee on Standards, and Psychological and Physiological Acoustics: New Loudness Standard

Rhona P. Hellman, Chair

Northeastern Univ., Speech, Language, Pathology and Audiology, 360 Huntington Ave., Boston, MA 02115-0000

Chair's Introduction—1:00

Invited Papers

1:05

2pNSa1. Rationale for a new loudness standard. Rhona P. Hellman (Dept. of Speech-Lang. Path. and Audiol. and Inst. for Hearing, Speech, & Lang. (106A FR), Northeastern Univ., Boston, MA 02115, hellman@neu.edu)

In 2005, a new loudness standard was approved by ANSI for calculating the loudness of steady sounds. The new standard is based on a recent loudness model [Moore *et al.*, J. Audio Eng. Soc. **45**, 224–240 (1997)] derived from an earlier model devised by Zwicker [Acustica **8**, 237–258 (1958)]. It replaces and supersedes the more limited computational procedure used in the old ANSI S3.4-1980 standard that had its origins in the protocol developed by S.S. Stevens [J. Acoust. Soc. Am. **33**, 1577–1585 (1961)]. Unlike the old ANSI S3.4-1980 procedure, the newly adopted ANSI standard can be applied to sounds with sharp line spectral components as well as to broadband spectra. The new standard also extends the dynamic range for loudness calculations to loudness levels below 40 phons. Hence, it can be used to estimate loudness, or loudness level, with reasonable accuracy at levels close to threshold. Moreover, in accord with the revision of ISO 226 [International Organization for Standardization, Geneva (2003)], it enables the loudness of complex sounds containing spectral energy below 500 Hz to be determined. Empirical examples that illustrate the broader and more accurate scope of the new loudness standard are presented.

1:30

2pNSa2. The Cambridge loudness model: Design principles and predictions. Brian C. J. Moore, Brian R. Glasberg, and Tom Baer (Dept. of Exp. Psychol., Univ. of Cambridge, Downing St., Cambridge CB2 3EB, UK, bcjm@cam.ac.uk)

The Cambridge loudness model, which forms the basis for the new ANSI standard (ANSI S3.4-2005), is similar in structure to loudness models proposed earlier by Fletcher and by Zwicker, but differs in many of the details. The input signal to each ear is specified in terms of its spectrum, which may be defined in 1/3-octave bands, or as the sum of a series of sinusoidal components and one or more pink-or white-noise bands. The first stage of the model is a filter to account for the transfer of sound from the sound field or headphone to the eardrum. The second stage is a filter to account for the transfer of sound through the middle ear. The third stage transforms the effective spectrum reaching the cochlea to an excitation pattern, calculated as the output of an array of level-dependent auditory filters. In the final stage, the excitation at each center frequency is converted to specific loudness, by a nonlinear frequency-dependent transformation. The specific loudness is summed across frequency to give the loudness for each ear, and is summed across ears to give overall loudness. The model accounts well for the effects of level, frequency, and bandwidth on loudness.

1:55

2pNSa3. Reexamining loudness at and near threshold. Michael Epstein (Auditory Modeling and Processing Lab., Inst. for Hearing, Speech & Lang., Dept. of Speech-Lang. Pathol. and Audiol., Northeastern Univ. Boston, MA 02115) and Mary Florentine (Northeastern Univ. Boston, MA 02115)

This talk reviews two sets of data that support the contention that loudness at threshold is not zero, consistent with the new loudness standard. Both sets of data examine the form of the loudness function in normal listeners at low levels. The first set of data was derived from loudness-matching experiments between equally loud tones and tone complexes (i.e., spectral integration of loudness). Listeners matched the loudness of tone complexes comprised of subthreshold components with pure tones above threshold. Results show that a tone complex composed of components at threshold can easily be heard and can be relatively loud. This result is only possible if all tones had some probability of detection (some contribution to loudness), even those below threshold. The second set of data consists of the slopes of the loudness function at threshold, which average 1.3. These data are consistent with loudness functions obtained with a wide variety of methods that yield a slope close to unity. A slope close to unity indicates a positive loudness at threshold in accord with the new loudness standard. [Work supported by NIH/NIDCD Grant R01DC02241.]

2:20

2pNSa4. The new International Organization for Standardization (ISO) standard for the equal-loudness contours: Comparison to earlier contours and procedures. Yôiti Suzuki (Res. Inst. of Elec. Commun., Tohoku Univ., Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan, yoh@riec.tohoku.ac.jp) and Hisashi Takeshima (Sendai Natl. College of Technol., Sendai 989-3128, Japan)

Equal-loudness contours exhibit the intensity-frequency characteristic of the auditory system. Since 1956 such contours based on the free-field measurements of Robinson and Dadson [Br. J. Appl. Phys. **7**, 166–181 (1956)] have been widely accepted as an international standard in ISO 226. However, in 1985 some questions about the general applicability of these standard contours were raised to ISO TC/43 WG1 by Fastl and Zwicker. Consequently, a new international effort to measure the equal-loudness contours was undertaken. In 2003, this effort produced a revision of ISO 226. The newer experimental data and their analyses were subsequently published by Suzuki and Takeshima, [J. Acoust. Soc. Am. **116**, 918–933 (2004)]. The revised contours more closely agree with the classic contours measured by Fletcher and Munson [J. Acoust. Soc. Am. **5**, 82–108 (1933)] than with those produced by Robinson and Dadson. This finding is especially true below 500 Hz at moderate loudness levels. Moreover, the revised contours are in good overall agreement with the contours calculated according to the procedure proposed by Moore *et al.* [J. Audio Eng. Soc. **45**, 224–240 (1997)]. In 2005, this procedure was adopted by ANSI for loudness calculations of steady sounds to replace ANSI S3.4-1980. Comparative examples are given.

2:45

2pNSa5. Future directions in loudness standardization: Time-varying loudness. Patricia Davies and Andrew Marshall (Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031, daviesp@purdue.edu)

The loudness of many sounds varies with time and the temporal processing of the human hearing system affects loudness perception. Existing loudness algorithms that are standardized address loudness of stationary sounds. Two algorithms that predict time-varying loudness have been developed: Zwicker's, and Moore and Glasburg's time-varying loudness; variants of the former are available in most sound-quality analysis systems. For continuous sounds with repetitive impulses or beating tones loudness exceeded 5% of the time correlates well with perceptions of overall loudness. The focus here is on predictions of time-varying loudness algorithms for isolated impulsive sounds caused by supersonic aircraft, because recent developments in supersonic aircraft design could result in aircraft producing much quieter booms that may be more acceptable to the general public. Loudness is a major factor in community response; thus, the research focus is to understand how sound modifications (amplitude reduction, signal shaping) affect the loudness of transients. Also, playback systems involve signal filtering; how this may affect loudness perception is of interest. The predictions of the two algorithms are compared for a series of sonic-boom recordings and modified recordings to determine how filtering and signal modification affect loudness predictions.

3:10–3:30

Panel Discussion

Session 2pNSb

Noise and Architectural Acoustics: Audio-Visual Design in Soundscapes II

Brigitte Schulte-Fortkamp, Cochair
Technical Univ. Berlin, Inst. of Technical Acoustics, Einsteinufer 25, 10587 Berlin, Germany

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

Continuation of session 2aNSa. Field measurements will be discussed and evaluated.

Session 2pPA

Physical Acoustics: Nonlinear Acoustics, Flow, and Miscellaneous Topics

Vera Khokhlova, Chair
Moscow State Univ., Acoustics Dept., Physics Faculty, Leninskie Gory, 119992 Moscow, Russia

Contributed Papers

1:30

2pPA1. Assessment of incipient material damage and remaining life prediction with nonlinear acoustics. Dimitri Donskoy (Intelligent Sensing Technologies, LLC, P.O. Box 6457, Fair Haven, NJ 07704-6457), Andrei Zagrai (Stevens Inst. of Technol., Hoboken, NJ 07030), Alexander Chudnovsky, Edward Golovin (Univ. of Illinois at Chicago, Chicago, IL 60607), and Vinod S. Agarwala (Naval Air Systems Command, Patuxent River, MD 20670)

Subjected to a wide spectrum of operational and environmental loads, engineering structures exhibit first signs of material deterioration at micro and meso scales. In contrast to nondestructive assessment of the macro-scale damage that resulted from material fracture, a limited number of technologies have tackled the complex task of detecting and monitoring the small-scale incipient damage. This paper discuses application of the nonlinear acoustic vibro-modulation technique for assessment of the micro/meso scale fatigue damage and remaining life prediction of metallic samples. A nonlinear acoustic damage index (DI), which indicates strength of the damage-induced nonlinear interaction of the high-frequency ultrasonic wave and the low-frequency structural vibration, was monitored during a set of strain-controlled fatigue tests. Stable growth of the damage index was observed at an increasing number of fatigue cycles showing correlation between DI and the micro/meso scale damage accumulation. This correlation allowed for introducing DI in the remaining life prediction methodology based on an entropy density criterion of local failure. Experimental testing confirmed the effectiveness of the proposed prediction approach. [Work supported by NAVAIR.]

1:45

2pPA2. Wave vector resonance in a nonlinear diffuse elastodynamic field. Alexei Akolzin and Richard L. Weaver (Dept. of Theoretical and Appl. Mech., Univ. of Illinois, 104 S. Wright St., Urbana, IL 61801)

Nonlinear coupling within an elastic body leads to spectral energy redistribution. The coupling strength can exhibit sharp discontinuities as a function of frequencies, due to wave vector coincidences between constituent waves of different wave speeds. The phenomenon is investigated for an ensemble of weakly nonlinear elastic bodies of a generic ray-chaotic shape, with the elastodynamic field inside being treated as diffuse and described in a statistical fashion. Comparison is made between theoretical estimates of the average mode coupling strength and results of direct numerical simulations performed for an ensemble of 2D plane-strain elastic squares with rough boundaries and a model quadratic nonlinearity representative of elastodynamic waves. Existence of the wave vector resonance and its dependence upon frequency ratios is verified.

2:00

2pPA3. Dynamic characteristics of nonlinear interaction in elastic medium. Nikolai Zagrai (Taganrog State Univ. of Radio-Eng. GSP-17A, 44 Nekrasovsky St., Taganrog, 347928, Russia)

Traditionally, an acoustic wave process is analyzed in terms of sound pressure and/or sound velocity. The dynamic distortion of the wave profile, however, is complicated by contribution of both quadratic and cubic nonlinearities of the elastic medium. This condition necessitates expanding a range of variables to include particle and wave accelerations. As dynamic characteristics, particle and wave accelerations fully complement

analysis of dynamic and inert properties of the elastic medium. The Riemanns solution, in which the quadratic and cubic nonlinearities are accounted for through respective derivatives of the sound velocity, allows for a new and different perspective on the aforementioned properties. For a simple wave with contribution of the quadratic nonlinearity only, the acceleration profile describing dynamics of wave deformation varies symmetrically relative to the discontinuity region. The cubic nonlinearity leads to more complex acceleration profile with asymmetric properties, which can be discerned and assessed using a combination of numerical and graphical methods. Such an approach is quite effective and vivid in separating contributions of quadratic and cubic nonlinearities. An integral nonlinear parameter N, introduced on the basis of this analysis, is proposed for estimating ability of the fluidlike medium to support effective propagation and nonlinear interaction of acoustic waves.

2:15

2pPA4. A frequency-domain algorithm for the propagation of finite-amplitude acoustic signals. Lauren E. Falco (Grad. Prog. in Acoust., Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802, lfalco@psu.edu), Philip J. Morris, and Anthony A. Atchley (Penn State Univ., University Park, PA 16802)

It has been shown that the quantity Q_{p^2p}, the cross-spectral density of the square of the pressure and the pressure, is related to the strength of nonlinear effects in a propagating acoustic signal. Morfey and Howell [AIAA J. **19**, 986–992, (1981)] derive from the Burgers equation an expression that relates Q_{p^2p} to the rate of change of the power spectral density (PSD) of a signal. They suggest a propagation scheme based on this expression that makes assumptions about the Gaussian nature of the signal. These assumptions are invalid for many finite-amplitude signals, so a new propagation scheme is proposed and validated. This scheme uses an expression for the rate of change of the phase of the signal as well as its amplitude. These two expressions are combined into a finite-difference equation that is used to step the signal to the desired propagation distance. The method is shown to compare favorably with the Fubini solution for lossless plane waves. Comparisons are also made with a numerical split-step algorithm for the nonlinear propagation of a broadband source, and with experimental data obtained in a plane wave tube. [Work supported by ONR, GE.]

2:30

2pPA5. Thermal gradient effects on the propagation of nonlinear pressure waves in a gas. Nicholas DiZinno, George Vradis, and Volkan Otugen (Mech. & Aerosp. Eng., Polytechnic Univ., 6 Metrotech Ctr., Brooklyn, NY 11201, dizinno@yahoo.com)

A numerical study of wave propagation through gases with nonuniform temperature distributions will be presented. The main objective of this study is to determine the impact of temperature gradients on high-intensity, initially sinusoidal pressure waves. Particular emphasis is paid to wave reflection, transmission, and any influence a high-temperature region may have on nonlinear behavior. Ultimately, the performance of thermal barriers in attenuating nonlinear waves is evaluated. The concept of using regions of hot gas inside an ambient environment has potential in aeroacoustic applications, such as jet screech mitigation. This analysis considers one-dimensional compressible unsteady Euler equations with an ideal gas state equation, applied over a uniform domain. The domain is composed of two regions with uniform and equal gas properties (ambient conditions) separated by a third region with higher gas temperature and, accordingly, lower density. Pressure is uniform throughout the domain. We introduce a high-intensity sinusoidal wave into this medium. The shape and extent of the high-temperature zone is varied to study the effect of this region on wave propagation. Further, wave reflection and transmission are studied for a range of wave and thermal field parameters. Results for nonlinear pressure waves are compared to linear acoustic waves.

2:45

2pPA6. Evanescent time reversal: Super-resolution and communications applications. Stephane G. Conti, Philippe Roux, and William A. Kuperman (SIO/UCSD, Marine Physical Lab., 9500 Gilman Dr., La Jolla, CA 92093-0238)

The spatial resolution of the focused field with a classical time-reversal mirror in the far field has a wavelength-order diffraction limit. Previously reported results for super-resolution beyond the diffraction limit require either the full knowledge of the original source, or the evanescent waves in the near field. Here, we show that subwavelength focusing can be achieved without *a priori* knowledge of the original probe source. In the wave-number domain, the contributions for the spatial localization and for the resolution of the focus can be separated from each other in the time-reversed field recorded within one wavelength of the original source. The evanescent time-reversal field is then obtained from the combination of the spatial localization contribution from the far-field time reversed with the spatial resolution for a point source. Further, we show that different signals can be transmitted simultaneously with subwavelength separation. Super-resolution of the order of 1/20 of the wavelength is experimentally demonstrated with acoustic wave fields. Applications for communications are presented using multiple receivers a fraction of the wavelength apart.

3:00–3:15 Break

3:15

2pPA7. Area oscillations of a bias-flow aperture, with application to phonation. Richard S. McGowan (CReSS LLC, 1 Seaborn Pl., Lexington, MA 02420, rsmcgowan@earthlink.net) and Michael S. Howe (Boston Univ., Boston, MA 02215)

A thin baffle plate in a mean flow duct partitions the tube into high- and low-pressure regions. The plate spans the whole cross section of the duct and is pierced by a small aperture, through which a nominally steady bias flow is maintained by the mean pressure differential. Sound is generated when the flow is disturbed by imposed, small amplitude cross-sectional area fluctuations of the aperture. For irrotational flow the radiation is determined by the input impedances of the tube on each side of the aperture, and energy flowing into the acoustic field is equal to the work done on the fluid by the suction force at the edge of the aperture. For high Reynolds number flow of a real fluid (modeled by the imposition of a Kutta condition) much of the work goes into the production of vorticity, the unsteady volume flow is greatly reduced, and at low frequencies there can be a substantial oscillating drag force on the plate [M. S. Howe, Proc. R. Soc. London, Ser. A **366**, 205–233 (1979)]. The application of these results to phonation indicates that the unsteady drag is most important in the lower harmonics of voice. [Work supported by NIDCD-004688 to UCLA.]

3:30

2pPA8. Simulation of fluid flow in and around a porous windscreen. D. Keith Wilson (U.S. Army Engineer Res. and Development Ctr., 72 Lyme Rd., Hanover, NH 03768, d.keith.wilson@erdc.usace.army.mil) and Michael J. White (U.S. Army Engineer Res. and Development Ctr., Champaign, IL 61826)

Fluid dynamic simulation of wind noise (turbulent pressure fluctuations) around porous microphone windscreens in an atmospheric flow is made difficult by the broad range of spatial scales, from large atmospheric eddies down to the pore size of the windscreen, that is involved. Here, we describe a simplified simulation approach that is based on two new insights: (1) the ability of linear fluid flow approximations to describe wind noise [Z.C. Zheng and B. K. Tan, J. Acoust. Soc. Am. **113**, 161–166 (2003)] and (2) a new set of macroscopic, time-domain relaxational equations for the linear interaction of a fluid flow with a porous material [D. K. Wilson, V. E. Ostashev, and S. L. Collier, J. Acoust. Soc. Am. **116**, 1889–1892 (2004)]. An upstream flow condition consisting of quasiwavelets

(analogous to turbulent eddies) is also used to create well-developed, realistic turbulence. This simulation approach is shown to provide detailed simulations of the production of pressure fluctuations in and around a porous windscreen.

3:45

2pPA9. Diffraction pattern of an axisymmetric aperture from axial measurements. Edward H. Pees (Dept. of Phys., Univ. of Rhode Island, 2 Lippitt Rd., Kingston, RI 02881-0817)

An axisymmetric aperture can be represented by a circular hole in an infinite rigid screen, a circular piston undergoing harmonic motion, or a more complicated source composed of annular ring transducers. Ordinarily, the radiation pattern for this type of aperture is measured by moving a sensor along a circular arc about the center of the aperture, measuring the intensity of the radiated field as a function of angle. In the present case, the angular spectrum of plane wave representation of the diffracted field is used to demonstrate that the diffraction effect leads to the presence of a wave number distribution, observable along the normal axis of the aperture, that can be used to reconstruct the far-field radiation pattern from measurements taken only along this angular view. Because this diffraction pattern has a direct relationship to the boundary condition on the aperture, it follows that measurements of the axial field can potentially provide useful information regarding the size and shape of the aperture.

4:00

2pPA10. Transient total reflection and transient evanescent wave. Hailan Zhang (Inst. of Acoust., Chinese Acad. of Sci., Beijing 100080 China)

The total reflection of a steady plan wave is well known. The transmitted wave is an evanescent wave that propagates along one direction and evanesces along a direction perpendicular to the propagation direction. Evanescent waves also appear in many other acoustic problems, such as the surface wave along a free boundary of a solid medium. In this report, the total reflections with different incident waves are studied, including an incident beam with a limited width, an incident pulse wave with a limited duration, and an incident bulletlike wave package. These incident waves are represented by superposition of plane waves based on the angle spectrum synthesis. Their transmitted wave, therefore, can be synthesized from the transmitted waves of the plane incident wave. Snapshots of the acoustic field are presented that show the procedure of the interaction of the incident wave with the interface of the media. [Work supported by National Natural Science Fundation of China, No. 10534040.]

TUESDAY AFTERNOON, 6 JUNE 2006 ROOM 556AB, 3:45 TO 5:20 P.M.

Session 2pPP

Psychological and Physiological Acoustics: Spatial Hearing

Norbert Kopco, Chair

Boston Univ., Cognitive and Neural Systems, 677 Beacon St., Boston, MA 02215

Chair's Introduction—3:45

Contributed Papers

3:50

2pPP1. Spatial audio target acquisition in the frontal horizontal plane: Effects of feedback, mobility, and index of difficulty. Georgios Marentakis and Stephen Brewster (Dept. of Computing Sci., Univ. of Glasgow, 17 Lilybank Gardens, G12 8QQ UK)

The results of an empirical study investigating the effect of feedback, mobility, and index of difficulty on a deictic spatial audio target acquisition task in the horizontal plane in front of a user are presented. Audio feedback marked display elements are found to enable usable deictic interaction that is performed with high accuracy and can be described using Fitts' law. Feedback does not affect perceived workload or preferred walking speed compared to interaction without feedback. Without audio feedback, spatial audio is not found to support adequately the corrective submovements necessary for high selection accuracies. Interaction with the display becomes therefore highly inefficient. Mobility is found to degrade interaction speed and accuracy by 20%. Participants were able to perform deictic spatial audio target acquisition when mobile while walking at 73% of their preferred walking speed. The proposed feedback design is examined in detail and the effects of variable target widths are quantified. Deictic interaction with a spatial audio display is found to be a feasible solution for future interface designs.

4:05

2pPP2. Effect of roving on spatial release from masking for amplitude-modulated noise stimuli. Norbert Kopčo, Jaclyn J. Jacobson, and Barbara G. Shinn-Cunningham (Hearing Res. Ctr., Boston Univ., 677 Beacon St., Boston, MA 02215, kopco@bu.edu)

A previous study investigating detection of broadband signals embedded in broadband noise showed that detection thresholds depended both on the temporal structure of the target and masker and their simulated spatial locations [Kopco and Shinn-Cunningham, J. Acoust. Soc. Am. **117**, 2396 (2005)]. The current study extends this work by introducing a rove in overall intensity to reduce the reliability of overall intensity cues for detecting the target. Detection thresholds were measured for a broadband noise target spectrally and temporally centered within a broadband noise masker. Thresholds were measured for all combinations of six spatial configurations of target and masker and five modulation conditions (all combinations of target modulated and unmodulated and of masker modulated and unmodulated; when both target and masker were modulated, the modulation could either be equal or π out of phase). The amplitude modulation, if present, had a rate of 40 Hz and depth of 0.5. The masker level was roved by ± 5 dB between intervals within a trial. Results suggest that the cues underlying target detection depend on both the spatial configura-

tion of target and masker and their temporal structure, but that, when available, overall intensity cues also contribute to a target detection. [Work supported by NSF and ONR.]

4:20

2pPP3. Experiments in speech localization using headphones and loudspeakers. Casey Canfield, Michelle Vu (Garrison Forest School, 300 Garrison Forest Rd., Owings Mills, MD 21117), Ilene J. Busch-Vishniac, and James E. West (Johns Hopkins Univ., Baltimore, MD 21218)

This experiment studies the ability of humans to localize sound as recording and playback methods are varied. Further, because subjects are asked to indicate personal preference, we are able to determine the importance of localization in setting personal preference. The stimulus used in this experiment was speech (6 stanzas of a poem, each recorded at a different location). Three recording methods were used: monaural, stereo, and stereo through an artificial head. Two playback methods were used: stereo speakers and stereo headphones. Results of the work are applicable to the design of conference facilities and the design of headphones and speakers. [This work is a Wise Women project. Wise Women partners high school students from Garrison Forest School with Johns Hopkins engineering faculty.]

4:35

2pPP4. Evaluating the principal spectral components positioning a virtual sound source on a cone of confusion. David H. Benson and William L. Martens (CIRMMT, Faculty of Music, McGill Univ., 555 Sherbrooke St. West, Montreal, QC H3A-1E3, Canada, david.benson@mcgill.ca)

Although many studies have attempted to identify spectral cues to sound-source direction for the entire sphere of possible incidence angles, few studies have focused their attention on directional hearing for virtual sources positioned on cones of confusion. These regions of space, defined by relatively constant interaural time and intensity differences, provide an ideal subset of incidence angles for testing the importance of the spectral cues to direction allowing for up/down and front/rear distinctions between sources at constant angular distance from the median plane. In this study, head-related transfer functions (HRTFs) measured on a well-lateralized cone of confusion were decomposed into principal spectral components putatively responsible for positioning a virtual sound source, and their related angle-dependent scores that describe the relative contribution of each shape to the total spectral variation. The findings can be summarized as follows: For the set of ipsilateral HRTFs the most significant set of scores shows sinusoidal variation with angle, and could potentially encode a front-back cue, though with the extrema rotated by about 30 deg. The most significant score from an analysis of interaural spectral differences seems to capture a head-shadowing effect which is most extreme in the lower rear.

4:50

2pPP5. The effect of spectral differences on the ability to judge the laterality of a simulated source followed by a simulated echo. Joseph B. Boomer, Raymond H. Dye, Jr., Stanley Sheft, and William A. Yost (Parmly Hearing Inst., Loyola Univ., 6526 N. Sheridan Rd, Chicago, Chicago, IL 60626)

The purpose of this study was to examine listeners' abilities to lateralize a source click based on interaural difference of time (IDTs) when followed by a spectrally different echo click. The first click was centered at 3000 Hz and the second was centered at 1500, 2000, 3000, 4000, or 5000 Hz. A correlational analysis was carried out to assess the relative weight given to the IDTs of source and echo clicks for echo delays of 8–128 ms. Also measured were proportions correct and proportions of responses predicted from the weights. Listeners were instructed to indicate the laterality of the source click. The IDTs of the source and echo clicks were selected independently from Gaussian distributions ($=0$ s, $=100$ s). When the second click had energy in a spectral region above that of the first, source weights and proportions correct were higher than when the spectral region of the second was lower than the first click. For some observers, this low-frequency dominance was evident for echo delays as long as 128 ms. The effect of echo frequency was constant across all echo delays, although source weights and proportions correct were highest at an echo delay of 8 ms.

5:05

2pPP6. Learning of interaural-level-difference discrimination with a 0.5-kHz tone in human adults. Yuxuan Zhang and Beverly A. Wright (Dept. of Commun. Sci. and Disord., and Northwestern Univ. Inst. for Neurosci., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208)

Sound-source location on the horizontal plane is primarily determined by two cues: interaural time differences (ITDs) for low-frequency sounds and interaural level differences (ILDs) for high-frequency sounds. Though not available in natural listening environments, ILDs in low-frequency sounds, when artificially delivered to listeners through headphones, also can be used to lateralize sounds. Little is known about the neural processing of such low-frequency ILDs. Here, the modifiability of this processing was examined by training human adults on ILD discrimination with a 0.5-kHz tone on the midline. After nine 1-h daily training sessions, trained listeners ($n=6$) improved significantly more than untrained controls ($n=8$) on the trained condition. This learning generalized to a 4-kHz tone and a 0.5-kHz tone amplitude modulated at 30 Hz, but appeared not to generalize to a 1-kHz tone, an off-midline position, or ITD discrimination with the trained tone. This overall learning pattern differs from those previously obtained for ILD at 4 kHz and ITD at 0.5 kHz using the same training paradigm [Wright and Fitzgerald, Proc. Natl. Acad. Sci. U.S.A. **98**, 12307–12312 (2001)], suggesting that the processes involved in low-frequency ILD discrimination differ from those encoding high-frequency ILDs and low-frequency ITDs, at least at the neural site(s) affected by training. [Work supported by NIH/NIDCD.]

Session 2pSAa

Structural Acoustics and Vibration: Structural Acoustics of Vehicles

Gerard P. Carroll, Cochair
Naval Surface Warfare Ctr., Carderock Div., 9500 MacArthur Blvd., West Bethesda, MD 20817

Hari S. Paul, Cochair
Acoustical Foundation, 101/20 First Main Rd., Gandhi Nagar Adyar, Chennai 600020, India

Contributed Papers

1:15

2pSAa1. Unsteady surface pressures on a ship hull. Michael Goody, Theodore Farabee, and Yu-Tai Lee (Signatures Directorate, Naval Surface Warfare Ctr., Carderock Div., 9500 MacArthur Blvd., West Bethesda, MD 20817-5700, Michael.Goody@navy.mil)

Unsteady surface pressure measurements at locations distributed on the surface of a ship model hull and flow visualization of the bow wave and free surface are discussed. The measurements were performed in the David Taylor Model Basin using a ship model that is approximately 40 ft. long and representative of a naval combatant at Froude numbers from 0.14 to 0.44. The pressure measurement locations are distributed over the hull surface from 15 surface pressure spectrum and limited spatial coherence measurements are discussed. Additionally, the measured surface pressure spectra are compared to predictions done using an RANS-based formulation and predictions done using an empirical formulation that is based on historical, flat-plate, surface pressure spectra. [Work sponsored by Office of Naval Research, Code 334.]

1:30

2pSAa2. Power flow and wave propagation characteristics of surface ship inner-bottom structure. Robert C. Haberman (BBN Technologies, Old Mystic Mill, 11 Main St., Mystic, CT 06355, rhaberman@bbn.com)

The construction of the inner bottom of a surface ship consists primarily of deck plating connected to hull plating by longitudinal vertical plates. Foundations supporting equipment typically line up with the vertical plates, thereby providing a direct path of mechanical excitation to the hull resulting in forced acoustic radiation. The purpose of this paper, therefore, is to provide an overview of the physics that controls wave propagation in surface ship inner-bottom structure along with a description of the power flow associated with direct force excitation of the vertical plating. Two theories will be presented: (1) an approximate theory based on the vertical plate modeled as a beam coupled to the deck and hull modeled as plates containing flexural waves and (2) an exact theory based on all three structures modeled as plates containing flexural, compressional, and shear waves. Numerical results will be presented that show the importance of full wave coupling along with some interesting phenomena associated with power flow and leakage of energy into deck and hull plating. The implication of these results in statistical energy analyses modeling will be briefly discussed. [Work partially funded by ONR, Code 334 Dr. S. Schreppler.]

1:45

2pSAa3. Radiated acoustic power of model ship hull sections estimated from scanning laser Doppler vibrometer vibration response measurements. Gerard P. Carroll (Naval Surface Warfare Ctr. Carderock Div., 9500 MacArthur Blvd., Bethesda, MD 20817)

This paper will present radiated acoustic power estimates for ship hull models which are based entirely on scanning laser Doppler vibrometer measurements of the underwater vibrating surface.The radiated power estimates are compared to direct measurements of the radiated acoustic power obtained using hydrophones in the reverberant tank. These radiation estimates are based on Rayleigh integral formulations, one for which the vibrating surface is assumed to be baffled, and the other for which it is unbaffled. For the latter case, the pressure in the plane of the vibrating surface is assumed to be zero outside the vibrating surface. This corresponds to a mixed boundary condition in which the velocity is known on the plate and the pressure is known off the plate. An iterative application of the Rayleigh integral is used to determine the velocity and the pressure over the entire plane. For the purposes of calculating the radiated power, this formulation is shown to iterate much more quickly by iteratively calculating the surface intensity. This is possible since the pressure and therefore the surface intensity is zero outside the plate boundary. Utilizing these surface radiation intensity calculations, it is possible to then investigate the sources and mechanisms of the hull radiation.

2:00

2pSAa4. Numerical investigation of small-scale surface ships by energy finite-element analysis. Kuangcheng Wu (Dept. of Signatures and Hydrodynamics, Northrop Grumman Newport News, 4101 Washington Ave., Newport News, VA 23607, kcwu@ngc.com) and Nick Vlahopoulos (Univ. of Michigan, MI 48109)

Experimental investigations have been performed to understand the structural acoustic characteristics of two small-scale surface ship models representing sections of conventional and advanced double-hull ship designs [Craun *et al.*, "Experimental and analytical investigation of surface shiphull structure acoustics using physical models," J. Acoust. Soc. Am. **116**, 2569 (2004)]. Recently, new developments have been added to the energy finite-element analysis (EFEA) for surface ships [Wu and Vlahopoulos, "Vibratory response of surface ships predicted by energy finite element approach," J. Acoust. Soc. Am. **116**, 2569 (2004)]. The presentation will compare hull responses from EFEA predictions to test data for the two small-scale models to validate the new development in EFEA. Furthermore, the impact of the hull responses due to the different model designs will be discussed. [Support for this research is provided by ONR code 334.]

2:15

2pSAa5. Application of a hybrid finite-element formulation for structure-borne vibration in an automotive vehicle structure. Sang-Bum Homg (Cummins, MC 50182, 1900 McKinley Ave., Columbus, IN 47201) and Nickolas Vlahopoulos (Univ. of Michigan, Ann Arbor, MI 48109)

The hybrid FEA method combines the conventional FEA method with energy FEA (EFEA) for analyzing systems that contain both flexible and stiff members. Conventional FEA models are employed for modeling the behavior of the stiff members in a system. Appropriate damping elements are introduced in the connections between stiff and flexible members in order to capture the presence of the flexible members during the analyses of the stiff ones. The component mode synthesis method is combined with

analytical solutions for determining the driving point conductance at joints between stiff and flexible members and for defining the properties of the damping elements which represent the flexible members when analyzing the stiff components. Once the vibration of the stiff members and the amount of power dissipated at the damping elements has been identified, an EFEA analysis is performed in order to determine the amount of vibrational energy in the flexible members. The new developments are validated by comparing results of the hybrid FEA with results from very dense conventional finite-element analyses for a typical body-in-white vehicle structure. [Work sponsored by the Automotive Research Center, supported by the U.S. Army NAC.]

2:30

2pSAa6. The role of structural acoustics in the space shuttle flowliner crack investigation. Michael J. Allen and Robert C. Haberman (BBN Technologies, Old Mystic Mill, 11 Main St., Mystic, CT 06355, mjallen@bbn.com)

During normal inspection of the Space Shuttle main propulsion fuel line system small cracks were identified in a slotted flowliner just upstream of the main propulsion engines. The cause is believed to be resonance excited by pressure fluctuations originating from the downstream fuel pump. This presentation highlights the analyses that were performed to characterize the structural acoustic environment surrounding the flowliner during waterflow and liquid hydrogen laboratory experiments, as well as in-flight conditions. Analyses employed a combination of analytical and finite-element methods to model fluid structure interaction and inverse methods to reconstruct equivalent acoustic sources representative of the fuel pump. Numerical results will be presented that compare experimental strain and pressure data with theoretical counterparts. Implications will be drawn regarding the applicability of the theory to the actual shuttle fuel line system. [Work was sponsored by NASA Marshall Flight Center.]

2:45

2pSAa7. Structural monitoring from noise cross correlation. Karim G. Sabra, William A. Kuperman (Marine Physical Lab., Scripps Inst. of Oceanogr., U.C. San Diego, 9500 Gilman Dr., La Jolla, CA, 92093-0238, ksabra@mpl.ucsd.edu), Eric S. Winkel, Dwayne A. Bourgoyne, David R. Dowling, Steve L. Ceccio, and Marc Perlin (Univ. of Michigan, Ann Arbor, MI 48109-2133)

It has been demonstrated theoretically and experimentally that an estimate of the Green's function between two receivers can be obtained from the long-time average of the cross correlation of ambient noise at these two receivers in various environments and frequency ranges of interest: ultrasonics, underwater acoustics, and seismology. This result provides a means for structural monitoring using the ambient structure-borne noise only, without the use of active sources. The coherent wave-fronts emerge from a correlation process that accumulates contributions over time from noise sources whose propagation path passes through both receivers. We present experimental results using random vibration data collected on eight accelerometers mounted on a hydrofoil or plate at the Navy's William B. Morgan Large Cavitation Channel. The tests were conducted at high Reynolds number, with the primary excitation source being turbulent boundary-layer pressure fluctuations on the upper and lower surfaces of the plate or foil. Identical deterministic time signatures emerge from the noise cross-correlation function based on robust and simple signal processing between passive sensors. These time signatures can be used to estimate and monitor the structural response of the test models.

TUESDAY AFTERNOON, 6 JUNE 2006 ROOM 551B, 3:15 TO 4:45 P.M.

Session 2pSAb

Structural Acoustics and Vibration: Damping and Materials

Courtney B. Burroughs, Chair
1241 Smithfield St., State College, PA 16801

Contributed Papers

3:15

2pSAb1. A strain energy approach to the flexural damping of layered plates. Richard J. Deigan and Kin Ng (NSWCCD, 9500 MacArthur Blvd, W. Bethesda, MD 20817-5700)

The flexural damping factor of multilayered plates is developed using a strain energy approach. The derivation is based on the exact, steady-state solution for a plate subjected to boundary conditions that are harmonic in time and position along the plate. The distribution of elastic strain energy contained within the individual layers of the plate is developed, from which an effective, flexural damping factor is determined. Comparisons to the widely used RKU analysis [D. Ross, E. E. Ungar, and E. M. Kerwin Jr., ''Damping of Plate Flexural Vibrations by Means of Viscoelastic Laminae,'' ASME, Section 3, pp. 49–87 (1959)] are made for selected cases of constrained and unconstrained layer damping. Good agreement with RKU theory is demonstrated for the selected cases. Advantages of the present analytic approach include: (1) The ability to model the effect of fluid loading; (2) Unlimited number of plate layers; and (3) The handling of compressible layers and associated loss mechanisms. Analytical estimates of damping factors are compared with measurements made on steel bars treated with an unconstrained damping layer. Stress-state assumptions are shown to be a major factor affecting theory-data agreement.

3:30

2pSAb2. Dynamic modulus measurement using the Dynamic Mechanical Analyzer. Kin Ng, Richard Deigan, and Samedi Koch (Naval Surface Warfare Ctr., Carderock Div., 9500 MacArthur Blvd., West Bethesda, MD 20817-5700)

The Navy has used the Dynamic Mechanical Thermal Analyzer (DMTA) as a tool to characterize material properties over the last 20 years. It is now evaluating the Dynamical Mechanical Analyzer (DMA) as a replacement for the DMTA, with the goal to insert it into operation in 2006 with minimal disruption to current acquisition programs. This paper compares the two instruments with regard to their capabilities and looks at some of the transition issues.

3:45

2pSAb3. Space-dependent damping for reducing vibration of continuous systems. Mark H. Holdhusen (Univ. of Wisconsin—Marathon County, 518 S 7th Ave., Wausau, WI 54401)

Structural vibration in continuous systems is a common source of problems for many mechanical systems. Much research has considered ways to reduce the vibration in continuous systems. Some research uses passive techniques that are simple to implement, whereas other methods

implement active control that can be quite complex. The research presented here investigates the effect space-dependent damping has on the vibration of continuous systems. Specifically, the work at hand will focus on the flexural vibration of beams. In a distributed damping beam system, the damping is variable as a function of position along the length of the beam. Such a system could be physically realized through the use of magneto-rheological or electro-rheological fluids. Along its length, a beam could be divided into several isolated sections containing such a fluid. By controlling the magnetic or electric field across each section of the beam, the damping would vary along the length of the beam. The work presented here optimizes the damping as a function of beam position to reduce the broadband frequency response of the beam. A number of support conditions for the vibrating beam are considered.

4:00

2pSAb4. Multiple-particle tracking rheometry of gelling polymer solutions. Ryan Braun (NSWCCD, 9500 MacArthur Blvd, W. Bethesda, MD 20817-5700) and Dennis Wirtz (Johns Hopkins Univ., Baltimore, MD 21218)

Many polymer materials used for damping are generated utilizing polycondensation reactions. These polymer reactions may start in the liquid phase, undergo gelation, and then solidify. The microscopic, three-dimensional, heterogeneous network produced during the chemical reaction greatly affects the bulk material properties. The polymer architecture is mainly locked in at the point of gelation, when polymer mobility is drastically reduced. This study investigates the steady-state microstructure and micromechanics of gelling polymer materials using the recently developed method of multiple-particle tracking microrheometry (MPTM). Unlike existing methods, MPTM quantitates, in real time, the microrheology and microheterogeneity of complex fluids undergoing gelation. Brownian trajectories of probe particles are transformed into families of mean-squared displacements, from which distributions in local, frequency-dependent viscoelastic moduli or time-dependent creep compliance are readily extracted. To understand the effects the microscopic, three-dimensional architecture of a gel has on the macroscopic, bulk properties, a model polysaccharide gel system was studied. The steady-state properties of the polysaccharide gels were evaluated as a function of concentration.

4:15

2pSAb5. Metal-polymer-ceramic constrained layer damping system. Liya Bochkareva (UIIP Natl. Acad. of Sci. of Belarus, Filatova 7-28, Minsk 220026, Belarus) and Maksim Kireitseu (Univ. of Sheffield, UK)

The focus of this paper is directed toward the investigation into multilayered metal-polymer-ceramic damping coatings and their FEM-based computer modeling that can provide an alternative spacer material in a constrained layer damping configuration so as to enable design of three and more layered damping structures, but with controlled structure and mechanical properties of the gradient damping layers. An advanced measuring system for the analysis of vibrating structures is based on a laser scanning vibration interferometer. The position of a ceramic layer in a constrained layer configuration does matter. The outward position is most favorable for high strength by ceramic layer and damping by constrained aluminum-polymer layer. Modal loss factors obtained in the 60%–80% partial-coverage test cases of cylindrical shell are higher than the maximum modal damping obtained in the case of 100% coverage. The modal damping for the target modes is higher than in the complete-coverage case. An effective damping design involves selecting a proper combination of area coverage, relative thickness and stiffness values of the layers of the damped configuration. [Dr. Bochkareva is currently continuing her research work under 2-year EU INTAS 2005-2007 postdoctoral fellowship Ref. No. 04-83-3067.]

4:30

2pSAb6. Carbon nanotube-reinforced damping materials: Design and experiments. Maksim Kireitseu (INDMASH Natl. Acad. of Sci. of Belarus, Filatova 7-28, Minsk 220026, Belarus)

The CNT-reinforced material damping phenomenon is complex because of the variety of energy dissipation/fracture mechanisms involved, the complex nature of the nanoparticles themselves, etc. that affect damping/dynamics. The focus of this paper is directed toward the experimental validation of the next-generation vibration damping materials, providing a road map to experimental calibration and measurement of these materials. Modified modal strain energy method and FEM-based model created in MSC.visualNastran is employed to support an experiment. SWNT-reinforced polymeric matrix composites are prepared using an extrusion procedure. SWNT particles produced by Dynamic Enterprises, UK (2–5 nm in diameter and 70–90 vol % sample purity) are added in concentrations of 3%–6% by weight along with a surfactant (polyoxyethylene) to aid in dispersion. Resonance frequencies, the mode shape, and damping at each mode were determined by laser vibrometry at standard vibration shaker tests and Bruel & Kjaer vibration system. Data acquisition and control of the electro-dynamic system is based on the computer measurement dystem that is highlighted. Resonance frequency is determined from peaks on a frequency response curve for each of the experimental methods. [Work supported by Marie-Curie Fellowship Ref. No. 021298-Multiscale Damping 2006-2008 at University of Sheffield, UK with Prof. G.R. Tomlinson.]

Session 2pSC

Speech Communication: Speech Production (Poster Session)

Rajka Smiljanic, Chair
Northwestern Univ., Dept. of Linguistics, 2016 Sheridan Rd., Evanston, IL 60208

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

2pSC1. Finite-element simulations of self-oscillation mechanisms in human phonation. Douglas Cook, Luc Mongeau, and Paul Smith (Purdue Univ., 585 Purdue Mall, West Lafayette, IN 47907)

The fluid-structure interactions which occur during human phonation were investigated using a commercially available finite-element program. Both lumped mass and continuous structural models were coupled with viscous fluid flow. The simulated behavior was analyzed. The addition of viscous flow was not sufficient to induce self-oscillation of a single degree-of-freedom lumped mass model but did contribute to the steady vibration of the continuous model. This observation reinforces the conclusion that structural complexity is an essential mechanism of self-oscillation. The influence of false folds on the oscillation patterns of the continuous model was also examined.

2pSC2. Characteristics of phonation in women who produce voiced versus voiceless /h/. Laura L. Koenig (Haskins Labs., 300 George St., New Haven, CT 06511 & Long Island Univ., Brooklyn, NY 11201)

Production of the consonant /h/ typically involves a vocal-fold abduction gesture. Although /h/ is traditionally described as voiceless in English, voiced allophones also occur. Our recent work has described the phonatory behavior of women who produce a mixture of voiced and voiceless /h/; our results suggested speaker-specific patterns of the factors that lead to devoicing. This study characterizes the behavior of normal women who produce exclusively voiced /h/ and compares it to those who produced both voiced and voiceless /h/. Measures of baseline (dc) airflow are used as an index of abduction degree; peak intraoral pressures during adjacent /p/ closures provide a measure of subglottal pressure; and pulse-by-pulse measures are made of $f0$ and pulse amplitude (ac flow) around the /h/ abduction gesture. Multidimensional analyses are carried out within speakers to determine how these vocal parameters are related. We expect that women who produce only voiced /h/ will differ from those who devoice some /h/ productions in the parameters that facilitate phonation, such as higher subglottal pressures, smaller abduction degrees, and reduced vocal-fold tension as reflected in $f0$ measures. Based on our previous findings, however, we also expect individual variation in the phonatory settings that produce voiced /h/. [Work supported by NIH.]

2pSC3. Contextual effects on the continuancy of /ð/. Sherry Zhao (MIT, 77 Massachusetts Ave., Rm. 36-545, Cambridge, MA 02139)

This research examined the segmental-contextual effects on the continuancy of the voiced dental /ð/ in American-English and whether certain acoustic attributes were preserved despite possible modification. Word-initial /ð/ cases, extracted from continuous speech of 146 speakers from the TIMIT database, were frequently (averaging 65%) stoplike when it was in utterance-initial position or when its preceding phoneme was voiceless and/or [-continuant]. This stoplike modification occurred less frequently (averaging 39%) when /ð/ was preceded by a voiced fricative and rarely (averaging 13%) when preceded by a vowel or liquid consonant. A comparison of stoplike /ð/ and /d/ cases under similar contexts showed that the burst peak location, burst spectrum shape, and F2 at vowel onset averages were all statistically different between the two groups. In addition, the acoustic data suggested that the dental place of articulation was preserved for the modified /ð/. Preliminary automatic classification experiments involving salient acoustic attributes seemed to indicate that F2 at vowel onset may be a reliable cue for the dental place of articulation in /ð/. [Work supported by NIH DC02978 and T32DC00038.]

2pSC4. Modeling the influence of acoustic loading on laryngeal self-sustained oscillations. Matías Zañartu, Luc Mongeau, and George R. Wodicka (Ray W. Herrick Labs. Purdue Univ., West Lafayette, IN 47907)

Three-way interactions between sound waves in the subglottal and supraglottal tracts, the vibrations of the vocal folds, and laryngeal flow were investigated. Nonlinear coupling was investigated analytically, with a focus on the effects of a realistic acoustic loading on the self-sustained oscillation of the vocal folds. MRI-based shapes for the supraglottal tract [Story *et al.*, J. Acoust. Soc. Am. **100**, 537–554 (1996)] and an equivalent acoustical model for the subglottal tract [Harper *et al.*, IEEE Trans. Biomed. Eng. **48**, 543–550 (2001)] were considered. The model was based on the wave reflection analog technique, along with a novel single degree of freedom model of the vocal folds to describe the source. The effects of a mucosal wave on the vocal folds were incorporated through a time-varying discharge coefficient and a Bernoulli-type flow model. The relative importance of acoustic loading and time-varying flow resistance for fluid-structure energy transfer was compared for various configurations. [Work supported by NIH Grant RO1-DC05788.]

2pSC5. Patterns of interarticulator coordination within the syllable. Patrizia Bonaventura and Joseph Sand (Dept. of Commun. Sci., Case Western Reserve Univ., 11206 Euclid Ave, Cleveland, OH 44106, patrizia.bonaventura@case.edu)

Interarticulatory coordination patterns were observed in movements of the crucial articulator for the production of place in selected syllables containing labiodental and alveolar consonants. The coordination of the lower lip with upper lip and tongue tip, and the coordination of the tongue tip with tongue blade and lower lip have been observed. The goals of the study were (a) to verify whether stable patterns of speed with respect to excursion of movement [Fujimura, O., "Temporal organization of articulatory movements as a multidimensional phrase structure," Phonetica, **38**, 66–83 (1981a); "Icebergs revised," ASA meeting, 2aSC (1996b)], were

found in concurrent articulators to the crucial one for place, under different prosodic conditions; (b) whether such stable portions would vary proportionally to the increase in amplitude of the vertical jaw movement, selected as a measure of the magnitude of the syllable; (c) to observe relative timing at the iceberg thresholds in the selected consonantal movements with respect to the center of the syllable. Analyses were performed on articulatory recordings from cinefluorographic data, based on read dialogues [Erickson *et al.*, ''Articulatory correlates of prosodic control: Emotion and emphasis,'' Lang. Speech, **41**, 399–417 (1998)]. Iceberg patterns were found in the concurrent articulator movements, showing a linear dependence of slope on the total excursion of the demisyllabic movement. Timing of concurrent articulator movements with respect to the center of the syllable is shown to be relatively stable.

2pSC6. Modulation of frication noise in a dynamic mechanical model of the larynx and vocal tract. Anna Barney (ISVR, Univ. of Southampton, Southampton, SO17 1BJ, UK, ab3@soton.ac.uk) and Philip J. B. Jackson (Univ. of Surrey, Guildford, GU2 7XH, UK)

A dynamic mechanical model of the larynx and vocal tract was used to investigate the aerodynamics and acoustics of voiced fricatives. The model had typical dimensions for a human adult male. Glottal excitation arose from driven shutters representing the vocal folds; sinusoidal and square-wave driver functions were investigated. Glottal amplitude, fundamental frequency ($f0$), and volume velocity were independently controlled. Frication noise was produced by an orifice plate with a sharp-edged obstacle on its downstream side. Two different positions of constriction within the vocal tract, each coupled with a different constriction-obstacle distance, were investigated. Vocal-tract wall pressure was measured at three locations and a microphone at the vocal-tract exit measured radiated sound. The sound radiated from the open end of the vocal tract was found to have both harmonic and noise components. The noise was modulated by the $f0$ of the glottal vibration. The depth of modulation was independent of the shape of the glottal excitation. It showed an overall decreasing trend with increasing $f0$ in the range 20 to 150 Hz, but with a local maximum of modulation depth around 80 Hz. The noise bursts had maxima coincident with the middle of the glottal open phase.

2pSC7. Acoustic characteristics of clearly produced fricatives. Kazumi Maniwa, Allard Jongman (Linguist. Dept., Univ. of Kansas, 1541 Lilac Ln., Blake Hall, Rm. 427, Lawrence, KS 66044-3177), and Travis Wade (Posit Science Corp., San Francisco, CA 94104)

Research suggests that speakers can adopt a speaking style that allows them to be understood more easily when confronted with difficult communication situations, but few studies have examined the acoustic properties of clearly produced consonants in detail. This study attempts to characterize the type and magnitude of adaptations in the clear production of English fricatives in a carefully controlled range of communication situations. Ten female and ten male talkers produced nonsense syllables containing the eight English fricatives in VCV contexts, both in a conversational style and in a clear style (elicited by means of feedback consisting of simulated recognition errors received from an interactive computer program). Acoustic measurements were taken for spectral, amplitudinal, and temporal properties known to influence fricative recognition. Results illustrate that (1) there were some consistent overall clear speech effects, several of which (consonant duration, spectral peak location, spectral moments) were consistent with previous findings and a few (notably consonant-to-vowel intensity ratio) which were not; (2) ''contrastive'' differences related to eliciting prompts were observed in a few key comparisons; and (3) talkers differed widely in the types and magnitude of acoustic modifications.

2pSC8. Articulation of fricatives: Evidence from X-ray microbeam data. Christine H. Shadle, Khalil Iskarous, and Michael I. Proctor (Haskins Labs., 300 George St., New Haven, CT 06511)

Articulatory changes in /s/ over time give rise to the onset and offset of turbulence noise, but the small tolerances involved make it difficult to separately consider the effects of movement, phonetic context, and subject variation. Aerodynamic estimates of the constriction area have been shown to demonstrate a marked asymmetry, with the /s/ onset more rapid than offset [C. Scully, Speech Commun. **11**, 1992]. Previous articulatory studies led us to expect an invariant tongue tip position, with jaw raising dependent somewhat on the following vowel. In order to investigate the timing of the constriction formation in fricatives, X-ray microbeam data [Westbury *et al.*, 1994] for citation forms of /sVd/ for 15 vowels were studied. In fact, the tongue marker patterns during [s] vary both by following vowel and subject; subjects seem to have different apical-laminal strategies. The jaw reaches maximum height during [s] for each subject studied; that height does not appreciably vary with vowel. This indicates that jaw height is important specifically for /s/ production. One possibility is that the constriction is formed by the tongue, and turbulence noise is then maximized by the lower teeth being brought into position. [Work supported by NIH NIDCD R01 DC 006705.]

2pSC9. A range of intonation patterns produced in an elicitation task. Alejna Brugos, Jonathan Barnes (Program in Appl. Linguist., Boston Univ., 96 Cummington St. #246, Boston, MA 02215, abrugos@bu.edu), Stefanie Shattuck-Hufnagel (MIT, Cambridge, MA 02139), and Nanette Veilleux (Simmons College, Boston, MA 02115)

A production task was designed to elicit a specific intonation contour, characterized by an exaggerated rise-fall-rise in $f0$ ($L+H^*L-H\%$ in ToBI terms), conveying an attitude of ''dismayed surprise.'' Speakers were given models, a short practice period, and a preceding context for each utterance. A targetlike rise-fall-rise contour (including $H^*L-H\%$ variants) was successfully elicited for each of the contexts, but in less than half the utterances: Of 1134 utterances, 631 (56%) were judged by a panel of 3 experienced ToBI labelers as different from the intended contour, e.g., $L^*H-H\%$. Of these, 261 were produced with contours perceptually similar to the target, with a sustained flat or minimally rising $f0$ after the fall from the pitch accent, or with the rise occurring early in the next phrase. Alignment characteristics and individual speaker data may determine whether these are one contour with a continuum of realizations, or separate contours. These results (a) suggest that speakers have more than one way of signaling a given attitude and resist being limited to using only one, and (b) support the need for careful postexperiment transcription to characterize which contours subjects have actually produced. [Work supported by NSF 0345627 and NIH DC00075.]

2pSC10. Factors affecting phonetic variation of American English /k/. Stefanie Shattuck-Hufnagel (Speech Group, Res. Lab. of Electron., MIT, Cambridge, MA 02139) and Nanette Veilleux (Simmons College, Boston, MA 02115)

Word-final consonant reduction is commonly observed in American English, but variation in word-initial consonants is less thoroughly studied. Several considerations suggest that initial consonants might resist reduction; a more nuanced hypothesis is that initial consonants resist reduction in prosodically strong positions (i.e., at phrase edges and pitch accents) but may be reduced in weaker positions (e.g., phrase medially, in nonaccented syllables), particularly in un-self-conscious speech. A corpus of 16 spontaneous dialogues (8 speakers), prosodically labeled using the ToBI transcription system, provides a testbed for this view: the speakers were pairs of close friends engaged in an absorbing task which distracted them from paying close attention to their speech. Preliminary inspection of initial /k/'s from 3 speakers revealed a range of pronunciations (as described in earlier work), from normal prevocalic /k/ (with a silent closure period followed by release noise), to ''leaky'' /k/'s with some noise during the closure period and higher-amplitude release noise, to fricated /k/'s with

constant-amplitude noise throughout, and offered some support for reduction in weak prosodic contexts. Further analysis will compare prosodic context with other factors governing phonetic variation (e.g., word frequency); implications for speech recognition and production will be discussed. [Work supported by NIH DC00075 and DC02978.]

2pSC11. Errors and strategy shifts in speech production indicate multiple levels of representation. Bryan Gick (Dept. of Linguist., Univ. of BC, E270-1866 Main Mall, Vancouver, BC V6T1Z1 Canada)

An earlier pilot study [Gick, J. Acoust. Soc. Am. **112**, 2416 (2002)] observed that speakers producing challenging sequences of articulatory movements exhibit categorical shifts in production strategies. For example, a speaker may initiate all two-flap sequences with an upward tongue movement, then shift to a downward movement, then shift again. The present study attempts to determine how these strategy shifts can inform phonologists as to levels of representation. Ten subjects produced sequences of flap allophones of English /t/ and /d/ in an experimental paradigm similar to the pilot study. Results show strategy shifts in the speech of 7/10 subjects, such that: (1) shifts are more likely to occur following speech errors; (2) errors are more likely to occur in sequences with more flaps and more conflicts; (3) errors in response to articulatory conflicts occur at multiple levels of representation (segmental, lexical, etc.); (4) strategy shifts occur at multiple levels of representation (subsegmental, segmental, whole word); (5) word frequency does not affect strategy shifts. These results suggest that speech processing is active in parallel at multiple levels associated with phonological representation. [Work supported by NSERC.]

2pSC12. 3-D acoustic analysis based on magnetic resonance imaging of a "retroflex" and "bunched" American English rhotic sound. Xinhui Zhou, Carol Espy-Wilson (Dept. of Elec. and Comput. Eng., Univ. of Maryland, 2205 AV Williams Bldg., College Park, MD 20742, espy@glue.umd.edu), Mark Tiede (Haskins Labs, New Haven, CT 06511), and Suzanne Boyce (Univ. of Cincinnati, Cincinnati, OH 45202)

A variety of articulatory configurations can produce an American English /r/. Common across all configurations is a large front-cavity volume that results directly or indirectly in the characteristically low third formant. Earlier work [Zhang *et al.*, ICASSP2005 1, pp. 893–896] using 2-D finite-element modeling of the front cavity with stylized bunched and retroflexed shapes showed that, even at 2000 Hz, cross modes are generated in the front cavity. Thus, the front cavity should be modeled as a single tube with a side branch, regardless of whether there is space underneath the tongue. This paper extends the earlier work by (a) considering the full 3-D vocal tract geometry reconstructed from magnetic resonance images, (b) performing 3-D finite-element analysis, and (c) using contrasting tongue shapes from similar-sized subjects. Harmonic analysis showed the frequency response is comparable to the spectrum of the acoustic signal, suggesting good agreement between the actual and reconstructed vocal tracts. As before, wave propagation in the front cavity is not planar. 3-D vocal tract analysis shows that the cross modes have 3-D properties. Application of the 3-D wave propagation property to area function extraction and simple-tube modeling of the vocal tract will be discussed. [Research supported by NIH.]

2pSC13. Syllable structure effects on velum-oral coordination evaluated with real-time MRI. Stephen Tobin, Dani Byrd (USC Linguist., 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, dbyrd@usc.edu), Erik Bresch, and Shrikanth Narayanan (USC Los Angeles, CA 90089)

Seminal work by Krakow (PhD. Diss., Yale University, 1989, Phonetics and Phonology **5**, 87–116, 1993) has indicated that intrasegmental coordination of articulatory gestures varies systematically as a function of syllable position. Specifically, onset and coda nasals display different lag times between the velum and oral gestural components. In onset nasals, velum lowering is achieved roughly synchronously with oral constriction, while in codas, velum lowering precedes oral constriction. However, Krakow's study was limited by its reliance on indirect measures of velum movement. Nor did her work examine nasals other than [m] or syllable organizations other than singleton onsets and codas. Our study employs real-time magnetic resonance imaging [Narayanan *et al.*, J. Acoust. Soc. Am. **115**, 1771–1776 (2004); Mathiak *et al.*, Int. J. of Lang. & Comm. Disorders **35**, 419–425 (2000)] to acquire direct information about the key vocal tract variables of velum aperture and oral constriction, thereby allowing the calculation of relative gestural timing. We investigate the effects of syllabification on interarticulatory timing in [n], and, in addition to examining codas and onsets, further consider juncture geminates spanning word boundaries. Preliminary real-time MRI results from one of three anticipated subjects do find the predicted timing difference for [n]. Further, we find that the relative timing in juncture geminates mirrors that of coda [n] but presents a velum aperture intermediate between coda and onset [n]. [Work supported by NIH.]

2pSC14. Adaptation of speech production to palatal perturbation: Kinematic analysis using electromagnetometry. Douglas M. Shiller (School of Commun. Sci. and Disord., McGill Univ., 1266 Pine Ave. West, Montreal, QC, Canada H3G 1A8, doug.shiller@mail.mcgill.ca), Mark K. Tiede (Haskins Labs., New Haven, CT), and Vincent L. Gracco (McGill Univ., Montreal, QC, Canada)

The extent to which adults are able to flexibly adapt their speech production to altered sensory-motor conditions has considerable implications for oral-motor rehabilitation, yet it remains poorly understood. Previous studies have explored the effects of different physical and perceptual perturbations on speech production, yielding mixed results with respect to the timecourse and extent of adaptation. A major limitation of these studies has been their reliance on acoustical measures as the sole index of motor adaptation. The difficulty with such an approach is the potentially complex relationship between acoustic changes and their underlying articulatory patterns. In the present study, speech adaptation to a palatal perturbation (5-mm thickened alveolar ridge) was investigated by directly measuring articulatory kinematics using electromagnetometry. In contrast to previous claims, subjects were found to produce a significant change in speech-related tongue movement immediately upon insertion and removal of the artificial palate. In addition, subjects gradually fine-tuned their tongue position with continued practice. The observed patterns are consistent with the idea of two independent adaptation processes: a rapid adaptation process that is driven by the mere presence of the false palate and a slower adaptation process driven by the continued presence of the palatal perturbation.

2pSC15. Production of American English word-initial stops in the vicinity of a voiceless stop. Elina Eydlin (Dept. of Speech-Lang. Pathol. and Audiol., New York Univ., 719 Broadway, Ste. 200, New York, NY 10003) and Nassima Abdelli-Beruh (New York Univ., New York, NY 10003)

This study examined how native-American English speakers produced the stop voicing distinction when word-initial stops were positioned before a voiceless stop (/t/) in the sentence repeat CVC ten times. Sentence durations, durations of vowel (/i/) before word-initial stops, number of produced stop closures (/t/, /c/), durations of voicing in closures and VOT were measured. Preliminary analyses on three speakers show that they produced only one closure in most cases. Speakers occasionally produced two closures when slowing down their speaking rate. Vowels preceding /b, d, g/ were barely longer than vowels before /p, t, k/. When speakers produced one closure, closures were occupied by voicing for both /b, d, g/ and /p, t, k/. However, the average durations of voicing in closure were consistently longer for /b, d, g/ than for /p, t, k/. In addition, /b, d, g/ were produced in most cases with short-lag VOT, except when there were two closures, in which case, /b, d, g/ were produced with lead VOT. All /p, t,

k/ were produced with VOT greater than 30 ms. Articulatory studies are needed to assess how the sequence of stops is produced and whether the gestures for /t/ are completely or partially eliminated.

2pSC16. Modeling the effects of the lower airway on vowel spectra. Steven M. Lulich (MIT Speech Commun. Group, 77 Massachusetts Ave., MIT Bldg. 36-511, Cambridge, MA 02139, lulich@mit.edu)

Computational models of the acoustics of the lower airway (below the glottis) have been implemented by researchers in the speech and respiratory sciences. In general, these models are symmetrical and are studied independent of acoustic coupling to the vocal tract. We have constructed a new model that is asymmetrical and based on human morphometric data. The acoustic properties of the model are explored and compared with symmetrical models, and the effects of the model on vowel spectra are studied for varying amounts of lower airway and vocal-tract coupling via the glottis. Cochlear excitation patterns are also computed for the vowel spectra, and the effects of the lower airway acoustics on these excitation patterns are discussed.

2pSC17. Effect of tone on voice onset time in laryngeal and esophageal speech of Mandarin. Manwa Ng (Commun. Sci. and Disord., Long Island Univ., Brookville, NY 11548-1300) and Hanjun Liu (Northwestern Univ., Evanston, IL 60208)

Defined as the time between the release of burst of a stop and the onset of the following vowel, voice onset time (VOT) directly indicates the interarticulator timing between the voicing source and the articulatory movement in the vocal tract. A number of studies have investigated the VOT characteristics of different languages. Yet, few have studied how VOT is affected by tone level in a tone language, and by the type of phonation. The present study investigated the possible effect of tone level and phonation type on such interarticulator timing characteristics. VOT values were measured from the syllables /p^ha/, /t^ha/, and /k^ha/ produced at four Mandarin tone levels by eight laryngeal (NL) and seven esophageal (SE) speakers who were native speakers of Mandarin. Results indicated that significant differences in VOT were found in the NL speaker group, but not in the SE speaker group. With respect to phonation type, SE speakers showed significantly shorter VOT values than NL speakers. Such differences may be related to the use of pharyngoesophageal (PE) segment as a new sound source. SE speakers appear to take a shorter time to start PE segment vibration as compared to NL speakers.

2pSC18. Intrinsic vowel pitch in tracheo-esophageal speech. Michael Kiefte, Natalie Didrachnik, and Jillian Roswell (School of Human Commun. Disord., Dalhousie Univ., 5599 Fenwick St., Halifax, NS B3H 1R2 Canada)

The tendency for high vowels to have higher fundamental frequency (f_0) than low vowels, or intrinsic vowel pitch (IF_0), has been found in both laryngeal and esophageal speech [e.g., Gandour and Weinberg, Phonetica **25**, 140–164 (1980)]. In the present study, eight tracheo-esophageal speakers were asked to produce five repetitions of each of the four point vowels and f_0 was measured. Results showed significantly higher f_0 for high vowels than for low vowels, consistent with previous findings for laryngeal and esophageal speakers. In addition, f_0 was measured from two laryngectomees over a period of 4 months immediately following surgery to determine whether IF_0 is produced automatically or whether it develops over time. Neither of these subjects produced differences in IF_0 at any time within the 4-month period. These preliminary data appear to indicate that IF_0 is not an automatic consequence of either speech physiology or acoustics. More data from additional subjects will be presented. [Work supported by SSHRC.]

2pSC19. Computational aeroacoustics of flow through static model of human vocal tract. Jungsoo Suh, Steven H. Frankel, Luc Mongeau, and Michael W. Plesniak (School of Mech. Eng., Purdue Univ., West Lafayette, IN 47907)

Computational aeroacoustics studies of flow through the human vocal tract, here modeled as a planar channel with an orifice, hence referred to as the glottis, were conducted using large-eddy simulation (LES). Comparisons were made between LES predictions and experimental wall pressure measurements and particle-imaging-velocimetry flow fields. Rigid models of both converging and diverging glottal passages, each featuring a 20-deg included angle and a minimal glottal width of 0.04 cm with transglottal pressure of 15 cm H_2O, were studied. The compressible Navier-Stokes equations were accurately and efficiently integrated for the low Mach number flow through the use of an additive semi-implicit Runge-Kutta method and high-order compact finite-difference schemes for spatial discretization. Characteristic-based nonreflecting boundary conditions were used together with an exit zone in the context of a multiblock approach. Asymmetry of the flow due to a Coanda effect, and transition to turbulence were observed. An acoustic analogy based on the Ffowcs Williams-Hawkings equation was applied to decompose the near-field acoustic source into its monopole, dipole, and quadrupole contributions to assess glottal geometry effects on far-field sound. The results showed that dipole sources due to the unsteady forces exerted on the duct wall are dominant. [Work funded by NIH Grant RO1 DC03577.]

2pSC20. Modal analysis of a time-varying vocal tract with losses: A comparison of left and right eigenfunctions. Gordon J. Ramsay (Haskins Labs., 300 George St., New Haven, CT 06511-6624)

Sound propagation during speech can be modeled by a linear wave operator describing the spatiotemporal evolution of acoustic pressure and velocity distributions on a bounded domain representing the interior of the vocal tract. The eigenvalues of the wave operator determine the formant frequencies and bandwidths. The right eigenfunctions describe the standing wave patterns associated with each formant, whereas the left eigenfunctions describe how an arbitrary source distribution projects onto each formant. In speech research, it has always been implicitly assumed that left and right eigenfunctions are identical. This only holds if the wave operator is self-adjoint, which cannot be true when losses due to viscous damping, radiation effects, and wall vibration are taken into account. To examine this issue, a time-domain finite-volume simulation of acoustic wave propagation in the vocal tract was developed, including the above loss mechanisms. Eigenvalues and eigenfunctions were calculated from the resulting state-space recursion, for time-varying geometries representing VCV sequences containing fricatives and stops. Left and right eigenfunctions are similar in shape, but there are differences in the zero-crossing points and the amplitudes of the different lobes. Predictions based on right eigenfunctions alone will be incorrect. [Work supported by NIH.]

2pSC21. The interaction of linguistic and affective prosody in a tone language. Chinar Dara and Marc D. Pell (Commun. Sci. and Disord., McGill Univ., 1266 Pine Ave. West, Montreal, PQ H3G 1A8, Canada, chinar.dara@mail.mcgill.ca)

To address how a common set of acoustic properties of speech prosody modulate to convey linguistic and affective meanings concurrently, this study investigated the influence of phonemic tones on the expression of emotion (happy, sad, angry) and linguistic modality (declarative, interrogative) in a tone language, Punjabi. Base stimuli consisted of neutral sentences with either of the three contrastive tones in Punjabi (falling, level, high-rising) varying at the object position of the sentence only. Each of these sentences was elicited as a statement or a question for each of the three emotions by 6 native Punjabi speakers. Utterances were validated for adequate representation of the target emotion, and reliable exemplars for each of the emotions and linguistic modality were subjected to detailed acoustic analysis. Fundamental frequency and duration measures were analyzed for the stressed vowel of the keywords (subject, object, verb) and

the whole utterance for each of the tokens. Results demonstrated ways that the acoustic correlates of tone at the word level interact with that of the intonation and emotion at the sentence level, which were further compared with previous results on the effect of focus on intonation and emotional prosody [M. D. Pell, J. Acoust. Soc. Am. **109**, 1668–1680 (2001)].

2pSC22. The scope of effect of prosodic boundaries in articulation. Jelena Krivokapic (Dept. of Linguist., USC, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693)

An articulatory study of the scope of prosodic boundary effects across boundaries of varying strengths is presented. Sentences in each of three prosodic conditions contain the following string: C1VC2VC3#VC4VC5VC6, where C is an alveolar consonant, V is a vowel, and # is a prosodic boundary. The boundaries are of three degrees of strength. Using articulator movement-tracking data (EMA) for the tongue tip articulator, consonant constriction formation and release duration, acceleration duration, and spatial magnitude are measured. Leftward and rightward temporal effects of the boundary on the consonants are investigated. Based on earlier studies (e.g. Cambier-Langeveld, Linguistics in the Netherlands, 13–24, 1997; Shattuck-Hufnagel and Turk, Proceedings 16th International congress on Acoustics and 135th Meeting Acoustical Society of America, 1235–1236, 1998; Berkovits, J. of Phonetics **21**, 479–489, 1993) and predictions of the prosodic gestural model (Byrd and Saltzman, J. of Phonetics, **31**, 149–180, 2003), it is expected that (1) the amount of articulatory lengthening will increase with boundary strength and (2) the degree of lengthening will decrease with distance from the boundary. Three subjects participated. The results show that at the boundary, gestures have longer constriction and acceleration durations, and the effect of the boundary increases with boundary strength, distinguishing up to three levels of boundary strength. The effect is local and diminishes with distance from the boundary. These results support the initial hypotheses. [Work supported by NIH.]

2pSC23. Acoustic and fiberscopic studies of the voice characteristics of school teachers under vocal abuse conditions with the loud voices of children. Noriko Kobayashi (School of Allied Health Sci., Kitsato Univ., 1-15-1 Kitasato, Sagamihara 228-8555 Japan, noriko@ahs.kitasato-u.ac.jp, Takashi Masaki, and Koichiro Nishiyama (Kitasato Univ, Sagamihara 228-8555, Japan)

It has been known that the incidence of vocal pathology was high in school teachers as they had to use a great amount of voice and loud phonation in noisy school settings with the children's loud voices. A proper vocal hygiene program and voice therapy are effective treatments for these cases. The final goal of the treatment should be to obtain the skills to use effective and healthy phonation even in noisy environments. In our study, three school teachers with vocal nodules, three speech therapists who were trained to produce nonconstricted voice even in noisy environments, and ten college students with no laryngeal pathology (control group) spoke eight sentences with one target word in two environments: a quiet environment and one with meaningful multi-talker babble (MMB). Acoustic analyses and fiberscopic examination revealed higher sound pressure and F0 levels and laryngeal constriction in MMB for the teachers before therapy and the normal talkers, but two speech therapists kept similar F0 levels and laryngeal conditions in both environments. After voice therapy, teachers produced less constricted voice under noise. These results suggested the efficacy of voice therapy with MMB for teachers to obtain effective and healthy voice in their noisy work environments.

2pSC24. Realizing question intonation in Mandarin with neutral tone. Fang Liu (Dept. of Linguist., Univ. of Chicago, 1010 E. 59th St., Chicago, IL 60637, liufang@uchicago.edu) and Yi Xu (Univ. College London, London, NW1 2HE, UK)

Previous research has found that postfocus pitch range is lowered in Mandarin even in questions, although the amount of lowering is smaller than in statements. This study investigates whether postfocus pitch range is raised in questions where the postfocus words carry the neutral tone, which, being conventionally considered targetless, may be better suited for manifesting intonation. Eight native speakers of Mandarin each produced 80 statements and 80 questions. The sentences ended either with a sequence of 5 neutral tones or with 3 neutral tones followed by 2 high tones. The sentences had two alternative focus locations, one on the full tone (high, rising, low, or falling) immediately preceding the neutral tone sequence, and the other on the low tone before the full tone. Results show that (1) postfocus lowering occurred in the neutral-tone-final sentences in both statements and questions, just as in the full-tone-final sentences, and (2) sentence-final neutral tone had falling F0 even in questions, thus contrasting with sentence-final high tone, which had rising F0 in questions. The findings are interpreted as indication that the neutral tone, being associated with weak articulatory effort, is not as effective as full tones in realizing the statement/question contrast.

2pSC25. Acoustic and articulatory correlates of contrastive focus. Melanie Thibeault, Lucie Menard, Annie Leclerc, Anthony Calabrino (Laboratoire de phonetique, UQAM, Case postale 8888, succursale Ctr. ville, Montreal QC, Canada, H3C 3P8, thibeault.melanie.2@courrier.uqam.ca), and Annie Brasseur (UQAM, Montreal QC, Canada, H3C 3P8)

This paper aims at examining production of French speakers in marking focus-induced prominence at the articulatory and acoustic levels. The corpus consisted of CVC syllables, where C corresponds to one of the stops /p t k/ and V is one of the vowels /i a u/. Target words were embedded in carrier sentences elicited in two prosodic conditions: neutral (unfocused) and under contrastive focus. Four adult speakers (all native speakers of French) pronounced ten repetitions of each sequence. The audio signal and tongue shapes were recorded using a digital camera and a SONOSITE 180 ultrasound. Formant frequencies, rms values, duration, and tongue contours corresponding to each vowel were extracted. Analyses show that prosodic context has a significant effect on acoustic and articulatory data for all vowels, increasing F1 and F2 under contrastive focus, compared to the neutral context. However, this stable acoustic pattern is achieved by various articulatory strategies across subjects. For example, 2 subjects produced /u/ under focus with a lower tongue body than in the neutral context, whereas the remaining 2 subjects had a higher tongue body in focused syllables. Results are compared to previous studies on articulatory and acoustic correlates of prosodic structure in French and English.

2pSC26. Rhythm in English clear speech. Rajka Smiljanic, Josh Viau, and Ann Bradlow (Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, rajka@northwestern.edu)

This study investigates the effect of hyperarticulated, intelligibility-enhancing clear speech on English speech rhythm. Ramus *et al.* [Ramus *et al.*, "Correlates of linguistic rhythm in the speech signal," Cognition **72**, 265–292 (1999)] showed that temporal properties of a speech signal such as the percentage of vocalic intervals (%V) and variability of consonantal and vocalic intervals (C, V) can be related to phonological properties such as presence/absence of unstressed vowel reduction and syllable structure complexity, and are consequently quite successful at grouping languages into the traditional rhythmic classes (stress-, syllable-, and mora-timed). Here, we explore whether/how clear speech affects stress-timed characteristics of English sentences in terms of these measures. Results revealed that the proportion of vocalic intervals (%V) within sentences remained stable across speaking styles, i.e., consonants and vowels

were lengthened equally in clear speech. However, variability of both vocalic and consonantal intervals (V, C) was higher in clear than in conversational speech. The increase in V and C was achieved primarily through insertion/strengthening of short vocalic and consonantal segments that were dropped or coarticulated with surrounding sounds in conversational speech, making word and syllable boundaries more salient. These results suggest that increased intelligibility of clear speech can in part be attributed to prosodic structure enhancement by means of enhanced syllable and word boundaries demarcation.

2pSC27. Sensitivity of Mandarin speech to loudness-shifted voice feedback perturbation. Hanjun Liu, Qianru Zhang, Charles R. Larson (Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208), and Yi Xu (Univ. College London, London, NW1 2HE, UK)

Recent studies have demonstrated that loudness-shifted voice auditory feedback leads to compensatory adjustments in voice amplitude during production of vowels. In this study we asked the question of whether Mandarin-speaking subjects would respond to loudness-shifted voice feedback during speech. Twenty-two native-speaking subjects with normal hearing and neurological background served as subjects. They said the Mandarin phrase ba ma, with the high tone on both syllables, as answers to three randomized prerecorded questions. The questions induced focus on the first or the second syllable, or on both syllables. During each production a ± 3 dB, 200- or 400-ms perturbation in voice feedback was presented 250 ms following voice onset. Randomized control trials were intermixed with test trials. Digitized voice records were converted to an rms trace representing voice amplitude. Traces were aligned with stimulus onset for event-related averaging. Results demonstrated that subjects responded to the loudness perturbation with compensatory adjustments in voice amplitude. The data extend recent findings indicating that voice feedback is used to help control amplitude during speech. [This work was supported by NIH Grant No. DC006243-01A1.]

2pSC28. Gradient alternations and gradient attraction. Katherine Crosswhite (Dept. of Linguist., Rice Univ., 6100 Main St., Houston, TX 77005, crosswhi@rice.edu)

Many phonological alternations have been shown to be sensitive to lexical ''gradient attraction,'' defined by Burzio [''Surface-to-surface morphology: When your representations turn into constrainsts,'' Rutgers Optimality Archive #341-0999 (http://roa.rutgers.edu) (1998)] as follows: (A) The overall structure of a word W (in both phonological and semantic components) is influenced by that of other words in the lexicon to which W is independently similar, and which are thought of as ''attractors'' of W. (B) Attraction is stronger where independent similarity is greater. Since most strong attractors will be morphologically related to W, this results in a limitation on phonological alternations that would produce allomorphy. So far, gradient attraction has been studied for categorical alternations, such as stress placement in English (cf. compArable versus cOmparable). However, many alternations previously believed to be categorical are in fact gradient [as reviewed in, e.g., Port and Leary, ''Against formal phonology,'' Language **81**, 927–964 (2005)]. This study documents the influence of lexical attraction on two gradient phenomena of English: polysyllabic shortening and flapping. Attractor strength was varied in two manners. First, each base used in the study was paired with both derivational and inflectional derivatives (i.e., rose-roses-rosy for polysyllabic shortening, and wait-waiting-waiter for flapping), and second by analyzing the effect of between-item variation in semantic relatedness for the derivationally related forms (cf. ice-icy versus nose-nosy).

2pSC29. Resonance tuning in soprano singing and vocal tract shaping: Comparison of sung and spoken vowels. Shrikanth Narayanan, Erik Bresch (USC Viterbi School of Eng., 3710 S. McClintock Ave, RTH 320, Los Angeles, CA 90089, shri@sipi.usc.edu), Stephen Tobin, Dani Byrd, Krishna Nayak, and Jon Nielsen (USC, Los Angeles, CA 90089)

Singing at high pitches by sopranos involves complex orchestration of vocal-fold mechanisms and vocal-tract shaping. A specific phenomenon is the tuning of vocal-tract resonances to the fundamental at high pitches. Measurement of vocal-tract resonances has shown that this is particularly true for the first resonance [Joliveau *et al.* (2004)]. Higher resonances show some increase in their values perhaps reflecting incidental effects rather than active tuning. We investigate vocal-tract shaping in soprano singing using a real-time MRI technique [Narayanan *et al.* (2004)] with synchronized noise-cancelled audio recording. A professionally trained Western-classical singer spoke and sang the vowels /a, i, u, o, e/ preceded by the consonant /l/ over an ascending scale (from B-flat below middle C, rising two octaves). Acoustic analysis shows clear evidence of F1 tuning to F0 in the higher-octave singing of the vowels. The vocal-tract shape was distinctly different at the higher pitches compared to vowels spoken at normal pitch and those sung at the lower octave, with fairly large oral and pharyngeal cavities, created by a highly bunched and retracted tongue, and a raised larynx. At the highest pitch, the vocal-tract shape differences between the vowels are considerably lessened, although some vestigial canonical tongue shapes appear to remain. [Work supported by NIH, USC Annenberg Center.]

Session 2pSP

Signal Processing in Acoustics: Topics in Seismic Signal Processing

Alan W. Meyer, Cochair
665 Dunhill Dr., Danville, CA 94506

Max Deffenbaugh, Cochair
Exxon Mobile Research and Engineering Co., 1545 Route 22 East, Annandale, NJ 08801

Chair's Introduction—1:00

Invited Papers

1:05

2pSP1. Short-term non-poissonian temporal clustering of magnitude 4+ earthquakes in california and western nevada. John E. Ebel (Weston Observatory, Boston College, 381 Concord Rd., Weston, MA 02493, ebel@bc.edu), Daniel W. Chambers, Alan L. Kafka, and Jenny A. Baglivo (Boston College, Chestnut Hill, MA 02467)

The M4+ mainshocks throughout California and western Nevada from 1932 to 2004 show non-Poissonian temporal clustering over time periods of a few days. The short-term clustering is independent of the distance between earthquake epicenters. It implies that some of the M4+ mainshocks are mutually triggered by some unknown regional cause. In southern California, more short-term clustering is found for M4+ earthquakes east of the San Andreas Fault. In central California, most M4+ mainshocks at Long Valley, CA have occurred within 10 days of M4+ mainshocks around the San Francisco Bay area. The clustering implies predictable behavior in the occurrences of M4+ mainshocks. We propose a hidden Markov model (HMM) as an earthquake forecast method for the region. Our HMM assumes a hidden sequence of interevent time states associated with observations of earthquake occurrences (times, locations, and magnitudes) with transition probabilities between states determined with the Baum-Welch algorithm and the past earthquake data. Given the seismic history up to the latest earthquake, the probability of another earthquake within the next few days is estimated. Tests of our HMM with two, three, and four temporal states show some modest success. We plan to extend the model to forecast magnitude and spatial parameters.

1:25

2pSP2. Trends in detection in earthquake seismology. David B. Harris (Lawrence Livermore Natl. Lab., P.O. Box 808 L-205, Livermore, CA 94551, harris2@llnl.gov)

Detection methods in earthquake seismology have been dominated in the past by simple energy detectors, often combined with array processing (beamforming) to detect at lower thresholds. Increasingly, correlation detectors now are being employed to exploit the additional gain possible with coherent detection methods. Coherent detection of transient seismic signals is complicated by the fact that earthquake signals generally cannot be predicted with certainty in advance, except for those cases where observations of previous events are available. One approach to solving this problem is to employ subspace detectors that allow variation in the correlation template to match the range of signals anticipated from a specific source region. Subspace templates have been produced empirically from past observations of repeating sources. The current challenge is to compute the templates in areas without past event observations. One approach to doing so calculates the target signal space by forward simulation through many earth realizations drawn from a stochastic earth model. In this talk, we present the results of empirical subspace detection with arrays of sensors (combining array gain with coherent matched-signal gain) and discuss preliminary results from the use of computed subspaces. [This work was performed under the auspices of the U.S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48.]

1:45

2pSP3. Electromagnetic-to-seismic conversion: A new direct hydrocarbon indicator. A. H. Thompson (Retired, 13602 Peachwood Ct., Houston TX 77077)

In electroseismic (ES) exploration, electromagnetic (EM) energy converts to seismic energy at depth in the Earth. The amplitude of the converted seismic wave is large at porous, permeable, resistive bodies such as hydrocarbon reservoirs. ES is the only known exploration tool that, applied from the surface of the Earth, may discriminate between water, gas, and oil in the pores of reservoir rock. ES measures rock electrical properties with seismic resolution. The seismic conversion may be linear, at the same frequency as the EM stimulus. Electrokinetic effects, involving the relative motion of the rock solids and fluids, describe linear ES well. Nonlinear conversion may create all harmonics of the EM stimulus. Conventional descriptions of ES and SE conversion do not account for nonlinear conversions. New models, that couple EM and seismic energies through modulation of electrochemical potentials in the Earth, do explain the observations. ES field tests detected the hydrocarbon-sensitive, linear response to 1000 m and the nonlinear response to 1500 m deep. The results suggest that greater depth penetration, to 7000 m, is feasible.

2:05

2pSP4. Waveform design for electroseismic exploration. Scott C. Hornbostel, Arthur H. Thompson, and Warren S. Ross (ExxonMobil Upstream Res. Co., P.O. Box 2189, Houston, TX 77252-2189, scott.c.hornbostel@exxonmobil.com)

Electroseismic (ES) exploration remotely identifies the presence of hydrocarbons using the conversion of electromagnetic energy to seismic energy. These conversions are relatively larger in a porous, permeable resistive body (such as an oil or gas reservoir) when compared with background conversions. Typical ES signals, however, may be several orders of magnitude below the expected ambient noise levels, thus requiring repetitions of long source sequences. In addition, the ES signals may vary linearly or nonlinearly with the input current. Two classes of coded waveforms are presented for the detection of these linear and nonlinear conversions. The linear conversions are detected using a Golay sequence pair with side lobes that cancel when the correlated pair is summed. A set of pseudo-random binary sequences (PRBSs) can be modified to detect a nonlinear squared response with minimal side lobes while removing undesired linear conversions. These source waveforms are implemented in the power waveform synthesizer (PWS) that is capable of switching three-phase power to create the specified waveforms. The PWS is an important part of the overall electroseismic exploration system that includes source electrodes, pickup-free receivers, and acquisition and processing features aimed at effective detection of small ES signals.

Contributed Papers

2:25

2pSP5. Small-scale geophysics inversion using surface wave extracted from noise cross correlation. Pierre Gouedard, Philippe Roux, and Michel Campillo (LGIT—BP 53—38041 Grenoble cedex 9, France, pierre.gouedard@ujf-grenoble.fr)

Green's functions between receivers can be retrieved from the correlation of ambient noise or with an appropriate distribution of sources. We demonstrate this principle in geophysics at small scales. We used a 22.5-m-long line of 16 geophones and recorded noise signals in the bandwidth of 5 to 150 Hz. Noise was generated by human steps for a 10-min walk in the alignment of the geophones line (5 min on each side). The correlations of the records gave us the Green's functions associated with each pair of geophones, which exhibit two Rayleigh modes. We applied a frequency wave number transformation to separate those two modes and we measured the dispersion curve of each mode. We used these dispersion curves to perform an inversion and obtain the 1D velocity profile of the near surface (about 10 m deep).

2:40

2pSP6. Passive correlation imaging of a buried scatterer. Oleg I. Lobkis, Eric Larose, and Richard L. Weaver (Dept. of Theoretical and Appl. Mech., Univ. of Illinois, 104 S. Wright St., Urbana, IL 61801)

Waveforms obtained by correlation of diffuse fields are now widely used to estimate times of flight and ballistic Green's functions. Theory, however, predicts that such passively obtained waveforms should equal the full Green's function of a medium, and thereby convey information on the details of a structure and the presence of isolated scatterers. Here, we investigate diffuse field correlations in a reverberant elastic body and demonstrate that their shear and Rayleigh waves allow to image internal and surface features. Resolution in the resulting images is comparable to standard echographic and seismic images. The experimental setup is analogous to a real seismic experiment.

2:55–3:10 Break

Invited Papers

3:10

2pSP7. Anisotropic seismic waves in systems with fluid-saturated fractures. James Berryman (Lawrence Livermore Natl. Lab., P.O. Box 808 L-200, Livermore, CA 94551-9900, berryman1@llnl.gov)

The seismic wave speeds observed in exploration reflection surveys of oil and/or gas reservoirs may be dominated by the elastic properties in the dry state of very flat cracks or fractures, containing relatively little porosity. In contrast, the majority of work on effective medium theories (both estimates and bounds) has concentrated on systems whose elastic behavior depends on the volume fractions of the fluid constituents. Although a considerable amount of work has already been done in this area, there remain controversies regarding the use of the differential and self-consistent schemes for property estimation in these systems, and these controversies are known and well documented. When the reservoirs contain vertical fractures, these systems exhibit anisotropic seismic wave speeds. Methods are known that permit relatively simple calculation of these anisotropic properties from theoretical estimates made on randomly isotropic systems of fractures. When these methods are appropriately combined with weak anisotropy approximations associated with the Thomsen seismic-wave parameters, it is possible (at least in principle) to estimate fracture orientations and densities. The relationships between these forward and inverse problems will be treated and discussed with special emphasis on the question of uniqueness of the resulting interpretations.

3:30

2pSP8. Finite-frequency seismic tomography. Yang Shen (Grad. School of Oceanogr., Univ. of Rhode Island, Narragansett, RI 02882)

The use of 3-D sensitivity kernels in tomographic inversions has received much attention in the seismological community. Conventional tomographic studies are based on ray theory, which assumes that the arrival time of a body wave depends only on the velocity along the ray path between the source and receiver. Observed seismic waves are finite frequency signals, and their travel times are sensitive to a 3-D volume around the geometrical ray and subjected to wavefront healing, scattering, and other diffractive effects. As a result, seismic waves passing through velocity heterogeneities with dimensions smaller than the width of the Fresnel zone undergo significant wave-front healing, which results in reduced travel-time shifts compared to the values predicted by ray theory. 3-D sensitivity kernels take into account the effects of wave diffraction and scattering. This presentation will provide a brief review of

finite-frequency tomography, which is in a stage of rapid development. With the advancement of computer technology, it has become increasingly practical to construct 3D sensitivity kernels based on 3D reference models using full waveform simulations. This holds the promise of a new level of the integration and joint interpretation of finite-frequency waves propagating in complex media.

3:50

2pSP9. High-resolution imaging of fault zone structures. Zhigang Peng and Yehuda Ben-Zion (Dept. of Earth and Space Sci., Univ. of California, 595 Charles E. Young Dr. East, Los Angeles, CA 90095, zpeng@ess.ucla.edu)

Damaged fault zone (FZ) rocks with high crack density are expected to produce several indicative wave propagation signals. These include scattering, anisotropy, nonlinearity, and FZ guided head and trapped waves. These signals can be used to obtain high-resolution imaging of the FZ structures and to track possible temporal evolution of FZ properties. Results associated with systematic analysis of such signals recorded at several large strike-slip FZs can be summarized as follows: The observed FZ trapped waves are generated by relatively shallow structures that extend only over the top $\sim 3--4km$ of the crust. The shallow trapping structure in the North Anatolian fault (NAF) is surrounded by broader anisotropic and scattering zones that are also confined primarily to the top 3 km. Systematic analyses of anisotropy and scattering around the NAF do not show precursory temporal evolution of properties. However, the scattering results show clear co-seismic changes and post-seismic logarithmic recovery. These effects are likely to reflect mostly changes in properties of the shallow crust in response to strong shakings of nearby major earthquakes. Analyses of FZ head waves along the San Andreas Fault point to the existence of material interfaces that extend to the bottom of the seismogenic zone.

4:10

2pSP10. Multichannel inversion of scattered teleseismic waves for migration imaging and separation of noise. Stephane Rondenay and Xander Campman (MIT, 77 Massachusetts Ave., 54-512, Cambridge, MA 02109, rondenay@mit.edu)

In recent years, a substantial increase in the availability and fidelity of broadband seismographs has led to new opportunities for multichannel processing of scattered teleseismic body waves, allowing for high-resolution imaging of subsurface structure. Here, we discuss the theoretical foundation and the implementation of such a technique. The problem is posed for forward- and back-scattered wave fields generated at discontinuities in a 2-D isotropic medium, with the backprojection operator cast as a generalized radon transform (GRT). The approach allows for the treatment of incident plane waves from arbitrary backazimuths, and recovers estimates of perturbations in material property (P and S wave velocities) about a smoothly varying reference model. In its current implementation, the method can be applied to low-pass filtered data (0.3-Hz cutoff) for imaging structure in the lower crust and lithospheric mantle. Examples from the Cascadia and Alaska subduction zones are shown. Ongoing work is being conducted to render the method applicable to higher frequency data by identifying and canceling noise generated by near-surface scatterers (e.g., mountains, sedimentary basins).

Contributed Papers

4:30

2pSP11. Multichannel near-array corrections for short-period teleseismic data. Xander Campman and Stephane Rondenay (Earth Resources Lab., Massachusetts Inst. of Technol., 42 Carleton St., Cambridge, MA 02142, xander@erl.mit.edu)

The increased availability of data from broadband and short-period seismometer arrays has stimulated the development of multichannel high-resolution imaging algorithms for scattered teleseismic waves. One of the main factors limiting the resolution of teleseismic images is the interference from unwanted phases at higher frequencies (0.3–1.5 Hz). These unwanted phases are generally associated to scattering at near-surface features. We describe initial results from a multichannel inversion-based method designed to suppress scattered waves from surface topography. Such an approach is possible because the dense spatial sampling of short-period arrays allows us to sample these phases well enough to characterize them. Using a subset of the data, we first estimate a model for the distribution of near-surface scatterers. Next, we predict scattered waves in the entire data set using this scatterer distribution model. Both the inversion and prediction steps are based on an integral representation of the scattered-wave field which takes into account the near-field character of body-to-surface wave scattering. Finally, we subtract the predicted scattered waves from the entire data set. Data from the LARSE93 experiment are used to illustrate and validate the present algorithm.

4:45

2pSP12. An earthquake detection, identification, and location system for the northeastern U.S. Based on the wavelet transform. John E. Ebel (Weston Observatory, Boston College, 381 Concord Rd., Weston, MA 02493)

In order to optimize the ability of the New England Seismic Network (NESN) to detect all local earthquakes to the smallest possible magnitude, Weston Observatory of Boston College has developed an automated event detection and identification system based on the wavelet transform. This software system performs a wavelet transform on the data from each station being received and uses the time, frequency content, and energy of the first arrival, the highest energy arrival, and the end of the detection to compute the Bayesian probability that the detection was a distant earthquake (teleseism), regional earthquake, local earthquake, quarry blast, Rg wave from a quarry blast, or transient noise. At each station, these measured parameters are used to estimate the origin time, epicentral distance, and magnitude for each detection. The systems attempts to associate the detections from different stations that have a common event identification as a teleseism, regional/local earthquake, or quarry blast. If three or more stations associate, an automatic event location and magnitude (if a regional or local earthquake) are computed. This wavelet-transform detector and identifier has greatly increased the sensitivity of detecting and locating small earthquakes and quarry blasts in the New England region.

Session 2pUW

Underwater Acoustics: Scattering of Sound at the Sea Surface

Duncan P. Williams, Chair
DSTL, Physical Sciences, Winfrith Technology Centre, Dorchester, DT2 8WX, UK

Chair's Introduction—1:00

Invited Papers

1:05

2pUW1. Measurements and modeling of low-frequency air-sea-interface scattering. Roger C. Gauss, Joseph M. Fialkowski, and Daniel Wurmser (Naval Res. Lab., Code 7144, Washington, DC 20375-5350)

The acoustic scattering response of the undersea air-sea interface can be quite sensitive to ocean surface conditions, scattering geometry, and acoustic frequency. This paper presents both measurements and modeling of low-frequency (<1500 Hz) scattering strength. The modeling is based on a small slope approximation (SSA) that accounts for directional surface-wave roughness spectra with low-wave-number cutoffs. To provide an operationally useful predictive tool, a set of surface-wave spectral parameters (strength, exponent, and peak) are determined by nonlinear, multi-parameter fits to broadband, vertically bistatic surface backscattering strength data, leading to a semi-empirical bistatic model environmentally parameterized solely by the wind speed. The same SSA theory is also used to estimate the coherent-reflection loss given the frequency, angle, and wind speed. These broadband models are then used to explore the dependence of both incoherent and coherent sea-surface scattering on frequency, angle, and wind speed. [Work supported by ONR and PEO C4I & Space (PMW-180).]

1:25

2pUW2. How significant is the wave height of a moving surface on low-frequency propagation? Duncan P. Williams (Dstl. Winfrith, Winfrith Technol. Ctr., Rm. N11 Bldg. A32, Dorchester DT2 8WX, UK)

The ocean can carry low-frequency acoustic waves to surprisingly long range, but the ocean is ever changing and acoustic propagation is rarely unaffected by its interaction with the moving sea surface. The clarity of low-frequency acoustic communications at long range can be corrupted by this interaction. This is unsatisfactory for reliable communication and it is necessary to explain the effects introduced by the nature of the sea-surface scattering. This paper looks at the relationship between the power spectra of the received surface scattered signal and the sea surface. The surface can be conveniently characterized by its swell and wind-sea parts that can, in turn, be expressed as a function of their significant wave height and peak frequency. Basic properties of a surface spectral model are reviewed. A new model-based approach to computing the scattering from a moving sea surface is introduced. This approach is based on ray theory and a simple physical model for the response function of the surface. The model is compared with long-range experimental data and then results are shown to illustrate the effects of wave height and peak frequency on simple broadband waveforms.

1:45

2pUW3. Gravity wave focusing in the littoral zone. Grant Deane (Scripps Inst. of Oceanogr., Code 0238, UCSD, La Jolla, CA 93092-0238)

One of the unique environmental factors impacting acoustic propagation in the littoral zone is the presence of shoaling surf. Surface gravity waves act as acoustic mirrors, creating high-intensity cusps and caustics in the reflected sound field. The focal points of sound move through the water column as the gravity waves shoal, and can disrupt the performance of underwater communications systems and impact sonar performance. Analytical and numerical calculations of the focusing properties of shoaling gravity waves are described and used to predict the motion of the focal regions through time. [Work supported by ONR.]

2:05

2pUW4. A brief review of modeling in support of bubble-enhanced scattering from the air/sea interface. Richard Keiffer (NRL-SSC, Stennis Space Ctr., MS 39529)

By 1995 or so, the long-standing disparity (at low to moderate frequencies) between sea surface (backscatter) reverberation data and interface scattering models was solved by the incorporating a bubble cloud scattering mechanism into reverberation models. At that time (and still today), a comprehensive description of the spatially and temporally varying bubble population was lacking, so critical environmental parameters needed for the bubble cloud scattering models were inferred from the differences in the scattering levels predicted by interface scattering models and measured scattering strengths. This modeling development generally ignored another mechanism through which the bubbles may influence the reverberation: the bubbles will also cause upward refraction near the air/sea interface and this upward refraction may significantly increase the scattering from the air/sea boundary. In this presentation, a brief review of the published modeling supporting the bubble-enhanced interface scattering phenomenon will be presented. [Work supported by ONR/NRL.]

2:25

2pUW5. Wavefront modeling of scattering from surface waves. Chris Tindle (Phys. Dept., Univ. Auckland, Auckland, New Zealand)

The scattering of sound from any particular smooth realization of ocean surface waves can be calculated using wavefront modeling. The method uses ray tracing to locate the wavefronts and stationary phase analysis to determine the phase and amplitude of the sound field. Waveforms due to reflection of short pulses can be found directly and compare well with experimental results. Reflection at a curved surface leads to foci and caustics and the field is determined numerically from Pearcey and Airy functions, respectively. The method uses the time domain and is fast and accurate. [Work supported by ONR.]

2:45–3:00 Break

3:00

2pUW6. Time domain modeling of acoustic Doppler from 2-D, developing, and swell-contaminated seas displaying a bimodal azimuthal dependence. Richard Keiffer (NRL-SSC, Stennis Space Ctr., MS 39529)

The typical sea-surface acoustic scattering experiment is often modeled invoking an environment and measurement scenario containing a number of idealizations. Almost universal is the assumption that the rough pressure-release air/sea boundary can be described using a roughness spectrum that belongs to spatially uniform, fully developed seas. Often this roughness spectrum is assumed to have a simple cosine-squared or some other canonical azimuthal dependence. The air/sea boundary is assumed to be an overly homogeneous isospeed ocean and the source and receiver are assumed stationary or in uniform rectilinear motion. Because the data collected at sea and the results from numerical simulations are often compared in order to infer the acoustic attributes of bubbles and bubble clouds, it is important to have a good understanding of the limitations that these common idealizations impose on the data/model comparison. In this paper, a newly developed and tested time domain model for acoustic scattering from moving 2-D surfaces [J. Acoust. Soc. Am. **118**(3), 1283–1299 (2005)] is used to examine the implications that wave growth, swell contamination of a bimodal azimuthal dependence, and near-surface upward refraction have on the Doppler of acoustic signals backscattered from the sea surface. [Work supported by ONR/NRL.]

3:20

2pUW7. Modeling of glints due to surface curvature using the virtual source technique. Ahmad T. Abawi, Michael B. Porter, and Martin Siderius (Heat, Light, and Sound Research, 12730 High Bluff Dr., San Diego, CA 92130)

The virtual source technique, also known as the method of superposition or the auxiliary source technique, has been widely used to model scattering from targets. In this technique the target is replaced by a collection of sources of unknown complex amplitudes. The source amplitudes are calculated by applying the appropriate boundary conditions at the surface of the target. In this paper, the virtual source technique is used to compute scattering from rough surfaces. For sound scattering from the rough ocean surface, this is done by solving the wave equation for a half-space ocean overlaying a bottom and forcing the solution to satisfy the pressure-release boundary condition at a desired surface. The method is applied to study the effect of focusing of energy (glints) due to reflection of sound from a gently sloping, but concave ocean surface for various ocean conditions. Additionally, the method is applied to the study of scattering of waves from an infinite periodic surface.

Contributed Papers

3:40

2pUW8. High-frequency limit under the Rayleigh hypothesis. Tanos Elfouhaily and Thomas Hahn (Univ. of Miami, RSMAS, 4600 Rickenbacker Cswy., Miami, FL 33149)

The Rayleigh hypothesis (RH) is traditionally invoked to compute the scattering amplitude in rough surface scattering. This hypothesis simplifies the analytic derivation of the scattering problem and has been proven to be accurate for the determination of the Volterra series for the small perturbation method (SPM). The SPM development physically corresponds to the scattering of low-frequency acoustical waves. In this study we focus on the high-frequency limit under the Rayleigh hypothesis (RH). We validate the consistency of the RH with the high-frequency limit to the order invoked in the approximation. This means that the first-order geometrical optics (GO1) is reached even though the Rayleigh hypothesis consists of the summation of solely upward scattered fields.

3:55

2pUW9. Scattering by surfaces with sub-Bragg structure. Frank Henyey and Dajun Tang (Appl. Phys. Lab., Univ of Washington, 1013 NE 40th St., Seattle, WA 92105)

The scattering from a rough pressure release surface, where the spectrum of the surface only contains wavelengths considerably shorter than half the acoustic wavelength, is considered. The amplitude of the surface may be larger than that for which perturbation theory would be appropriate. We show that the surface can be replaced by an effective surface containing waves longer than half the acoustic wavelength. This effective surface is close to the envelope of the actual surface; this effective problem is likely to be much easier to solve than the original problem. We investigate the relationship between the problems and adjust parameters of the effective surface, by solving both problems with the integral equation.

4:10

2pUW10. Finite-difference time-domain simulation of acoustic propagation through a dispersive oceanic bubble cloud. Guy Norton (Naval Res. Lab., Acoust. Div., Stennis Space Ctr., MS 39529, guy.norton@nrlssc.navy.mil)

Accurate modeling of pulse propagation and scattering is a problem in many disciplines. For the case of an acoustic wave propagating in a two-dimensional non dispersive medium, a routine second-order in time and space finite-difference-time-domain (FDTD) scheme representation of the linear wave equation can be used to solve for the acoustic pressure. However, when the medium is dispersive, one is required to take into account the frequency-dependent attenuation and phase speed. Until recently, to include the dispersive effects one typically solved the problem in the frequency domain and then Fourier transformed into the time domain. However, by using a theory proposed by Blackstock [D. T. Blackstock, J. Acoust. Soc. Am. **77**, 2050 (1985)], the linear wave equation has been modified by adding an additional term that takes into account the disper-

sive nature of the medium. In this work a fourth-order (in time and space) 2-D FDTD scheme, which includes attenuation and dispersion via the convolution operator, is used to model backscattering of broadband signals from a dispersive composite subsurface bubble cloud composed of two different plumes underlying a random rough sea surface. [Work supported by ONR/NRL.]

4:25

2pUW11. A time domain approach to sea-surface scattering. Martin Siderius, Michael B. Porter, and Ahmad T. Abawi (HLS Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130)

Solutions to scattering problems are often formulated in the frequency domain which, for ocean acoustics, implies the surface is frozen in time. This may be reasonable for short-duration signals but breaks down if the surface changes appreciably over the time of the transmission. Additionally, frequency-do-main solutions are impractical for source-receiver ranges and frequency bands typical for applications such as acoustic communications (e.g., several kilometers, 3–50-kHz band). The time-varying nature of the sea surface adds to the complexity, often leading to a statistical description for the variations in received signals. A purely statistical description likely limits the insight that modeling generally provides. In this presentation we describe a time-domain modeling approach to the sea-surface scattering problem. The Helmholtz integral equation is used with the Kirchhoff approximation. The Green's functions representing the impulse response from all scattering points on the surface to the receivers are computed in terms of arrival amplitudes and time delays. This time-domain framework allows for including both broadband signals and the time-varying sea surface. The method will be compared with exact solutions for a subset of problems (frozen, rough surfaces) to determine limitations. Further, scattering results will be compared with those using Gaussian beam tracing.

4:40

2pUW12. Further work on the relationship between the coherent-to-incoherent ratio (CTIR) and the sea surface wave height for one-way propagation in a shallow ocean channel. R. Lee Culver, David L. Bradley (Appl. Res. Lab and Acoust. Grad. Program, The Penn State Univ., P.O. Box 30, State College, PA 16804, rlc5@psu.edu), Philippe Roux (LGIT-Maison des Gosciences, Grenoble cedex 9, France), William A. Kuperman (Univ. of California at San Diego, La Jolla, CA 92093-0238), and Stevenson Mark (NATO Undersea Res. Ctr., La Spezia, Italy)

A 29-element source array and 32-element receiver array have been deployed 8.6 km apart in a 120-m-deep ocean channel. Both arrays spanned most of the water column. Linear frequency-modulated (LFM) pulses 150 ms long and 4 kHz wide centered at 3.5 kHz were transmitted from one source at a time and recorded at all receivers during a several day period. During the recordings, the sound speed profile was approximately iso-velocity and the significant wave height varied from 0.3 to 1.1 m. We average the received signal coherently and incoherently for each source-receiver pair, average over all pairs, and take the coherent-to-incoherent ratio (CTIR). Log (CTIR) is found to depend linearly on time after the initial arrival, with slope proportional to the significant wave height. An earlier talk showed that the average reflection coefficient times the number of surface bounces N was a good predictor of the slope, but subsequently we find that it alone is not sufficient. Here we report a more complete theory that combines N, the average reflection coefficient, and the density of the arrivals. A predictive model for slope versus surface wave height is presented and compared with the measurements. The method can incorporate ray tracing and be extended to non-iso-velocity environments.

4:55

2pUW13. Detection in a nonstationary waveguide using orthogonal codes to measure the invariants of the time reversal operator. Franck Philippe, Claire Prada, Dominique Clorennec, Julien De Rosny (LOA, ESPCI, 10 rue Vauquelin, 75005 Paris, France), and Mathias Fink (LOA, 75005 Paris, France)

In underwater acoustics, time reversal (TR) techniques using an array of acoustic sources are commonly known for their efficiency in applications such as communication and detection. However, as TR methods assume stationary conditions, they are clearly affected by fluctuations of the medium. Among them, the decomposition of the time reversal operator (DORT method), although highly efficient for separating the echo from two close targets through a complex medium, suffers from the long acquisition time required. Indeed the whole interelement responses are needed meaning that N signals are acquired for an N-transducer array. The measurement is then at least N times longer than travel time. To reduce the acquisition time, orthogonal codes can be used [Folgot *et al.*, J. Acoust. Soc. Am. **117**, 3757–3765 (2005)]. These codes are tested in a laboratory detection experiment: a small-scale model of a shallow-water wave guide in which we add artificial surface waves of different frequencies and different amplitudes. The perturbations induced by the surface waves on the invariant of the time reversal operator are analyzed. Different methods which greatly reduce the acquisition time like Kasami and AIRS codes are compared.

Meeting of Accredited Standards Committee (ASC) S1 Acoustics

J. P. Seiler, Chair S1
U.S. Dept. of Labor, Mine Safety and Health Admin., P.O. Box 18233, Bldg. 38, Cochrans Mill Road, Pittsburgh, PA 15236

G. S. K. Wong, Vice Chair S1
Inst. for National Measurement Standards, National Research Council, Montreal Road, Bldg. M36, Ottawa, Ontario K1A OR6, Canada

Accredited Standards Committee S1 on Acoustics. Working group chairs will report on the status of standards currently under development in the areas of physical acoustics, electroacoustics, sonics, ultrasonics, and underwater sound, etc. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note – those meetings will be held in conjunction with the Standards Plenary meeting at 10:30 a.m. on Tuesday, 6 June 2006.

Scope of S1: Standards, specifications, methods of measurement and test, and terminology in the field of physical acoustics, including architectural acoustics, electroacoustics, sonics and ultrasonics, and underwater sound, but excluding those aspects which pertain to biological safety, tolerance and comfort.

Meeting of Accredited Standards Committee (ASC) S12 Noise

R. D. Hellweg, Chair S12
Hewlett Packard Co., Acoustics Lab., MR01-2/K15, 200 Forest St., Marlborough, MA 01752

W. J. Murphy, Vice Chair S12
NIOSH, 4676 Columbia Parkway, Mail Stop C27, Cincinnati, OH 45226

Accredited Standards Committee S12 on Noise. Working group chairs will report on the status of noise standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 43/SC 1 Noise, take note – that meeting will be held in conjunction with the Standards Plenary meeting at 10:30 a.m. on Tuesday, 6 June 2006.

Scope of S12: Standards, specifications and terminology in the field of acoustical noise pertaining to methods of measurement, evaluation and control, including biological safety, tolerance and comfort and physical acoustics as related to environmental and occupational noise.

Session 3aAA

Architectural Acoustics, Musical Acoustics, and Engineering Acoustics: Composed Spaces

Alexander U. Case, Chair
Fermata Audio + Acoustics, P.O. Box 1161, Portsmouth, NH 03802

Chair's Introduction—7:55

Invited Papers

8:00

3aAA1. "Flowers and Wreaths": Radio theater and surround. Ean White (15 Channel Ctr. St., #514, Boston, MA 02210-3426, ean@incendiaryarts.org)

Originally produced in late 1992, ''Flowers and Wreaths'' is a setting for poems by Jacques Prvert (*paroles* Prvert) in French and in English translations by Lawrence Ferlinghetti and me. Prvert's success in France after World War II was largely due to his unabashed use of street language and his staunch antiwar, antirational political stance. The performance featured much original instrumentation: an amplified skull-resonator dental floss violone, a 35-ft.-long infrasonic electric monochord, a naval signal lamp, mechanized moving audience seating, the amplified electrocution of vegetables, unconventional speaker placement, etc. Mari Novotny-Jones and I continued to perform the piece through 1993. ''Flowers and Wreaths'' was decidedly a kind of live radio theater, but far too contingent, too phenomenological to expect that it could actually be recorded and distributed in any satisfactory way. Nevertheless, this work continued to be pursued through the mid-nineties, finally producing *Radio Etudes* (with Sarah Hickler and technical help from the late Robert Moog), which addressed directly the issues of reproduction before being given up entirely. Now, on the brink of a new era in broadcast technology, imagine a radio theater renaissance.

8:20

3aAA2. "Passing Landscapes." James Moses (Music Dept., Brown Univ., Orwig Music Bldg., 1 Young Orchid Ave., Providence, RI 02806)

This 15-min piece is composed from five-channel surround sound environmental and naturalistic concert hall recordings of original music. Cross-synthesis and resonant filtering techniques are used to create harmonic and melodic content from the environmental recordings. The synthesized components move in the surround soundfield against the stable spatial image of the source recordings. The environmental recordings include bucolic settings, a downtown cityscape, a highway traffic scene, a public gathering, and a composed ambience/montage of sounds from our electronic media culture. The various environmental recordings are juxtaposed with each other and with concert hall material to create the form of the piece. The piece intends to provide a contextualized view of aspects of modern life and technology.

8:40

3aAA3. Vis-a-vis: Reading texts in real-time. Joseph Rovan (Music Dept., Brown Univ., 1 Young Orchard Ave., Providence, RI 02912, rovan@brown.edu)

This presentation addresses the formal and psychological aspects of Vis-a-vis as a way of framing the attitude of the technology itself. Based on a text by Rilke, the work situates a solo singer within a dense sonic landscape whose electronic sounds are based entirely on her voice. The interactive score, written in MaxMSP, makes the computer a virtual performer, listening and responding to the changing musical events in its own terms. In this new 5.1 mix a variety of real-time processes are used on the voice, shifting from scene to scene, according to the tone of the text. Among the real-time processes used are spectral analysis/resynthesis, granular sampling, harmonization, frequency shifting, and envelope tracking, among others. In Vis-a-vis the computer functions like a virtual performer, listening to the changing musical events phrasing, dynamics, and register of the singer, and responds accordingly, in its own terms. The point is to capture something of the improvisatory spirit that is part of any live performance, and to create an environment in which the singer and computer can interact as a duo, responding to each other's musical decisions in an immersive sonic environment.

9:00

3aAA4. Convergence. Andrew Carballeira (Driftwood Audio, 311 Concord Ave., Lexington, MA 02421)

convergence seeks to show the potential beauty of a world at peace through the manipulation of dissonance/consonance and spatial information. As a noncomposed experimental piece, the principal task is sifting through a wealth of improvised free music to gleam the sections most pertinent to the aims of the work. Thereby, convergence is, in many ways, a continuation of the concept adopted by Columbia editors on Miles Davis later works (''Bitches Brew'' and ''Live/Evil'' among them) which places the engineer/editor at the forefront of the creative process. In the spatial domain, this work draws heavily on the multi-channel spatialization techniques developed by IRCAM (*et al.*) and convolution with impulse responses, sampled and synthesized, to create the environment in which sound sources interact and ultimately join together with one voice.

9:20

3aAA5. Sparrows swarm and sing—a sonic collaboration. Eli D. Cohn, Robert M. Mallory, and Alexander U. Case (Sound Recording Tech., Dept. of Music, UMass Lowell, 35 Wilder St., Ste. 3, Lowell, MA 01854, alex_case@uml.edu)

Untitled Number 3 was written and performed by the Massachusetts-based sextet, Sparrows Swarm & Sing. The work combines orchestral string instruments and percussion, distorted guitars, rock drums, voices without words, and an array of handmade noisemakers. The first movement, ''Quartet'' (recorded live), begins with underlying chimes and noisemakers present throughout the piece. Acoustic cello and violin quietly emerge from the disciplined electric guitar line. Harmony and dissonance evolve into a lush reverberant noise created from each note played in the preceding string passage. The second movement, ''Come on Camaraderie,'' emerges from the reverberant decay into an uplifting dance song of which the climax is highlighted with a choral vocal line. This immediately propels itself into the brooding and relentless noise of guitar melodies/detunings. The quiet vocal melody introduces the third movement, ''The Processional.'' Electric guitar and bass beckon a distant cello melody followed by layers of bowed glockenspiel, cello/violin pizzicato, chimes, and snare drum.

9:40

3aAA6. Improvising with spaces. Pauline Oliveros (Rensselaer Polytechnic Inst., 110 Federal St., Troy, NY 12180, paulineo@deeplistening.org)

This paper explores qualitative changes that occur in voices and instruments in relationships with changing spaces that ordinarily are held in a stationary paradigm of performance practice, spatial transformations, and the effect on sounds in multi-channel speaker systems. Digital technology allows one to compose and improvise with acoustical characteristics and change the apparent space during a musical performance. Sounds can move in space, and space can morph and change, affecting the sounds. Space is an integral part of sound. One cannot exist without the other. Varieties of sounds and spaces combine in symbiotic relationships that range from very limited to very powerful for the interweaving expressions of the music, architectures, and audiences.

10:00–10:10 Break

10:10

3aAA7. The creation of a surreal five-channel sound environment using human mobile microphone stands. Thomas Plsek (Berklee College of Music, MS-1140 Brass, 1140 Boyslton St., Boston, MA 02215, tplsek@berklee.edu)

The notion of creating a piece for a discrete five-channel audio environment calls for one to consider ways in which a piece could be created that could only exist in this environment, i.e., it would not be a documentation of a performance, nor could it even exist in the real acoustic world. Considering low-tech solutions to these problems led to the idea of doing a recording of solo trombone using human mobile microphone stands (or HMMS). This allows for real-time movement of both the source and the receivers during the recording process. This gives the possibility of creating a surreal audio environment that could only be created with extreme difficulty (if at all) any other way. As source material for composition, the poem, ''Rather the flight of the bird passing and leaving no trace,'' by Alberto Caeiro, who was one of many heteronyms of the Portuguese poet Fernando Pessoa, was chosen. Espcially notable is the line, ''What's past is nothing and remembering is not seeing.''

10:30

3aAA8. Temporal/spatial convergence in surround sound. Stephan Moore (Troy, NY 12180, stephan@oddnoise.com)

A discussion of the musical composition ''Thaw/Twist at Bailey's,'' in which a field recording of the 2004 spring thaw at Thomas Bailey's farmhouse in Richland, MI is used to explore the phenomena of temporal/spatial convergence and dispersal. This presentation will cover the creation of the piece, its initial installation for a site-specific 16-channel array of custom-built speakers, and the challenges and merits of adapting it for 5.1 surround sound.

10:50

3aAA9. The evolution of new counterpoint through electro-acoustic music. Eric Chasalow (Music Dept., Brandeis Univ., MS051, Waltham, MA 02454-9110)

It is widely believed that the advent of electronic music during the mid-20th century redefined music fundamentally. It may be shown, however, that a number of composers found rather more evolutionary ways of elevating basic acoustical elements to enrich the long tradition of music making. Examples from Davidovsky and the author demonstrate this history. Additional examples demonstrate the continuing evolution enabled by historical artifacts and surround technology.

11:10

3aAA10. Issues pertaining to the composing of "minimalist" loudspeaker music for home theater arrays. David Moulton (Moulton Labs., 39 Ames Rd., Groton, MA 01450, dmoulton@moultonlabs.com)

Loudspeaker music composed for home theater loudspeaker arrays is subject to a different set of aesthetic principles and production problems than is music (live or otherwise) prepared for concert presentation. Issues of scale, timing, dynamic range, spectral range, sonic materials, reverberance, the treatment of spatial issues, and listener expectations will all be discussed.

11:30

3aAA11. Spatial treatment of sound sources in John Mallia's "Anastasis." John Mallia (New England Conservatory of Music, 290 Huntington Ave., Boston, MA 02115)

This presentation will outline the spatial techniques and treatment of sounds in my five-channel acousmatic composition entitled ''Anastasis.'' The orchestration of sound objects in the field of five loudspeakers adheres to an elastic gravitational plan assigned to each point in space. Generally speaking, the field of five speakers is weighted so that rear point sources support sounds lower in pitch than those occurring in the front of the projection plane where frequencies are raised. The relationship between each point source and its corresponding interval of transposition is in constant flux and its size serves to emphasize the degree of spatial depth present at any given moment. No attempt is made to precisely emulate the way in which we hear sounds move in the real world, but, rather, sound sources are intuitively, and experimentally, decorrelated to support and exaggerate the listeners experience of space. The presentation of this working method will raise questions regarding the compositional effectiveness of experimental spatialization techniques informed by general observations of the physical properties of sound versus methods that attempt to model, more closely and accurately, the predictable behavior of sound in space.

11:50

3aAA12. Different space conceptions in two electroacoustic compositions: Rojo and human-space-factory. Hans Tutschku (Harvard Univ., Music Bldg., Cambridge, MA 02138, tutschku@fas.harvard.edu)

Having used multichannel speaker setups for my electroacoustic compositions since 1989, I have experimented with many approaches to compose different sound-space relationships over the years. In human-space-factory, composed in 1999, the movements of the sounds in space are directly correlated with spectromorphologic changes of the sounds themselves. The composition tends to create distinguishable different spaces, sometimes even simultaneously. Rojo, composed in 2004, evokes the aural image of rituals of different cultures. Voices come from all directions and create a musical happening, surrounding the audience to give the impression that the listener is actually participating in the performance. The paper will outline the different compositional approaches and some technical solutions encountered during the realization of both compositions. Sound examples will be played during the lecture.

WEDNESDAY MORNING, 7 JUNE 2006 ROOM 550AB, 8:25 A.M. TO 12:00 NOON

Session 3aAB

Animal Bioacoustics, Signal Processing in Acoustics, and Engineering Acoustics: Mechanisms of Biosonar

James A. Simmons, Chair
Brown Univ., Neuroscience, Box 1953, Providence, RI 0291

Chair's Introduction—8:25

Invited Papers

8:30

3aAB1. The echolocating bat controls the direction and distance of its acoustic gaze. Cynthia Moss and Kaushik Ghose (Dept. of Psych., Inst. for Systems Res., Univ. of Maryland, College Park, MD 20742)

Adaptive vocal control by echolocating bats may serve to spatially filter acoustic information in a complex environment. It is well documented that FM bats actively adjust the duration of sonar vocalizations to avoid overlap of emissions and echoes. FM bats also control the direction of the sonar beam, which restricts the direction and volume of space sampled by the sonar system. Therefore, the temporal characteristics of the bat's sonar signals can be used to estimate its window of attention along the range axis, and the sonar beam aim as an index of the direction of its acoustic gaze. Here, we describe a series of studies that investigated adaptive sonar behavior of the big brown bat performing in a variety of echolocation tasks. We used high-speed infrared video to reconstruct the flight path of the bat and its insect prey. An array of ultrasonic microphones recorded the timing and spatial pattern of each sonar emission. Quantitative analyses of the data suggest that the bat's directional aim and temporal control of sonar signals may reduce information processing demands on its auditory system, enabling adaptive behaviors that must occur rapidly as the animal flies. [Work supported by NIMH and NIBIB.]

8:45

3aAB2. Strategies for bat echolocation estimated through characteristics of pulse and echo measured by a telemetry microphone. Hiroshi Riquimaroux (Doshisha Univ., 1-3 Tatara Miyakotani, Kyotanabe, Kyoto, Japan, hrikimar@mail.doshisha.ac.jp)

How echolocating bats extract important information from multiple targets during flight was investigated by a small and very light telemetry microphone system (Telemike) mounted on the head. A high-speed video camera system was also adopted together with Telemike to trace positions in space of the flying bat for analyzing spatio-temporal pulse emission patterns and flying velocity. With those devices, how the bats executed parallel time-sharing real-time processing during their flight could be examined. Taiwanese

leaf-nosed bats (Hipposideros terasensis), CF-FM bats, were used as subjects. The Doppler-shift compensation was confirmed during flight where the frequency of the dominant second harmonic of constant frequency component, CF2, of returning echoes was kept constant by adjusting CF2 frequency of emitting pulses. Estimated frequencies of CF2 of pulses from the bats flight speed strongly agreed with actually observed values. Pulse frequencies were also estimated using echoes returning directly ahead of the bat and from its sides for two different flight conditions: landing and U-turning. Temporal patterns for Doppler-shift compensation, interpulse interval, and echo delay showed that bats in flight alternately changed their attended direction from one to another. Bats were also observed to change the intensity and emission rate of pulses depending on the distance from the landing site. [Work partly supported by the Innovative Cluster Creation Project promoted by MEXT and by a grant to RCAST at Doshisha University from MEXT.]

9:00

3aAB3. Perception of object features by echolocating big brown bats. Caroline M. DeLong, Rebecca Bragg, and James A. Simmons (Dept. of Neurosci., Brown Univ., Box 1953, Providence, RI 02912, caroline_delong@brown.edu)

Big brown bats can discriminate among objects using wideband FM sonar sounds. They can perceive only the acoustic dimensions of echoes (e.g., delay, frequency, amplitude), or they can use those dimensions to reconstruct object features (e.g., distance, shape, size). To investigate this question, bats were presented with two-alternative (left/right) forced-choice sonar discrimination tasks. In experiment 1, the stimuli were a one-cylinder monopole target and a two-cylinder dipole target that were presented at all aspect angles. In experiment 2, the stimuli were multiple two-cylinder dipole targets presented at all aspect angles. In both experiments, the positive target was the dipole with a 5-cm separation (negative targets had separations of 1–7 cm). Acoustic dimensions of the dipole echoes change depending on the aspect angle. If the bats represent only acoustic dimensions, they should have difficulty selecting the positive target at many aspects. If the bats represent object features, they should be able to select the positive target independent of aspect angle. Bats perform well, suggesting that they reconstruct object features from echo acoustic dimensions. Errors appear due to masking of either the monopole glint with the dipole glints or the dipole glints with each other. [Work supported by NIH and ONR.]

9:15

3aAB4. How well does a whale hear its own biosonar pulses? Alexander Ya. Supin (Inst. of Ecology and Evolution of the Russian Acad. of Sci., 33 Leninsky Prospect, Moscow, Russia), Paul E. Nachtigall, and Marlee Breese (Univ. of Hawaii, HI)

During echolocation, whales and dolphins perceive faint echoes shortly after emitting loud outgoing sound pulses. Why do the outgoing sounds not mask the faint echoes? Is it because the sound energy is effectively concentrated in the transmitted sonar beam and isolated from the animal's own ears? To answer this question, we recorded a false killer whale's emitted echolocation pulses and auditory-evoked potentials provoked by its pulses and compared them to similar responses produced by external acoustic signals of known intensities. A few thousand of the individual AEP records were sorted according to the spontaneously varied amplitude of the synchronously recorded biosonar pulses. In each of the sorting bins (in 5-dB steps), AEP records were averaged to extract AEPs from the noise. For comparison, AEPs were recorded to external sound pulses with waveforms, spectra, and amplitudes similar to those of the whale's biosonar pulses. A comparison has shown that, depending on the target presence or absence, the sensitivity of the whale's hearing to its own transmitted biosonar pulse was 30 to 45 dB lower than might be expected in a free acoustic field

9:30

3aAB5. Detection and discrimination of fish prey by echolocating odontocetes. Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734), Kelly J. Benoit-Bird (Oregon State Univ., Corvallis, OR 97331) and Ronald Kastelein SEAMARCO, 3843 CC Harderwijk, the Netherlands)

The echolocation detection range of the Atlantic bottlenose dolphin *Tursiops truncates* and the harbor porpoise *Phocoena phocoena* have been either measured or estimated in nonmasking and noise-masking environments using human-made targets. These experiments provide little information on the detection range for fish prey in a natural environment. Backscatter and corresponding target strength measurements were performed on four species of fish, Atlantic cod, gray mullet, pollock, and sea bass, using simulated dolphin and harbor porpoise echolocation signals. Backscatter measurements were also performed on several species of snappers using simulated bottlenose dolphin echolocation signals. The fish subject was anesthetized and placed in a monofilament net bag which was attached to a monofilament curtain attached to a rotor. The target strength and backscatter characteristics were determined as a function of angle in the lateral plane. The results of detection experiments using real targets were used to estimate echolocation detection range for the different fish species. In a nonmasking environment, a bottlenose dolphin should be able to detect a cod between 84 and 135 m and a porpoise detection range should be between 12 and 23 m. We also examined the cues echolocating animals could use to discriminate fish species.

9:45

3aAB6. Numerical insights into the mechanisms of biosonar beamforming. Rolf Müller (School of Phys. & Microelectronics, Shandong Univ., 5 Hongjia Lou, 250100 Jinan, China)

The pronounced interspecific variability in the outer ear and nose shapes of bats suggests that the beamforming action of these structures is functionally important to the animals' biosonar systems. Numerical methods bring three major advantages to the study of the underlying mechanisms: efficiency in producing high-resolution directivity estimates, exploration of wave properties in the near field, and integration with digital shape manipulations. High-resolution directivity estimates reveal the shape of main- and sidelobes as well as their interplay over frequency. In two case studies, sidelobes were found to be prominent directivity features with significant sensitivity. In one case, they were even part of a dynamic pattern in which the maximum sensitivity alternates between mainlobe and sidelobes over frequency. Shape manipulations can be applied to the study of such features as the tragus or antitragus and prominent ledges around the pinna rim. These features can either be digitally removed or continuously transformed (e.g., rotated or scaled). The effects of multiple features can be studied by generating all combinations of presence or absence of the individual features. All prominent morphological features studied so far were found to have an effect mainly on sidelobes.

10:00–10:15 Break

10:15

3aAB7. Cepstral processing for bat biosonar. John R. Buck (Dept. of Elec. and Comput. Eng. & School for Marine Sci. and Technol., Univ. of Massachusetts Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747), James A. Simmons, and Caroline M. DeLong (Brown Univ., Providence, RI 02912)

Bats routinely discriminate multiple glint targets whose interglint time spacings are much less than the integration time of the bat auditory system. This suggests that the bats are exploiting spectral cues in the received echos to resolve the glints. Closely spaced glints cause notches and peaks in the received signal spectrum with frequency spacings characteristic of the glint separations. A biologically inspired signal processing algorithm for resolving these echos into target images is described. The algorithm combines a coarse energy detector to locate targets with a windowed short-time cepstral analysis, or cepstrogram, to resolve the fine structure of the detected targets. The algorithm's performance is demonstrated using data recorded in a laboratory experiment. The experiments transmit an ultrasonic chirp and record the echos from a target consisting of a pair of cylinders spaced 5 cm apart, presented at aspect angles ranging from 0 to 90 deg. These recordings are processed with the cepstrogram algorithm using parameter choices commensurate with the performance of the bat auditory system. The cepstrogram algorithm accurately estimates the spacing of the cylinders. Combining the results from the left and right receiver channels produces a binaural image of the target. [Work supported by ONR.]

10:30

3aAB8. Evaluation of performance of the echolocation model for the restoration of an acoustic image from a single-emission echo. Ikuo Matsuo (Faculty of Liberal Arts, Tohoku Gakuin Univ., 2-1-1 Tenjinzawa, Izumi-ku, Sendai 981-3193, Japan), Takashi Magara, and Masafumi Yano (Tohoku Univ., Sendai 980-8577, Japan)

Using frequency-modulated echolocation sound, bats can form a fine acoustic image of an object. The acoustic image is an impulse response, known as a reflected-intensity distribution, which is composed of amplitude and phase spectra over a range of frequencies. However, bats detect only the amplitude spectrum due to the low-time resolution of their peripheral auditory system, and the frequency range of emission is restricted. The amplitude spectrum varies with the changes in the configuration of the reflected-intensity distribution, while the phase spectrum varies with the changes in its configuration and location. By introducing some reasonable constraints, we previously proposed a method for restoring an acoustic image from the echo produced by a single emission. The configuration is extrapolated from the amplitude spectrum of the restricted frequency range by using the continuity condition of the amplitude spectrum at the minimum frequency of the emission and the minimum phase condition. The location is estimated from the temporal changes of the amplitude spectra computed by the convolution of the Gaussian chirplet with a carrier frequency compatible with bat emission sweep rates. In this paper, the performance of this model is evaluated by measuring and analyzing the echo reflected from the object.

10:45

3aAB9. A computational model of bottlenose dolphin sonar: Feature-extracting method. Tengiz V. Zorikov (Inst. of Cybernetics, 5 S. Euli, 0186, Tbilisi, Georgia, zorikov@hotmail.com)

A process of echo-image formation in bottlenose dolphin perception was investigated in our experiments, in which animals were trained to differentiate passive actual echoes recorded beforehand purposely, as well as echo-like simulated impulses. The results suggest that dolphin, processing echoes within critical interval of time ($\sim$0.2 ms), utilizes the string of three independent, hierarchically organized features (echoes subjective characteristics) being defined by different scales of spectral density oscillation, and energy. Herewith, dolphin is capable to assess features average value over a series of echoes. Distinguishing stimuli, dolphin estimates successively features values from senior to minor, terminating the process at the feature, which contains detectable differences in compared stimuli (the distinctive feature). A procedure of subsequent identification of the reference stimulus can be described by the following decision rule: If the dolphin utilizes some feature as the distinctive one, then in order to preserve the image of the reference stimulus in the animal's perception, it is necessary and sufficient to preserve the same values of the distinctive feature and all higher ones in order of hierarchy. These data were formalized mathematically and used in our computational model, comparative testing of which has revealed critical capabilities comparable with those of bottlenose dolphins.

11:00

3aAB10. A spike burst coding scheme for environment representation by FM bats. Herbert Peremans and Bertrand Fontaine (Dept. Environment, Tech. and Tech. Management, Universiteit Antwerpen, Prinsstraat, 13, 2000 Antwerpen, Belgium, herbert.peremans@ua.ac.be)

All environmental features of interest to bats, i.e., spatial arrangement of reflectors as well as reflector type, are coded into the patterns of spikes that travel up the auditory nerve into the bat's brain. Realistic scenes containing multiple, closely spaced reflectors give rise to complex echo signals consisting of multiple, highly overlapping filtered copies of the bat's own vocalization. Some of this filtering is self-induced, i.e., directivity of the bat sonar system, and some of it is beyond the control of the bat, i.e., sound absorption and reflection process. This analysis concentrates on the conspicuous ridges these filter operations give rise to in the time-frequency representation of the echo as produced by the FM bat's inner ear. Assuming multiple threshold detecting neurons for each frequency channel, it is shown how the time dispersion of the generated spike bursts is linked to the presence and characteristics of these notches. A neural network decoding the spike bursts in terms of notch locations and types is described. After separating the notches due to the HRTF from the reflector generated ones, they are separately interpreted in terms of the environmental features of interest, i.e., target location and target structure. [Work funded by the EC, IST-FET Initiative.]

11:15

3aAB11. Simulating neural responses to biosonar signals. David C. Mountain (Boston Univ. Hearing Res. Ctr., 44 Cummington St., Boston, MA 02215)

The mammalian auditory system is a highly evolved acoustic signal-processing system that performs well even in highly reverberant and cluttered acoustic environments. In echolocating species, the auditory system is even more highly evolved and is generally more important than vision for navigation and foraging. Since invasive physiological experiments using echolocating species is challenging at best and is illegal for protected species, other methods must be sought if we are to understand how biosonar systems function. Biophysically based computer models provide a method by which we can extrapolate from physiological experiments performed on species that can be used for invasive experiments to those species that cannot be used. Results will be presented for simulations of physiological responses in the auditory nerve and brainstem to biosonar signals for several species. The simulations were done using the EARLAB (http://earlab.bu.edu) desktop simulation environment. Model parameters for species of interest were estimated from behavioral audiograms and from other available data. The models can be used to predict how different types of biosonar signals are represented in neural firing patterns and how the neural representation degrades in the presence of anthropogenic noise. [Work supported by ONR and NSF through the NOPP program.]

Contributed Papers

11:30

3aAB12. Bat sonar and the role of frequency diversity. Kenneth G. Foote (Woods Hole Oceanogr. Inst., 98 Water St., Woods Hole, MA 02543) and James A. Simmons (Brown Univ., Providence, RI 02912)

Species of echolocating bats that forage in foliage, in the presence of clutter with attendant multiple, often overlapping echoes, use 3–5 simultaneous harmonic sweeps to cover a broad bandwidth, whereas species of bats that hunt only in the open use a single sweep to cover the same broad bandwidth. It is hypothesized that the frequency diversity evident in the presence of several harmonically related frequencies is vital to success in clutter. It is additionally hypothesized that the bat processes harmonic FM echoes in a manner tantamount to split-beam processing. Simultaneous determination of phase at several related frequencies enables the bat to distinguish between single and multiple targets, as in the theory of coincidence echoes [K. G. Foote, J. Acoust. Soc. Am. **99**, 266–271 (1996)]. The plausibility of the second hypothesis is gauged by computations based on empirical knowledge of the signal parameters and timing acuity of the big brown bat (*Eptesicus fuscus*).

11:45

3aAB13. Echolocating big brown bats shorten interpulse intervals when flying in high-clutter environments. Anthony E. Petrites (Ecology & Evolutionary Biol., Brown Univ., 80 Waterman St., Box G-W, Providence, RI, 02912), Donald S. Mowlds, Oliver S. Eng, James A. Simmons, and Caroline M. DeLong (Brown Univ., Providence, RI, 02912)

Insectivorous big brown bats (*Eptesicus fuscus*) use frequency-modulated ultrasonic echolocation calls to locate and capture prey, often while navigating through highly cluttered areas of vegetation. To test how their calls change while flying through different clutter densities, a matrix of vertically hanging chain links was constructed in a 4.5-m-wide, 10.5-m-long, and 2.6-m-high flight room. Three different clutter densities (low, medium, and high) were created by varying the number of chains in the matrix (9, 114, and 150, respectively). Four wild-born bats were trained to fly through curved gaps in the chain network. These flights were recorded in the dark with a stereoscopic pair of thermal imaging cameras and a heterodyne bat detector. The bats were flown first through the high-, then the low-, and finally the medium-density configuration over a period of 40 days. Preliminary analysis of the pulse intervals of the bats sounds during flight reveals that the interpulse intervals shorten considerably under high-clutter conditions as compared to medium or low clutter. The data suggest that bats flying through high clutter become limited to shorter interpulse intervals in order to ensonify their immediate environment at the expense of larger-scale navigation. [Work supported by NIH.]

WEDNESDAY MORNING, 7 JUNE 2006 BALLROOM D, 8:45 TO 10:00 A.M.

Session 3aAO

Acoustical Oceanography: Acoustical Oceanography Prize Lecture

James F. Lynch, Chair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Chair's Introduction—8:45

Invited Paper

9:00

3aAO1. Acoustic species identification: When biology collides with physics. John K. Horne (UW School of Aquatic and Fishery Sci., P.O. Box 355020, Seattle, WA 98195)

Acoustic species identification remains a long-term goal of fishers, researchers, and resource managers that use sound to locate, map, and count aquatic organisms. Interpreting acoustic data from aquatic organisms requires an understanding of how pressure waves interact with biological targets to produce single or multiple echoes. Intensity and variability of backscattered energy is influenced by physical factors associated with energy propagation through a dense medium, and biological factors including the morphology,

ontogeny, physiology, and behavior of targets. Efforts to quantify relative magnitudes and importance of biological factors have resulted in an approach that integrates field and laboratory measurements, numeric backscatter models, and computer visualizations. Acoustic visualizations are used to incorporate representations of the acoustic beam; the shape and composition of targets; target orientation (tilt, yaw, roll) and behavior (individuals and aggregations); and the resulting frequency-dependent backscatter. Exclusive use of acoustics to identify aquatic organisms requires metrics that discriminate among species with similar body types and implementing these metrics in routine data processing. Increased frequency bandwidth and angle-dependent backscatter measurements increase the information content of acoustic data to help achieve this goal. [Work supported by ONR and NOAA Fisheries.]

WEDNESDAY MORNING, 7 JUNE 2006 ROOM 554AB, 8:30 A.M. TO 12:00 NOON

Session 3aBB

Biomedical Ultrasound/Bioresponse to Vibration: Imaging, Wave Propagation, and Other Topics

Yuri A. Pishchalnikov, Chair
Indiana Univ., Anatomy and Cell Biology, 635 Barnhill Dr., Indianapolis, IN 46202

Contributed Papers

8:30

3aBB1. Nonlinear Rayleigh waves in soft tissue. Evgenia A. Zabolotskaya, Yurii A. Ilinskii, and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

Linear and nonlinear Rayleigh waves in soft elastic, tissuelike media are investigated theoretically. When the shear modulus is much smaller than the bulk modulus, as it is for soft tissue, the medium can be modeled as being incompressible provided shear stress alone is sufficient to support a particular mode of propagation. Such is the case for Rayleigh waves. Propagation speeds and penetration depths of the transverse and longitudinal components of Rayleigh waves are evaluated under these conditions. Differences between Rayleigh waves in ordinary solids and in soft elastic media are discussed. In tissuelike media, the transverse component obeys a wave equation, and the longitudinal counterpart obeys the Laplace equation. It is also shown that Rayleigh waves in soft elastic media are degenerate in the sense that the potential and solenoidal components cannot be separated. The analysis is based on Hamiltonian mechanics, following an approach developed by the authors for modeling nonlinear surface acoustic waves in both isotropic and anisotropic solids. Nonlinearity coefficients are estimated that characterize the rate of waveform distortion and harmonic generation. Numerical simulations of nonlinear propagation effects in soft elastic incompressible media are simulated in the quadratic approximation. [Work supported by NIH Grant EB004336.]

8:45

3aBB2. Coupling between elastic strain and interstitial fluid flow: Ramifications for poroelastic imaging. Ricardo Leiderman, Paul E. Barbone (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215), and Assad Oberai (Rensselaer Polytechnic Inst., Troy, NY 12180-3590)

Elastography refers to a host of imaging techniques that enables the quantitative measurement of tissue displacement and strain *in vivo*. In its most common modality, ultrasound is used to map the displacement, and an inverse problem to find the tissue biomechanical properties results, based on the tissue's constitutive equation, and the law of conservation of momentum. In most previous applications, the constitutive model for the tissue is assumed to be elastic or viscoelastic. It is well recognized, however, that soft tissues contain both fluid and solid phases. Furthermore, the fluid may flow within and between several different compartments. This paper addresses the question: Can ultrasound be used to image and quantify interstitial fluid flow in soft tissues from spatio-temporal patterns of elastic strain? To answer this, we use the mathematical model of vascularized soft tissue, which includes the effects of fluid flow and the possibility of exchange between fluid compartments in conjunction with finite-element modeling to predict the effects of fluid fluid flow on the spatio-temporal patterns of soft-tissue elastic strain.

9:00

3aBB3. Transcranial longitudinal and shear sound speeds in animal models. P. J. White, G. T. Clement, and K. Hynynen (Dept. of Radiol., Harvard Med. School, Brigham and Women's Hospital, Boston, MA 02115)

For nearly every experimental procedure or device developed for clinical application, animal models are vigorously pursued to not only ensure efficacy and safety of the tool, but as research models for the sequential observation of phenomena arising from simple systems to the complex. When novel procedures are introduced or proposed for study, it is imperative to ensure that effects observed in animal models are congruous to the expected effects in humans. In the case of transcranial ultrasound, the recently proposed application of ultrasound through the skull bone at incident angles to induce shear mode transmission required the quantitative assessment of shear sound speeds in human and animal skull bone. Using time-of-flight measurement methods with fresh and formalin-fixed skulls, measurements of longitudinal and shear sound speed were performed on pig, sheep, rabbit, and rat skull bone specimens. Discrepancies were observed in some cases between sound speeds in animal skull bone and those obtained from earlier studies of human skull bone. Results from these measurements will give indication as to whether certain animal models are sufficiently suited for transcranial longitudinal and shear mode ultrasound studies. [Work supported by NIH: R01EB003268 and R21EB004353.]

9:15

3aBB4. Ultrasonically determined thickness of long cortical bones. Petro Moilanen, Maryline Talmant, and Pascal Laugier (Laboratoire d'Imagerie Paramtrique, CNRS UMR 7623, Univ. Paris 6, 15 Rue de l'Ecole de Medecine, 75006 Paris, France, pemoilan@cc.jyu.fi)

Recent experimental studies have shown that an ultrasonically determined thickness (U-Th), based on a fit between measured and theoretical velocity of a fundamental flexural guided wave, correlated significantly with the cortical thickness (C-Th) of human radius bones. In this study, suitability of the theoretical model for fitting with experimental results for bones was evaluated. Numerical model, based on 3-D x-ray reconstructions of human radii ($n=29$), was developed, and wave propagation was simulated using the finite-difference method. In spite of variations in cortical thickness along the propagation path, a wave mode consistent with A0 Lamb mode was identified, and U-Th (2.7 ± 0.6 mm) was indeed strongly correlated ($r^2=0.90$, $p<0.001$) with C-Th (3.0 ± 0.6 mm) (2D

results). Three-dimensional simulations were then performed yielding similar results. These results were in good accordance with the previous *in vitro* experiments. Particularly, slight variation in geometry of bone did not affect too strongly the dispersion of fundamental guided waves, and simple analytical theory can be used as a satisfactory model for the average geometrical properties of cortical bone.

9:30

3aBB5. Sensitivity of 1-MHz ultrasound compression and shear waves propagating on cortical bone porosity: A numerical study. Cecile Baron, Maryline Talmant, and Pascal Laugier (Laboratoire d'Imagerie Parametrique, UMR 7623, 15 rue de l'Ecole de Medecine, 75006 Paris, France)

The objective of this study was to describe how the porosity of cortical bone affects ultrasound propagation (wavelength: mm scale). The bulk compression and shear velocities were determined for 40 specimens of human radius using finite-difference computation of transient ultrasound bulk waves in axial and transverse directions. The 3D microstructure of the specimens were obtained using high-resolution synchrotron radiation micro-computed tomography (voxel size $=10\mu$m). The pores were filled with a nonviscous fluid and the model assumed the bone material to be elastic and transversely isotropic or isotropic. All bulk wave velocities decreased quasi-linearly with increasing porosity, e.g., typically 28 m/s/%por for the axial compressional wave velocity. For the radial compressional wave velocity, the rate depended on the assumption on the bone material (for isotropy 56 m/s/%por and for transverse isotropy 19 m/s/%por). In case of transversely isotropic bone matrix, cortical bone tended to become more acoustically isotropic when its porosity increases in agreement with previous experimental studies. This study may help to develop more accurate models of cortical bone used to design improved ultrasound devices and to provide guidelines for establishing methods for the ultrasonic determination of bone properties.

9:45

3aBB6. A constrained optimization technique for the design of a uniformly excited $N\times N$ ultrasonic transducer micro-array. James Whitney, III and Ralph Etienne-Cummings[a] (Morgan State Univ., 1500 E. Coldspring Ln., Baltimore, MD 21251)

A method for obtaining the optimal and specific sensor parameters for an ultrasonic transducer micro-array is presented. The ultrasonic transducer array sensor parameters, such as frequency of operation, element size, interelement spacing, number of elements, and transducer geometry, are adjusted using a quadratic programming method to obtain maximum directivity while being constrained by MEMS design rules, a total array size of 4 mm^2, and the required resolution for microsurgery. The technique is used to determine whether a uniformly spaced $N\times N$ transducer array is capable of resolving structures as small as 2 μm. The parameters extracted from the proposed technique will be used in the development of tool-mounted sensors for imaging and robotic guidance for retinal microsurgery. [a] Presently at Johns Hopkins Univ.

10:00–10:15 Break

10:15

3aBB7. A single-channel low-frequency method for ultrasound imaging. P. J. White and G. T. Clement (Dept. of Radiol., Harvard Med. School, Brigham and Women's Hospital, Boston MA 02115)

Transcranial ultrasound imaging of brain structures and pathologies have traditionally suffered from compromised signal-to-noise ratios due to the bone-mediated increase in signal attenuation with increasing frequency. Yet, established modalities of image reconstruction with ultrasound measurements rely on higher frequencies for radial resolution and large temporal bandwidths for axial resolution. A new method that substantially reduces the dependency of image resolution on ultrasound frequency is proposed. Using an asymmetric transducer design, an unfocused, broadband, and frequency-separated ultrasound field is created in a region-of-interest (ROI). In this fashion, a unique complex spectral label can be assigned to each point in the ROI with a spatial resolution independent of source frequency. Ultrasound scattering sites within the ROI are then spatially localized by identifying the unique frequency-phase signature detected by either a point receiver or by the transmitting transducer. A cross-correlation method is used to reconstruct the echogenic profile of the ROI using *a priori* measurements of the field. Since the method's axial and radial resolutions are dependent on bandwidth rather than frequency, deeper penetration into tissues and application in highly attenuation tissues is feasible. Also, the prospect of multiple-dimension imaging with a single channel will allow for reduced emitter and electronic complexity. [Work supported by NIH: R21EB004353 and U41RR019703.]

10:30

3aBB8. Three-dimensional ultrasound image registration and shear elastic modulus reconstruction. Michael S. Richards (Dept. of Biomed. Eng., Boston Univ., 44 Cummington St., Boston, MA 02215, msrich@bu.edu), Nachiket H. Gokhale, Paul E. Barbone (Boston Univ., Boston, MA 02215), and Assad A. Oberai (Rensselaer Polytechnic Inst., Troy, NY 12180)

It is widely recognized that tissue pathologies often change biomechanical properties. For instance, neoplastic tissue is typically highly vascularized, contains abnormal concentrations of extracellular proteins (i.e., collagen, proteoglycans) and has a high interstitial fluid pressure compared to most normal tissues. The aim of this work is to develop and evaluate an ultrasound technique to quantitatively measure and image the mechanical properties of soft tissues in three dimensions. The intended application of our work is in the detection and diagnosis of breast cancer and other soft tissue pathologies. The specific goal of this project is the accurate measurement of the elastic shear modulus distribution of a tissue-mimicking ultrasound phantom, using a 3-D ultrasound imaging system. This requires the design and characterization of algorithms to provide three-dimensional motion estimates from ultrasound images and to solve the three-dimensional inverse problem to recover shear elastic modulus. Modulus images, reconstructed using our algorithms, are presented and compared to values determined by independent mechanical testing. The effect of the boundary conditions and regularization methods on the reconstructed modulus images and the uniqueness of the solution is also discussed.

10:45

3aBB9. Motion detection for vibro-acoustography. Matthew Urban and James Greenleaf (Dept. of Physiol. and Biomed. Eng., Mayo Clinic College of Medicine, 200 First St. SW, Rochester, MN 55905)

Vibro-acoustography is a method that uses ultrasound radiation force to investigate the mechanical response of an object [Fatemi and Greenleaf, ''Ultrasound-stimulated vibro-acoustic spectrography,'' Science, **280**, 82–85 (1998)]. The dynamic radiation force causes the object to vibrate within the focal region of the transducer, which creates a sound field called acoustic emission measured by a hydrophone. To gain a better understanding of the objects investigated with vibro-acoustography, a method is proposed that measures the displacement of the local vibration. A pulse-echo ultrasound method is implemented with the same transducer used to create the radiation force to measure the displacement of the object. A Kalman filter approach is used to process the pulse-echo data to obtain estimates of the local displacement [Zheng *et al.*, ''Kalman filter motion detection for vibroacoustography using pulse-echo ultrasond,'' 2003 IEEE Ultrasonics Symposium, pp. 1812–1815 (2003)]. Results using this ultrasound method to measure vibration of a metal beam compare well with laser vibrometer results. This method can reliably measure displacements with amplitudes of 120–2000 nm with error less than 9%. A pulse-echo method is described that provides displacement detection with simultaneous radiation force generation for vibro-acoustography applications using a single transducer. [Work supported by Grants EB002640 and EB002167 from NIH.]

11:00

3aBB10. Simulation of B-mode image synthesis with the fast near-field method and time-space decomposition. James F. Kelly and Robert J. McGough (Dept. of Elec. and Comput. Eng., Michigan State Univ., 2120 Eng. Bldg., East Lansing, MI 48824)

Rapid calculations of scattered and pulse-echo acoustic fields are required for simulations of ultrasonic imaging systems. Due to the large computational grids, broadband pulses, and complex scattering geometries, these simulations are computationally intensive with traditional tools. A new computational framework, based on the fast near-field method and a time-space decomposition technique, is developed for transient acoustic scattering problems. By analytically decomposing the temporal and spatial dependence of the pulse-echo field, this time-space decomposition method computes the first-order scattered field in an efficient and flexible manner. Due to the bandlimited nature of the time-space decomposition approach, sampling frequencies commensurate with the bandwidth of the excitation pulse are used in the scattering simulations. B-mode imaging is simulated by computing the scattered fields due to a collection of scatterers for multiple steering angles. The pulse echo field generated by a 128-element phased array of rectangular elements is derived and numerically evaluated. The computed transient fields are validated and compared with the FIELD II software in terms of computation time. Preliminary results show that the combination of the fast near-field method and time-space decomposition is significantly faster than FIELD II for numerical simulations of diagnostic imaging systems.

11:15

3aBB11. Transcranial ultrasound tracking of brain shift. Anthony B. Ross and Greg T. Clement (Brigham and Women's Hospital, Harvard Med. School, 221 Longwood Ave., Rm. 514, Boston, MA 02115)

Transcranial ultrasound provides a portable, inexpensive, rapid, and nonionizing technique to measure brain shift and structural changes. This technique may have important implications in neurosurgery, emergency room diagnosis, and long-term or frequent patient monitoring. The ability of transcranial ultrasound to measure motion inside the skull was investigated in *ex vivo* human skulls. The skulls were submerged in degassed water and a 2-D velmax positioning system was used to position a small needle within the skull. An array of low-frequency transducers (0.5–2 MHz) was positioned in a frame around the skull. Ultrasound pulses were sent in longitudinal and shear wave modes through the skull. The digitized receive signals of all transducers were then analyzed with Matlab. Using geometric and time-of-flight analysis the magnitude and direction of the needle shifts were determined. The transcranial ultrasound tracking system was able to accurately (within 0.2 mm) detect the needle motion in the skull. Overall, transcranial ultrasound using both shear and longitudinal waves was shown to be an effective method to track motion in the skull and may be a useful real time method to correct for misregistration of presurgical images. [Work supported by NIH Grants R21EB004353 and U41RR019703.]

11:30

3aBB12. Singular spectrum analysis applied to ultrasonic imaging of brachytherapy seeds. Jonathan Mamou and Ernest Feleppa (F. L. Lizzi Ctr. for Biomed. Eng., Riverside Res. Inst., 156 William St., 9th Fl., New York, NY 10038)

Singular spectrum analysis (SSA) shows promise for reliably imaging radioactive seeds implanted in the prostate gland. Brachytherapy using small titanium-shelled radioactive seeds is proving to be a well-accepted means of treating prostate cancer. Unfortunately, implantation using needles inserted transperineally causes gland movement, and the resulting distortion may cause seed misplacement and dosimetry errors. If actual seed locations can be determined during the brachytherapy procedure, then dosimetry errors can be corrected immediately in the operating room. SSA utilizes pairs of eigenvalues derived from the autocorrelation matrix of envelope-detected radiofrequency echo signals to identify seed-specific signal repetitions. The power spectrum associated with a repetition signal is computed to derive a P-value indicative of the likelihood of the presence of a seed at the location of that repetitive signal. Simulations to assess performance as a function of different levels of white and speckle noise and the presence of signals at repetition periods not associated with seeds. Experiments in an ideal scattering environment were performed, using seeds implanted in beef. Results show an encouraging potential of SSA for seed detection and imaging, and illustrate the basic trade-off considerations for optimization of SSA for clinical applications. [Work supported by NIH Grant R01 CA098465.]

11:45

3aBB13. Cavitation and acoustic streaming in ultrasonic fields: Applications in bio-, leather-, and chemical processing. R. Nagarajan, N. P. Dhanalakshmi, A. Krishnan (Dept. of Chemical Eng., IIT Madras, Chennai, India 600036, nag@iitm.ac.in), V. Sivakumar, G. Swaminathan (Central Leather Res. Inst., Adyar, Chennai, India 600036), and M. Goodson (Crest Ultrasonic Corp., Trenton, NJ 08628)

High-frequency, high-intensity ultrasound influences chemical reactions and transport phenomena via two principal mechanisms, acoustic streaming and cavitation. Bubble implosions associated with acoustic cavitation generate very high temperatures (1000s of Kelvin) and pressures (100s of atmospheres) which can accelerate many biochemical reaction rates by several orders of magnitude; acoustic streaming unidirectional high-velocity flow induced by the pressure gradient can substantially enhance heat-, mass-, and momentum fluxes. These aspects of acoustically driven process intensification are investigated here for three different applications: (1) growth characteristics, enzyme production, and ROS (reactive oxygen species) of *Bacillus subtilis* under varying conditions of ultrasonic frequency (58–192 KHz) and input power (maximum 500 W, variable between 40% to 100%); (2) dynamics of dye uptake by leather under these conditions; and (3) sono-chemical production of a fluorescent product from the reaction of two nonfluorescing species. Experimental data indicate that a combination of cavitation and acoustic streaming, e.g., by using dual-frequency systems that combine a low frequency (e.g., 58 kHz) with a higher frequency (e.g., 192 kHz), is optimal for accelerating processes that involve simultaneous heat-, mass-, and momentum transfer; purely kinetically limited processes, however, are most sensitively affected by cavitation characteristics of the applied field.

Session 3aED

Education in Acoustics: Undergraduate Research Poster Session

E. Carr Everbach, Chair
Swarthmore College, Dept. of Engineering, 500 College Ave., Swarthmore, PA 19081-1397

Contributed Papers

All posters will be on display from 9:00 a.m. to 12:00 noon. Contributors will be at their posters from 9:00 a.m. to 10:30 and posters will remain on display until 12:00 noon.

3aED1. Quantifying rain noise on three different roof structures. Sunil Patel and Lily M. Wang (Architectural Eng. Prog., Univ. of Nebraska-Lincoln, 1110 S. 67th St., Omaha, NE 68182-0681)

A system for measuring rain noise has been constructed, consisting of a 3.4-m^3 structure with a slanted roof and an artificial rain generation apparatus using PVC pipes and an appropriate spray nozzle. Three different roof structures have been tested for the level of rain noise generated within the structure: a shingle roof, one made with an EPDM membrane, and a corrugated metal panel roof system. Measurements show that the shingle and EPDM membrane roofs perform approximately the same and outperform the corrugated metal panel system by at least 13 dB across the frequency bands of 250 to 5000 Hz. The test system and detailed results are presented, as well as compared to guidelines provided in a draft ISO standard on measuring rain noise. [Work supported by UNL UCARE Grant.]

3aED2. Quantitative evaluation of architectural acoustics computer models. Christopher J. Beers and Robert D. Celmer (Acoust. Program & Lab., Mech. Eng. Dept., Univ. of Hartford, 200 Bloomfield Ave., West Hartford, CT 06117)

Computer modeling has enhanced many professional fields, including that of architectural acoustics. The accuracies of these models should be known if they are to be used in the design process. While some calculations in architectural acoustics software have been well established (such as the Sabine and Eyring reverberation time), the development of other higher-order algorithms is more recent. The purpose of this investigation is to examine possible methods for evaluating accuracy in architectural acoustics computer models. Recordings were made in an existing space, and analogous audio files were generated through one such software package. The recorded and calculated audio files were then compared via multiple means. The effectiveness and appropriateness of each metric was analyzed. Conclusions were drawn regarding the metrics and subsequent research implications were explored. [Work supported by ASA Robert W. Young Award for Undergraduate Student Research in Acoustics.]

3aED3. Wavelet analysis of active cavitation detector output. Kofi A. Anguah and E. Carr Everbach (Eng. Dept., Swarthmore College, Swarthmore, PA 19081)

Active cavitation detectors (ACDs) have provided useful information concerning the dynamics of microbubbles exposed to intense pulsed ultrasound. In an examination of the effects of nearby solid surfaces on the asymmetrical collapse of bubbles, ACD signals consisting of 60-cycle 30-MHz tone bursts of <0.1 MPa peak negative pressure returned complex echoes for individual bubbles and groups of bubbles. The Gauss-Harmite wavelet transform was used to analyze these signals to differentiate between symmetrical and asymmetrical collapse signatures, with the goal of identifying the percentage of asymmetrical bubble collapses in a cavitation field. [Work supported by NIH R-15 EB004630-01.]

WEDNESDAY MORNING, 7 JUNE 2006 BALLROOM D, 10:25 TO 11:30 A.M.

Session 3aID

Interdisciplinary: Distinguished Lecture

James F. Lynch, Chair
Woods Hole Oceanographic Inst., Woods Hole, MA 02543

Invited Paper

Chair's Introduction—10:25

10:30

3aID1. Status of acoustics in Russia. Nikolay A. Dubrovskiy (Andreyev Acoust. Inst., Shvernika Ul, 4, Moscow 117036, Russia)

Acoustics is developed mainly in Moscow, Saint Petersburg, Nizhny Novgorod, and Vladivostok. Studies are performed in different directions including both fundamental studies and applied research connected with the needs of Russian industry and the developing Russian market. Coordination of works on acoustics and dissemination of knowledge are implemented mainly by the Russian Acoustic Society (RAO), East European Association of the acousticians (EEAA), and the Council for Acoustics of the Russian Academy of Sciences. RAO carries out 1 or 2 scientific sessions yearly. This year RAO conducted two sessions. The basic periodical of Russian acousticians is *Acoustical Physics*. Science (and acoustics, in particular) in Russia is slowly reviving after a crisis happened after the collapse of the Soviet Union. This crisis manifests itself in significant reduction in funding; loss of prestige in being a scientist; young people leaving science and the aging of personnel; capable acousticians leaving the country; lack of demand for scientific innovations by Russian industry; lack of experience to work within the market economy; and absence of adequate tools for financing R&D. Wide international collaboration has become a positive factor for Russian acoustics. ASA provided valuable financial support for young Russian acousticians in recent years. [This lecture is supported by the ASA.]

WEDNESDAY MORNING, 7 JUNE 2006 BALLROOM E, 9:00 TO 11:30 A.M.

Session 3aMU

Musical Acoustics and Structural Acoustics and Vibration: Finite Element and Finite Difference Methods in Musical Acoustics II

Rolf Bader, Cochair
Univ. of Hamburg, Inst. of Musicology, Neue Rabenstr. 13, 20354 Hamburg, Germany

Uwe J. Hansen, Cochair
Indiana State Univ., Dept. of Physics, Terre Haute, IN 47809

Invited Paper

9:00

3aMU1. Contact problems with friction applied to musical sound synthesis. Joel Bensoam (Acoust. Instrumental Team, Ircam, 1, place Igor Stravinsky, 75004 Paris, France)

Modalys, a sound synthesis software developed at Ircam for research and musical applications, allows one to build virtual instruments based on physical models to obtain the most entire range of expressive variations in the instrument in response to intuitive controls. An instrument, as a complex structure, is described by the mechanical/acoustical interaction of its components (strings, tubes, resonators, soundboard, . . .). Some new research has been done recently to extend the sound prediction to three-dimensional objects with the help of numerical methods. In particular, theoretical and numerical treatment of the unilateral dynamic contact with friction between two arbitrary elastic bodies is studied. In addition to the classical variational statement that arises from static problems, a dynamic contact condition is needed and found by adjusting the energy balance law to the impenetrability condition. In the context of infinitesimal deformation, a reciprocal formulation is then used to reduce this well-posed problem to one involving Green functions

defined only on contact surfaces. This allows us to use considerably fewer unknowns compared to finite difference algorithms. The ability of the method to predict the contact interaction between two elastic bodies, irrespective of the material constitution and geometry, is highlighted by analytical and numerical simulations.

Contributed Papers

9:30

3aMU2. Whole-body finite-difference formulation of the violin. Rolf Bader (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstrasse 13, 20354 Hamburg, Germany, R_Bader@t-online.de)

The whole body of a violin was modeled with the finite-difference method in terms of a transient calculation. The model, like a previous guitar body model, takes bending or transversal and longitudinal or in-plane motion of the wooden body into consideration and couples these two kinds of vibrations. Here, the curvature of the violin body also enhances this coupling according to the amount of curvature. The string/bow mechanism was modeled in terms of sticking/slipping, depending on the displacement difference of the bowing point to a point next to this bowing point. The different parts of the violin show different behavior within the steady state of one bowing cycle. Most noticeable, the ribs showed a chaotic motion, causing it to be quite small and greatly influenced by the surrounding parts. As this is practically the only part with noisy components, we can say the violin bowing noise comes from the ribs. Additionally, initial transients were modeled by changing the bowing pressure and velocity over the intial transient phase for a soft and a hard attack, resulting in reasonable descriptions of the violin playing attack.

9:45

3aMU3. A 3D finite-element method (FEM) of the body of a classical guitar. Jan Richter (Jan Richter, Gefionstrasse. 11, 22769 Hamburg, Germany)

A 3D finite-element method (FEM) of a body of a classical guitar is performed. It consists of the rib, back plate, top plate with bridge, soundhole and fan bracing, and wooden block. The idea of this project is to compare the eigenfrequencies of different body shapes and make assumptions about the way the elements of the body influence the sound of a classic guitar. The aim is to calculate the influence of the thickness of the bottom, the amount of bracing, and the appearance of the block. As a submittal a real guitar was taken and the geometrical pararmeters are precisesly modeled in the computer. These solutions are found by using the finite-element method, which allows calculation of the different eigenfrequencies and mode shapes of the guitar body.

10:00–10:15 Break

10:15

3aMU4. Modeling of steeldrum sounds using finite element and finite difference methods. Michael Steppat (Institut fr Musikwissenschaft, Universitt Halle, Kleine Marktstr. 7, 06099 Halle(Saale), steppat@cs.tu-berlin.de)

The Caribbean steeldrum, which was developed over 60 years ago in Trinidad, is one of the most novel music instruments. It is made of leftover oil drums. On the bottom of these drums, dents can be formed with a specialized technique, to get playing areas with a definite pitch. The tonal range of a whole orchestra can be reached with different models, covering a tonal range from bass to soprano. Despite the usage of the same material and even if they are made by the same tuner, the instrument can sound very different. Good instruments can be played in various striking strengths, without producing interfering noise. Instruments of weaker quality tend to produce interfering noice, even at small striking strengths. The sound of the instrument can be simulated with the finite difference method and finite element method . With the finite element method, a model of the vibration system can be formed by small rectangular plate elements. To simulate the sound of a note, a transient response analysis can be calculated. The simulation takes various production techniques into account, e.g., different shapes of the dent.

10:30

3aMU5. Tuning parameters of a Nigerian slit gong. Jacob Skubal, Nathan Hall, and Thomas Moore (Dept. of Phys., Rollins College, Winter Park, FL 32789, tmoore@rollins.edu)

The Nigerian slit gong is a wooden idiophone, hand-carved from a single log, that is characterized by its ability to produce two distinct notes when struck. The resonant modes of the gong have been studied using a finite-element analysis model, which was validated by comparing the predicted mode shapes to those observed by electronic speckle pattern interferometry. Using this model, the effects of changing the parameters available to the artisan when carving the instrument have been studied. Results of the model are compared with a simple theory, and the physics of tuning the Nigerian slit gong is discussed.

10:45

3aMU6. A three-dimensional, time-dependent simulation of a Turkish cymbal using the finite element method. Malte Muenster (Mansteinstr. 46, 2053 Hamburg, Germany, m.muenster@arcor.de)

We built computer-aided simulations of several musical instruments to study microstructural behavior of initial transients and eigenvalues. Our finite element model (FEM) of a cymbal is an example for the possibilities of this method. Proceeding from a three-dimensional mindlin plate model, we discovered some former unknown, sometimes moving fundamentals, depending on striking positions, hammering during production process, sizes, forms, and alloys. Special movements of the bell of the cymbal have been investigated. Moreover, experimental studies and simulation are able to be compared directly concerning their variety of dynamic behavior. This leads to more and more precise formulation of requirements of musical instruments. Cymbals are among the oldest instruments. The production techniques, some were real secrets, were given from father to son to the next generations of instrument builders. But only recent developments in musical acoustic research include computational modeling and signal processing power to show that most parts of musical sounds of a cymbal are caused by the features mentioned above, in regard to styles, expression, and instrument qualities important to musicians. Our studies try to give an idea of formulating and influencing sound parameters of musical instruments using FEM simulations. The target is to describe how to build a cymbal fulfilling any condition of sound.

11:00

3aMU7. Turbulent finite difference model of simplified musical geometries. David Linke and Berit Janssen (Lohkoppelstrasse 32c, 22083 Hamburg, Germany, david_linke@web.de)

A turbulent flow in a two-dimensional space is modeled using a finite difference method of the Navier-Stokes equation taking the k-epsilon turbulence model into account. In this formulation the flow is divided into a laminar and an averaged turbulent part. Mathematically this changes the normal Navier-Stokes equation into the Reynolds averaged Navier-Stokes equation. Kolmogorov's idea was that the energy of the laminar flow is transverted into eddies which get smaller down to the smallest eddy size, where this enery is completely dissipated. This introduces two new parameters, the turbulent energy k and the turbulent dissipation epsilon. Generally this turbulence model adds additional damping through the Reynolds stress tensor. This investigation compares turbulent and nonturbulent flows using different musical relevant geometries, like surfaces and edges.

11:15

3aMU8. Conservative numerical methods for sound synthesis. Stefan Bilbao (Univ. of Edinburgh, Rm. 7306B, James Clerk Maxwell Bldg., King's Bldg., Mayfield Rd., Edinburgh, EH9 3JZ UK)

Direct numerical simulation techniques, such as finite difference and finite element methods, are playing an increasingly important role in sound synthesis based on physical models of musical instruments. While they can lead to enormous improvements in sound quality over existing methods, a problem of mounting importance is that of maintaining numerical stability when the system to be simulated is strongly nonlinear, as is the case for many musical instruments. This problem is especially acute if real-time synthesis is to be an eventual goal. Standard stability checking machinery, such as frequency domain (von Neumann) analysis, does not immediately apply, and, often, the most straightforward numerical methods may exhibit such instabilities in an unpredictable manner. In this paper, various techniques for the construction and analysis of numerical methods for strongly nonlinear musical systems are presented; they are based on the translation of the principle of conservation or dissipation of energy to discrete time and can lead directly to sufficient nonlinear stability conditions, without invoking frequency domain concepts. Special attention is paid to the nonlinear string, bar, and plate models. Numerical results and sound examples will be presented.

Session 3aNS

Noise, Architectural Acoustics, and Speech Communication: Fifty Years of Speech Privacy I

Gregory C. Tocci, Chair
Cavanaugh Tocci Associates, Inc., 327F Boston Post Rd., Sudbury, MA 01776-3027

Chair's Introduction—8:15

Invited Papers

8:20

3aNS1. From articulation index to speech privacy: A review. Leo L. Beranek (975 Memorial Dr., Ste. 804, Cambridge, MA 02138, beranekleo@ieee.org)

Communication by voice over wired and radio systems is ever more important. Voice recognition by machine is commonplace today, whether it involves a person telephoning into an organization or the government spying on international communications. In these examples, context is not yet highly developed, so that the advantage a person has in understanding a sentence even if all words are not perceived does not ordinarily exist. Thus, correct understanding of isolated words is crucial. Two papers pioneered in establishing a metric for determining the degree of understanding of words in the presence of noise [French and Steinberg, ''Factors governing the intelligibility of speech sounds,'' J. Acoust. Soc. Am. **19**, 90–119 (1947); Beranek, Proc. I.R.E., **35**, 880–890 (1947)]. This paper will review the basic principles taught in those papers. Very low values of AI (i.e., little or no intelligibility) proved critical in studies on speech privacy in buildings by BBN scientists some 50 years ago. Those early studies also confirmed the relative importance of masking sounds in offices and other spaces, sometimes referred to as acoustic perfume, in achieving desired degrees of speech privacy.

8:40

3aNS2. Speech privacy in buildings: A review. William Cavanaugh (Cavanaugh Tocci Associates, Inc., 327F Boston Post Rd., Sudbury, MA 01776) and Parker Hirtle (Acentech Inc., Cambridge, MA 02138)

Studies on speech privacy problems in buildings at Bolt Beranek and Newman begun in the mid 1950s showed convincingly that a person's sense of acoustical privacy was directly related to the intelligibility of the intruding speech over the continuous ambient sound present [Cavanaugh *et. al.*, J. Acoust. Soc. Am. **34**, 475–492 (1962)]. Both in laboratory and in real-world settings people feel they have confidential privacy when the articulation index (AI) of the intruding speech is 0.05 or less. Higher values of AI are acceptable for less demanding tasks, i.e., normal privacy. The method quantifies each of the significant variables involved and compares a summary of these metrics with response data for prior case histories both to predict, and, in the case of existing situations, to evaluate speech privacy conditions. R. W. Young reviewed the analysis method and showed that sufficient accuracy was preserved if simple, commonly used A-scale sound levels were used in place of more cumbersome frequency-based metrics [R. W. Young, J. Acoust. Soc. Am. *38*, 524–533 (1965)]. Later in the 1960s the method was successfully applied to open plan offices which were becoming more prevalent [Hirtle *et. al.*, J. Acoust. Soc. Am. **46**. 91A (1969)]. The lessons learned in applying the speech privacy calculation over the years are discussed.

9:00

3aNS3. Postoccupancy assessment of speech privacy in offices. Charles Salter, Randy Waldeck (Charles M. Salter Associates, 130 Sutter St., Ste. 500, San Francisco, CA 94104), and Kevin Powell (General Service Administration, San Francisco, CA 94111)

The General Services Administration (GSA) Workplace 20.20 initiative includes acoustical design analysis for offices. Internet questionnaires about acoustics and occupant satisfaction have been filled out by hundreds of GSA office workers. These results have been correlated with acoustical testing that our firm, Charles M. Salter Associates, Inc., has conducted during the last 4 years around the country. Survey responses, which are correlated with the results of our acoustical testing, will be presented.

9:20

3aNS4. New developments in speech privacy design. Kenneth P. Roy (Armstrong World Industries, 2500 Columbia Ave., Lancaster, PA 17604)

The design of architectural spaces for the purpose of speech privacy attainment has been quietly practiced over many years. In the late 1960s and 70s, with the introduction of open-plan office layouts, the need for speech privacy became an issue of productivity, as well as one of confidentiality. Today, with the introduction of government regulations such as HIPAA, the need for confidential speech privacy in the healthcare industry has hit the mainstream of our consciousness, even as individuals. And, in this day of (in)security the issues of espionage and terrorism have implications for architectural acoustic design of spaces for speech security. This paper considers both the architectural design trends and issues, and the measurement and evaluation methods being developed in response to these needs for speech privacy, as we move boldly forward into the next 50 years.

9:40

3aNS5. Open-plan privacy revisited. Rein Pirn (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

Driven by a desire for economy, the open office has been with us since before World War II. However, it was not until the 1960s that largely through the work of a Quickborn (Germany)-based team of space planners it became an art form, touting all manner of organizational benefits, and flexibility. Though speech privacy was recognized as an issue, there was no systematic approach to resolve it. Against this backdrop, five variables that would affect speech privacy in an open plan were identified and their relative contributions quantified [R. Pirn, ''Acoustic variables in open planning,'' J. Acoust. Soc. Am. **49**, 1339–1345 (1971)]. This work and its impact on acoustical design of the 1970s is reviewed, as are the trends in and the failings of current open-plan design practices.

10:00–10:20 Break

10:20

3aNS6. The speech intelligibility index standard and its relationship to the articulation index, and the speech transmission index. Chaslav Pavlovic (Sound ID, 3040 West Bayshore Rd., Palo Alto, CA 94301)

The history of the Speech Intelligibility Index (SII) Standard (ANSI S3.5-1997) and its relationship to the Articulation Index (AI) and the Speech Transmission Index (STI) are reviewed. The SII Standard defines a method for computing a physical measure (SII) that is highly correlated with the intelligibility of speech. The SII is calculated from acoustical measurements of speech and noise. The standard is a revision of the ANSI S3.5-1969, ''American National Standard Methods for the Calculation of the Articulation Index.'' In that sense the AI and the SII are one and the same thing. In addition, however, SII incorporates the Speech Transmission Index procedures appropriate for many conditions of use where the traditional AI does not apply [H. J. M. Steeneken, and T. Houtgast, ''A physical method for measuring speech-transmission quality.'' J. Acoust. Soc. Am. **67**, 318–326 (1980)]. Other changes of the Standard are due to new data which have been accumulated since 1969 for various parameters and procedures used in the calculations. These include spread of masking, standard speech spectrum level, and relative importance of various frequencies to speech intelligibility. In addition, the SII calculation details have been adapted to computer implementation rather than to manual chart-type computations.

10:40

3aNS7. Speech privacy: What's new? What's next? John S. Bradley and Bradford N. Gover (Inst. for Res. in Construction, Natl. Res. Council, Montreal Rd., Ottawa, K1A 0R6 Canada)

This paper will review recent work that has considered speech privacy criteria in open-plan offices and the development of new procedures to assess the architectural speech security of closed offices and meeting rooms. For open-plan offices, new subjective studies have confirmed the validity of a speech privacy design goal of SII no more than 0.20 (AI 0.15) and that an ideal masking noise level is ~45 dBA. For the speech security of closed meeting rooms, subjective studies have developed new weighted signal-to-noise ratio measures for rating the privacy of meeting rooms and a new measurement procedure has been developed to assess conditions close to the outside boundaries of a meeting room. While much progress has been made, further work is still required. For example, it is now clear that in many situations, privacy is significantly influenced by both signal-to-noise and reverberation in the rooms. Both factors must be included when assessing the privacy of situations with significant amounts of reverberation.

11:00

3aNS8. Speech privacy: New laws require changes in professional practice. David Sykes Remington Partners, 23 Buckingham St., Cambridge, MA 02138)

We take privacy for granted but changing technology and heightened government security concerns have altered the landscape, putting privacy back in the headlines. Privacy laws like HIPAA, GLBA, and PIPEDA compete against tech-based crimes like identify theft as well as security, laws like the Patriot Act, Homeland Security, and FISA. What constructive role can acoustics professionals play? Speech privacy is covered by most privacy laws, explicitly or implicitly—though enforcement procedures remain unclear. But, thanks to 50 years of research and standards development, a basis for objective, measurable enforcement of speech privacy exists and will soon be integrated into professional practice by the AIA. New policies and improved understanding by architects as well as legislators, regulators, judges, lawyers, and juries are beginning to stimulate demand for speech privacy engineering services including research, measurement, design, monitoring, site certification, expert testimony, and training. The TCAAN joint subcommittee on speech privacy (soon to be an S.12 Work Group) was set up in November 2005 to lead this work. Subcommittee members are addressing two issues: Writing new Guidelines for Acoustics in Healthcare Facilities for the AIA and American Hospital Association, and developing a uniform speech privacy enforcement guideline for use in healthcare environments.

11:20

3aNS9. Noise in hospitals redux. Lewis S. Goodfriend (Lewis S. Goodfriend & Assoc., 760 Rte. 10, Whippany, NJ 07981, lsg@lsga.com)

A major result of the patient survey, conducted as part of the study, Noise in Hospitals, published in 1963, and funded by the U.S. Public Health Service, was that patients in private and semi-private rooms could hear conversations from the corridor and, on occasion, discussions of their diagnosis and condition by hospital staff. Also noted was the audibility of conversations from adjacent rooms. The study covered eight eastern hospitals of varied size. The results showed background ambient sound kevels in patient rooms, in the low-to-mid 40 dB(A) range. In general, the spectra, with low mid- and high-frequency levels, show why speech in the corridor, at normal levels, was highly intelligible in patient rooms in a number of hospitals. Much of the noise in hospitals is not continuous and that which was continuous, often had undesirable spectral or cyclic characteristics that added to the annoyance of the patients, but did not interfere with speech intelligibility, even with high, annoying, low-frequency levels. More recent information from one of the original hospitals and work by others indicates that the problem of noise in hospitals has not yet been well addressed.

11:40

3aNS10. Quieting Weinberg 5C: A case study in reducing hospital noise on a patient ward. Mark MacLeod, James E. West, Ilene J. Busch-Vishniac, Jeffrey Dunn (Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218), and Anita Reedy (Johns Hopkins Hospital, Baltimore, MD 21218)

Although noise in hospitals is among the top complaints of patients, visitors, and staff, little has been done to address the problem. Further, hospitals have become increasingly worried about using materials that might harbor bacteria, shunning traditional sound absorbers and making the problem even more acute. In this study, we investigated an oncology unit in a relatively new building at Johns Hopkins Hospital, with the aim of demonstrating a noise reduction. We found a series of problems at Weinberg 5C, including architectural design that focuses sound at the nurses station, hallways that create an effective waveguide around the unit, and the total absence of sound-absorbing materials. We were able to mitigate these problems through judicious placement of sound-absorbing panels on the ceiling and high on hallway walls. The panels were constructed of noncompressed fiberglass insulation wrapped in an antibacterial fabric and suspended with Velcro. Results show a drop in sound level of about 5 dB(A), a drop in reverberation time of roughly a factor of 3. A survey of patients and staff before and after the intervention shows a remarkable shift from the perception that the unit is noisy to a perception that it is reasonably quiet.

Session 3aPA

Physical Acoustics: Acoustic Microscopy at the Nanoscale

Joseph A. Turner, Cochair
Univ. of Nebraska-Lincoln, Dept. of Engineering Mechanics, Lincoln, NE 68588

Donna C. Hurley, Cochair
National Institute of Standards and Technology, Materials Reliability Div., 325 Broadway, Boulder, CO 80303

Chair's Introduction—7:55

Invited Papers

8:00

3aPA1. Characterization of elasticity, friction, anelasticity, and adhesion with nm lateral resolution, using the ultrasonic vibration modes of atomic force microscope cantilevers. W. Arnold, A. Caron, D. Rupp, K. Schwarz, H. Hirsekorn, and U. Rabe (Fraunhofer Inst. for Non-Destructive Testing (IZFP), Saarbrcken, Germany)

Ultrasound is combined with atomic force microscopy to achieve the lateral resolution of scanning probes for imaging. Atomic force acoustic microscopy (AFAM) and ultrasonic friction force microscopy (UFFM) exploit the vibrational modes of AFM cantilevers which extend from 10 kHz to several MHz. In the AFAM mode the cantilever vibrates in contact mode in one of its flexural resonances. Images can be obtained with the contrast depending on the local indentation modulus, which can be obtained quantitatively from the contact resonance frequencies. The lateral resolution is defined by the tip-sample contact radius and is 10 nm. At large ultrasonic amplitudes nonlinear effects that are related to adhesion become important. In UFFM shear stiffness and friction phenomena can be examined by evaluating the torsional cantilever resonances. At low lateral surface amplitudes, the sensor tip remains in elastic contact and the cantilever behaves like a linear torsional oscillator with damping. If the amplitude exceeds a critical value, the resonance develops a plateau due to stip-slick in the tip-sample contact, allowing one to monitor friction. Ultrasonic C-scan images and the corresponding contact resonance spectra will be shown and compared to theory. Applications of the AFAM techniques to measure local piezoelectricity and anelasticity will be discussed as well.

8:25

3aPA2. Quantitative nanomechanical measurements and imaging with atomic force microscopy methods. D. C. Hurley, M. Kopycinska-Mueller, R. H. Geiss, E. D. Langlois (NIST Mtls. Reliability Div., 325 Broadway, Boulder, CO 80305, hurley@boulder.nist.gov), and A. B. Kos (NIST Electromagnetics Div., Boulder, CO 80305)

We describe our work to evaluate the nanomechanical properties of surfaces, thin films, and nanostructures using the atomic force microscope (AFM). Our approach is based on atomic force acoustic microscopy (AFAM) and involves the resonant vibrations of an AFM cantilever in contact with a sample. We first describe the basic principles of AFAM, in which local elastic-property values are determined from the contact-resonance frequencies using models for the cantilever dynamics and for the tip-sample contact mechanics. Quantitative results for a variety of thin supported films ranging in stiffness from 50 to 250 GPa and as thin as 50 nm are given to illustrate our methods. Studies related to measurement accuracy involving the effects of film thickness and tip shape are also presented. Finally, we discuss new frequency-tracking electronics for rapid (typically 20 min per image), quantitative imaging of nanoscale elastic properties. We describe how such methods not only provide modulus mapping capabilities, but also have enabled us to detect variations in adhesion between a thin film and substrate.

8:50

3aPA3. Nanoscale acoustical imaging—deep or superficial? Andrzej J. Kulik, Gerard Gremaud, Laszlo Forro (Ecole Polytechnique Federale de Lausanne, IPMC-Station 3, CH-1015 Lausanne, Switzerland), Robert Szoszkiewicz (Georgia Inst. of Technol., Atlanta, GA 30332), and Nancy A. Burnham (Worcester Polytechnic Inst., Worcester, MA 01609-2280)

Several ideas of near-field detection of ultrasonic vibrations of the sample surface using atomic force microscopy (AFM) have been developed. Two major techniques are used: contact and nonlinear detection (intermittent contact). Most of the experiments were performed using low-MHz frequencies, meaning that wave propagation can be neglected in thin and small samples. In this talk we will analyze and present data obtained using two techniques: (a) contact measurements, where surface vibrations are sensed by the AFM tip, mapping local stiffness of the sample [Burnham *et al.*, J. Vac. Sci. Technol. B **14**, 794–796 (1996)], and (b) ''mechanical diode'' measurements, where high-amplitude surface vibrations are detected through nonlinearity of tip-sample interactions. This method gave access to local adhesion hysteresis [Szoszkiewicz *et al.*, Appl. Surf. Sci. **210**, 54–60 (2003); J. Appl. Phys. **99**, 014310 (2006)].

9:15

3aPA4. High-resolution ultrasound: A survey. Bernhard R. Tittmann (Penn State Univ., 212 Earth & Eng. Sci. Bldg., University Park, PA 16802)

This presentation provides a survey of high-resolution ultrasonic techniques and examples. With the discovery by Baranski, followed by the work of Boemmel and Dransfeld, the field of hypersonic wave propagation began to come into its own. The measurements over a range of frequencies from 100 MHz to 2 GHz, first in superconducting lead and then in type-II superconductors by Tittmann, took place under cryogenic temperatures. These and other results provided impetus for the development of a room temperature scanning acoustic microscope (SAM) by Quate. The licensing to three organizations for manufacturing the SAM led to the sale of many instruments which have since been used by many researchers around the world for materials characterization. This work was enhanced by Weglein with the interpretation of the $V(z)$ signature of materials. The room temperature SAM requires fluid for ultrasonic coupling and provides about 1 micron resolution at 1 GHz. With the onset of nanotechnology even finer resolution was required and the atomic force microscope was adapted by several groups around the world for use with ultrasound with the development of U-AFM and AFAM. This presentation will provide several examples of methods and results throughout the development of high-resolution measurement capability.

9:40

3aPA5. Recent advances in nanoindentation-based nanoscale dynamic mechanical analysis. Oden L. Warren and S. A. Syed Asif (Hysitron, Inc., 10025 Valley View Rd., Minneapolis, MN 55344)

Nanoindentation-based nanoscale dynamic mechanical analysis is rapidly gaining popularity as a quantitative small-volume characterization technique for soft matter such as polymers. In addition, the so-called nanoDMA technique can be coupled to scanning probe microscopy to quickly generate images showing the spatial distribution of storage and loss moduli, a time-saving advantage for the study of heterogeneous materials such as phase-segregated polymer blends. This presentation will address recent advances in nanoDMA, including instrumentation improvements, new measurement/analysis techniques, and efforts to push beyond the widely used linearized equation of motion to begin accounting for sources of nonlinearity not previously appreciated. If time permits, this presentation will also address the use of nanoDMA in combination with scanning probe microscopy for rapid screening of combinatorial polymer libraries.

10:05–10:20 Break

10:20

3aPA6. Resonance ultrasound microscopy for quantitative determination of local stiffness and local damping of solids. Hirotsugu Ogi and Masahiko Hirao (Grad. School of Eng. Sci., Osaka Univ., Toyonaka, Osaka, Japan 560-8531)

Home-built resonance ultrasound microscopy is presented for quantitatively determining local Young's modulus and local internal friction of materials. It evaluates the elastic properties from resonance frequencies and attenuation of an isolated langasite oscillator, touching the examined material. Noncontacting measurement (wireless and electrodeless) of the resonance frequency and attenuation of the oscillator has been achieved with a line antenna by the piezoelectric effect. The oscillator is supported by a fixture exactly at nodal points, and it touches the material at an antinodal point. Thus, acoustical contact on the oscillator surfaces occurs only at one point with the material, providing high sensitivity to the modulus and damping in the local area. Formulation of relationships between material's stiffness and internal friction, and oscillator's resonance frequency and attenuation is proposed using the generalized Hertzian-contact model. Results for polycrystalline copper, dual-phase stainless steel, and NbTi/Cu superconducting wire are shown, giving important knowledge including stiffness distribution within a crystallite. The results are compared by electron-backscattering-pattern measurements.

Contributed Papers

10:45

3aPA7. Acoustic microscopy inspection of layered structures. Jikai Du and Benhard R. Tittmann (Dept. of Eng. Sci. and Mech., The Pennsylvania State Univ., Univeristy Park, PA 16802, jxd338@psu.edu)

Thin films are widely applied in a variety of fields and, for example, are utilized in semiconductor devices, biological measurements, optical components, and micro-electro-mechanical-structures (MEMS). Thin films can serve for conductive and capacitive effects, tissue elastic properties evaluation, and control of optical properties. Therefore, the evaluation of thin film integrity is necessary and important. A thin film component is a layered structure and usually referred to as a film-substrate system. Acoustic microscopy is a nondestructive technique, which can generate surface acoustic waves (SAWs) in the thin film/substrate system. The energy of these SAWs is confined to the system surface and thus suitable for the evaluation of film properties. In our research, theoretical modeling of wave propagation in soft thin film/substrate system with different material properties and bond strength is established. A relationship between dispersive wave propagation and film integrity is discussed. A new experimental system is reported for measuring dispersion curves up to 1 GHz. Preliminary data are presented and compared to the theory. Good agreement was observed.

11:00

3aPA8. Correlation between elastic constant and perpendicular magnetic anisotropy in cobalt/platinum superlattice thin films studied by picosecond ultrasounds. Nobutomo Nakamura, Hirotsugu Ogi, Takeshi Yasui, and Masahiko Hirao (Grad. School of Eng. Sci., Osaka Univ., Machikaneyama 1-3, Toyonaka, Osaka 560-8531, Japan)

The easy magnetization direction of a Co monolayer thin film is always parallel to the film surface, but it can be perpendicular to the film surface for Co/Pt multilayer thin films when the Co layer consists of a few atomic layers. This property is called perpendicular magnetic anisotropy (PMA), achieving ultrahigh-density magnetic recording media. PMA occurs because of the large lattice misfit (near 10%) at the Co-Pt interfaces, and large elastic strains must affect the elastic constants through lattice anharmonicity. Thus, elastic constants and magnetic properties of Co/Pt superlattice thin film should show a correlation, which was investigated by picosecond ultrasound. The pulse-echo signals of the longitudinal wave

traveling in the thickness direction were detected by the laser pump-probe technique, which gave the elastic constant C33. The film thickness was determined by the X-ray reflectivity measurement. Co/Pt superlattice thin films were prepared by the magnetron-sputtering method, changing the Pt-layer thickness between 0.2–2 nm and keeping the Co-layer thickness to be 0.4 nm. Total film thickness was near 80 nm. C33 showed good correlation with the degree of PMA and it differed from the modulus predicted from the elastic constants of bulk materials.

11:15

3aPA9. Comparison of atomic force acoustic microscopy and dynamic nanoindentation methods for compliant polymer systems. Matthew Stagemeyer, Joseph A. Turner (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588), and Mark R. VanLandingham (US Army Res. Lab., Aberdeen Proving Ground, MD 21005-5069)

Nanomechanical characterization of soft materials such as polymers and polymer composites represents a unique and challenging research topic of current interest. In this presentation, two characterization methods are compared directly using a set of model PDMS polymer systems. Experiments were conducted using both nanoindentation and atomic force acoustic microscopy (AFAM) to determine the dynamic mechanical properties. The experiments were conducted to determine reduced elastic modulus E^* via AFAM measurement and storage modulus E via nanoindentation. The AFAM analysis was done using both a Hertzian contact model and a flat punch model. While the two methods showed marginal agreement, the resulting AFAM moduli were dependent on the model used in the analysis. Further investigation is needed to improve the accuracy of these measurements. In addition, further modeling is needed to improve the understanding of the tip-sample interaction in such systems. The results are anticipated to lead to the development of improved characterization methods for compliant materials. [Work supported by ARL.]

11:30

3aPA10. Capture of *in vivo* biological images below the surface of aqueous media with ultrasound atomic force microscopy. Sahar Maghsoudy-Louyeh, Jikai Du, and Bernhard R. Tittmann (Dept. of Eng. Sci. & Mech., The Penn State Univ., University Park, PA 16802, sum179@psu.edu)

Atomic force microscopy (AFM) is a useful tool for measuring sample topography and has been widely applied to imaging of biological cells. The Park Scientific M5 AFM in our laboratory, in addition to measuring sample topography, can also give images of spatial distributions of sample elastic properties through applying an ultrasonic AFM (U-AFM) technique. But the M5 AFM is normally a dry operation AFM, and this makes it unsuitable for imaging *in vivo* biological cells in aqueous media. To modify the M5 AFM for imaging in aqueous media, both mathematical modeling and physical experimentation were established. A modified AFM system was built based on the theory. As a result, both topography and elastic property images are obtained on biological samples in aqueous media. The details of the modification will be described and examples of under-liquid imaging will be presented.

11:45

3aPA11. Dissipation mechanisms associated with atomic force acoustic microscopy. Joseph A. Turner, Matthew Stagemeyer, and Kangzhi Shen (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588)

The impact of tip-sample dissipation mechanisms in atomic force acoustic microscopy (AFAM) and other techniques is evident in the width (and corresponding Q-factor) of the resonances observed. The Q-factors of resonances while in contact are often several orders of magnitude lower than those for air alone. Dependence on applied load, humidity, and vibration mode have also been observed. Here, two dissipation mechanisms are discussed with application to AFAM. First, frictional losses associated with the tip-sample oscillations are examined. The cyclic energy loss that accompanies tip vibrations is calculated for the combined normal and shear forces present during AFAM measurements using an extension of a Mindlin contact model. The dependence on mode, applied load and friction coefficient, and other parameters is seen explicitly. Second, the dissipation associated with losses from viscoelastic contact is calculated using simple viscoelastic materials models that are coupled with the vibrating-beam boundary value problem. In this case, the dependence of the dissipation on sample properties and tip geometry is determined. The role of each of these mechanisms for specific tip-sample systems is discussed. The results are expected to impact quantitative aspects of nanomechanical characterization techniques. [Work supported by ARL.]

Session 3aPP

Psychological and Physiological Acoustics: Detection, Discrimination, Identification, and More

Beverly A. Wright, Chair
Northwestern Univ., Dept. of Communication Sciences and Disorders, 2240 Campus Dr., Evanston, IL 60208

Chair's Introduction—7:35

Contributed Papers

7:40

3aPP1. An auditory perceptual change with repeating noise bursts. Patrick M. Zurek (Sensimetrics Corp., 48 Grove St., Somerville, MA 02144, pat@sens.com)

This talk describes a change in perception that is experienced when listening to repeating noise bursts. When a brief (10–20 ms) frozen noise burst is presented repeatedly at a rate of about 1/s, the perceived quality of the bursts changes. Initially, the noise bursts have the expected perceptual texture of white noise (like the consonant ''ch'' spoken in isolation). At some later time, usually 10–30 s after the start, that noise texture disappears and the perceptual quality becomes that of a dull click or thump. Initial observations of this perceptual change under a variety of stimulus manipulations will be described. Unlike perceptual changes involving repeated speech (e.g., semantic satiation and the verbal transformation effect), which are usually labile and involve multiple transitions among multiple perceptual forms, the observed change with noise bursts appears (thus far) to be only a two-state, one-way change with constant stimulus repetition.

7:55

3aPP2. Two types of learning revealed by evaluation of the best and worst daily performance during auditory-discrimination training. Beverly A. Wright, Yuxuan Zhang, and Christopher C. Stewart (Dept. of Commun. Sci. and Disord., Hugh Knowles Ctr., and Inst. for Neurosci., Northwestern Univ., 2240 Campus Dr., Evanston, IL 60208-3550)

Performance on many auditory discrimination tasks improves with practice. This learning could result from improvements in fundamental processing capacity (as is implicitly assumed), from improvements in the ability to consistently access already existing capacities, or from a combination of the two. To help distinguish among these types of improvement, we reanalyzed learning data on three basic auditory tasks, examining the best and worst, rather than the mean, thresholds on each of 6–10 days of training. On one task (interaural-level discrimination), the best and worst thresholds decreased in parallel, indicating an improvement in fundamental processing capacity. On another task (interaural-time discrimination), the best thresholds remained unchanged but the worst thresholds decreased, suggesting an increased consistency in accessing already existing capacities. Demonstrating a combination of these two types of improvement, on the third task (duration discrimination), both the best and worst thresholds improved, but the worst thresholds improved more. These results indicate that learning on auditory tasks can result not only from refinements in fundamental stimulus encoding, but also from increases in the ability to use the best information that was already available before training. [Work supported by NIH/NIDCD.]

8:10

3aPP3. Different developmental patterns for average performance and performance consistency on basic auditory tasks during adolescence. Julia Jones Huyck and Beverly A. Wright (Dept. of Commun. Sci. and Disord., 2240 Campus Dr., Northwestern Univ., Evanston, IL 60208-3550)

Learning patterns on auditory discrimination tasks have been observed to differ markedly between average performance (average threshold) and performance consistency (within-subject standard deviation), suggesting that these two measures of performance may be governed by different processes [Wright *et al.*, ARO Abstr. **27**, 274 (2004)]. Further supporting this idea, different developmental patterns for average performance and performance consistency are reported here. Naive 8- to 10-year-olds ($n=6$), 11- to 14-year-olds ($n=16$), and adults (≥ 18 years old; $n=17$) were tested on four auditory detection conditions in which a 10-ms 1-kHz tone was presented (1) immediately before a 300-ms broadband noise, (2) at noise onset, (3) 200 ms after noise onset, or (4) immediately after noise offset. On two of the four conditions (conditions 1 and 4), both average performance and performance consistency improved with increasing age. However, on the remaining two conditions, either average performance improved significantly with increasing age while performance consistency did not (condition 2) or vice versa (condition 3). These results suggest that, in development, the processes underlying average performance may differ from those underlying performance consistency, and, more generally, that different processes may underlie these two measures of performance in many aspects of behavior. [Work supported by NIH/NIDCD.]

8:25

3aPP4. Estimating the rate of response growth using above characteristic-frequency maskers. Magdalena Wojtczak and Neal F. Viemeister (Dept. of Psych., Univ. of Minnesota, 75 East River Rd., Minneapolis, MN 55455, wojtc001@umn.edu)

Growth of forward masking was measured to determine the relative rate of basilar-membrane (BM) response growth at the characteristic-frequency (CF) place that was stimulated with a higher-frequency tone. Two signal frequencies (f_s) were used, 1 and 4 kHz. The masker frequency was $1.1f_s$ or $1.2f_s$. A lower-frequency masker ($0.6f_s$) was also used to replicate previously published data. Previous data obtained using a method for deriving the BM input-output function based on temporal masking curves suggested a linear growth of the BM response to a tone with a frequency equal or higher than $1.1f_s$ and a compressive growth for the CF tone. This study demonstrates that the growth-of-masking method for deriving the BM response provides a similar result (indicating a compressive growth for the CF tone) when a low-frequency masker is used but leads to a fundamentally different result (indicating an expansive response growth) when a higher-frequency masker is used. The expansive growth is observed even when a low-pass noise is used to limit potential off-frequency listening. The data will be related to mechanical BM data and the equivalence of the two psychophysical methods for deriving BM compression will be discussed. [Supported by NIH Grants DC00683 and DC006804.]

8:40

3aPP5. Forward masking of spectral modulation. Elizabeth M. O. Borchert, Magdalena Wojtczak, and Neal F. Viemeister (Univ. of Minnesota, 75 E. River Rd., Minneapolis MN 55455)

Some physiological and psychophysical evidence indicates that spectral shape is processed by means of a Fourier-like analysis of the spectral envelope. A forward masking paradigm in the spectral envelope domain was used to further investigate the properties of such spectral analysis. Detectability of a 50-ms sinusoidal modulation (ripple) in the amplitude spectrum of a broadband noise was measured when it was preceded by either uniform-spectrum noise, or by a noise with a same-frequency ripple (the forward masker). Data showed considerable forward masking of ripple detection when the phase of the masker and that of the signal were the same, but little or no masking when the masker and signal ripples were out of phase. Such results could be explained either by perceptual confusion between the signal and masker or by the adaptation of phase-sensitive, tuned ripple channels. In a follow-up study, the masker consisted of a broadband noise with both an on-frequency and an off-frequency (nonmasking) ripple. Little or no masking was observed in this study, which is consistent with an explanation based on the perceptual similarity between signal and masker. [Work supported by NIH Grant DC00683.]

8:55

3aPP6. Temporal integration functions of AM detection and AM depth discrimination. Jungmee Lee (Dept. of Speech, Lang., and Hearing Sci., Univ. of Arizona, P.O. Box. 210071, Tucson, AZ 85721-0071, jmlee6@email.arizona.edu), Glenis Long, and Changmo Jeung (CUNY, New York, NY 10016)

Previous studies (Lee, 1994; Lee and Bacon, 1997) suggested that both AM rate and AM depth discrimination were influenced by the number of AM cycles, instead of by the duration of stimulus. AM detection and AM depth discrimination were measured as a function of the number of AM cycles for the modulation rates of 10, 20, 40, 80, 125, 160, and 320 Hz. The carrier was a broadband noise (10 kHz low pass), and the carrier was either gated with the modulator (gated carrier) or presented 250 ms earlier and 250 ms later than the modulator (continuous carrier). The overall level of each presentation was randomized within 6-dB range from 65 dB SPL. For the lower modulation rates (10, 20, 40, and 80 Hz), there was a significant improvement with the continuous carrier at smaller number of cycles (two and four modulation cycles) as compared to the gated carrier condition. For the higher modulation rates (125, 160, and 320 Hz), however, the thresholds were not affected by carrier type. These results suggest that there might be a different temporal integration process for lower and higher modulation rates. [This work was supported by NIDCD Grant No. R03 DC06605-02.]

9:10

3aPP7. Response-related patterns in discrimination of FM narrow-band noise. Jinyu Qian and Virginia Richards (Dept. of Psych., Univ. of Pennsylvania, 3401 Walnut St., Ste. 302C, Philadelphia, PA 19104)

Noise stimuli were used to investigate weighting patterns for detecting dynamic frequency changes. For one stimulus set, a tonal glide was modulated by a narrow band of noise (gliding narrow-band noise, GNBN, condition). For the second set, the glide was preceded and followed by steady-state tones before modulation (SS condition). The task was to indicate whether the stimulus had a frequency glide or not. Weighting patterns (as a function of time) were derived for each subject based on a linear classification model. First, the deviation of the instantaneous carrier for each stimulus was calculated and then weighted by the squared envelope of the stimulus. This was followed by the linear classifications of the squared-envelope-weighted deviation of instantaneous carrier (SEWDOIC) according to the subjects responses [A. Ahumada, J. Vis., **2**, 121–131 (2002)]. The weighting patterns were consistent across subjects although individual differences in pattern details were observed. The comparison between the patterns from the GNBN and SS conditions indicates the use of dynamic information rather than only comparing the beginning and ending frequencies, consistent with Lyzenga *et al.* [J. Acoust. Soc. Am. **116**, 491–501 (2004)]. Similar weighting patterns were also estimated when the center frequency of the stimuli was random.

9:25

3aPP8. Revisiting psychometric functions for frequency discrimination. Huanping Dai (Div. of Commun. Disord., Univ. of Wyoming, Laramie, WY 82071)

In studies of auditory detection and discrimination, psychometric functions not only describe listeners sensitivity to changes in the stimulus, but often provide information regarding the auditory neural coding mechanisms involved in these tasks. In the present study, psychometric functions for frequency discrimination were systematically measured with a group of normal-hearing young adults. Frequency difference limens were in agreement with previous reports in their dependence on frequency. On the other hand, unlike that implied in previous analyses, the slopes of the psychometric functions, expressed in $\log(d)$ versus $\log(\Delta F)$, were found to be consistently shallower than unity at frequencies below 4 kHz. Only at higher frequencies did the slopes average around unity. These results will be discussed in light of auditory neural computational models based on either place or temporal coding. [Work supported by the Division of Communication Disorders, University of Wyoming.]

9:40

3aPP9. Suppression of sports chants through delayed feedback of sound. Sander J. van Wijngaarden and Johan S. van Balken (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands)

A novel approach for disrupting offending chants at sporting events is proposed, based on attacking synchronization between individuals. Since timing is crucial for coordination between chanters, disruption of timing is expected to be effective against chants. Delayed auditory feedback is known to disrupt timing in individuals; hence, delayed feedback may also be effective in a group context. It is shown that the timing of individuals joining in with sports chants can be severely disrupted by also presenting an artificially delayed version of this chant. This effect is reduced if the individual is given more cues (direction, fidelity) to differentiate between original chant and distracter. Unsurprisingly, the effectiveness also depends on delay time and the target-to-distracter ratio. The feedback signal has to be at least as loud as the original chant, and the delay should be greater than 200 ms. Conventional auditory feedback shows a pronounced peak in effectiveness around 300 ms. Such a peak does not occur here; up to approximately 1 s, effectiveness increases monotonically with delay time. Field experiments show that it may be hard to implement the approach suggested here in practice, since the required feedback signal levels tend to lead to instability of the feedback loop.

9:55–10:10 Break

10:10

3aPP10. Testing the role of harmonic number in the contribution of a nonsimultaneous mistuned harmonic to residue pitch. Hedwig E. Gockel, Robert P. Carlyon (MRC Cognition and Brain Sci. Unit, 15 Chaucer Rd., Cambridge CB2 2EF, UK, hedwig.gockel@mrc-cbu.cam.ac.uk), and Christopher J. Plack (Lancaster Univ., Lancaster LA1 4YF, UK)

Ciocca and Darwin [J. Acoust. Soc. Am. **105**, 2421–2430 (1999)] reported a surprising finding: Mistuning the fourth harmonic produced the same residue pitch shift when presented after the remainder of the complex as when it was simultaneous. They concluded that partials are integrated over a period of 170–250 ms for the calculation of pitch. Gockel *et al.* [J. Acoust. Soc. Am. **118**, 3783–3793 (2005)] failed to replicate this result, but mistuned the third harmonic. Here, subjects adjusted a matching sound (harmonic complex) so that its pitch equaled that of a subsequent 90-ms complex tone (12 harmonics of a 155-Hz $F0$) containing a mistuned ($\pm 3\%$) third or fourth harmonic. Pitch shifts were significantly

larger when: (i) the mistuned harmonic was presented simultaneously with, rather than after, the remainder of the target complex; (ii) the fourth rather than the third harmonic was mistuned; (iii) the nominally mistuned harmonic was present in, rather than absent from, the matching sound. The results again contrast with Ciocca and Darwin's findings and contradict the idea of obligatory integration over a long time period for the formation of residue pitch. [Work supported by EPSRC Grant EP/D501571/1.]

10:25

3aPP11. Onset transient significance in listener identification of pipe organ stops. Alastair C. Disley and David M. Howard (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK, acd500@ohm.york.ac.uk)

Pipe organs typically contain many ranks of pipes, but it is uncertain whether listeners can discern the often subtle differences between different ranks. The importance of onset transients in timbral identification is well established, but in many pipe organ listening situations, the transient is obscured. This paper explores the significance of organ pipe onset transients in an experiment based on Berger's work on orchestral timbre (K. W. Berger, J. Acoust. Soc. Am. **36**, part II, 1888 (1963)] Listeners familiar with the pipe organ attempted to identify recordings of single organ pipes, representing commonly found types of organ stop. Each pipe was presented on different occasions with its own transient and with artificial transients created by amplitude ramps. The results demonstrate that when onset transients are present, expert listeners are adept at identifying the general tonal nature of a stop, but are less adept at precise identification. Replacement of transients with amplitude ramps resulted in a decrease in both precise and general correct identification. The subjective mapping demonstrated between pipe names and their spectral features is explored. Onset transients are shown to be important in listener identification of pipe organ stops.

10:40

3aPP12. Spectral correlation of timbral adjectives used by musicians. Alastair C. Disley, David M. Howard, and Andrew D. Hunt (Dept. of Electron., Univ. of York, Heslington, York, YO10 5DD, UK, acd500@ohm.york.ac.uk)

As part of a project to develop a synthesis interface which nontechnical musicians should find intuitive, the adjectives musicians use to describe timbre have been studied in a large-scale listening test. Timbral adjectives were gathered from both musicians and previous studies and refined in a pilot experiment to a set of 15 (bright, clear, dull, evolving, gentle, harsh, metallic, nasal, percussive, pure, rich, ringing, thin, warm, and wooden). Fifty-nine English-speaking musicians used these words as rating scales (e.g., bright to not bright) in a listening experiment involving 12 single-note samples chosen to explore the timbre space of Western musical instruments. Twenty-three of those listeners were in a control group who took the test ten times, showing good consistency in their answers. The results demonstrate a good common understanding of adjectives between listeners, and multivariate data analysis techniques were used to exclude poorly understood or highly correlated adjectives. The relationship between the consistently understood adjectives and the spectral features they describe is explored by looking for correlation with a large number of time and frequency domain analyses of the 12 instrument samples. Significant correlation is found between some timbral adjectives and spectral features.

10:55

3aPP13. Speeded detection of sound signals based on temporal differences. Clara Suied, Laurent Pruvost, Patrick Susini, Nicolas Misdariis (IRCAM-CNRS, 1 place Igor Stravinsky, 75004 Paris, France, clara.suied@ircam.fr), Sabine Langlois (Renault, 78288 Guyancourt, France), and Stephen McAdams (McGill Univ., Montreal, QC, Canada H3A 1E3)

Previous research has identified acoustic parameters modulating the perceived urgency of alarms. Such approaches have focused mainly on the subjective relationship between characteristics of the alarms and the perception of urgency. We suggest that an objective measurement will ensure with more certainty the effectiveness of the acoustic parameters and will help us to understand more precisely the underlying processes. We have to the improve the ability of warning signals to increase the probability of a faster reaction under emergency conditions. Thus, we performed two experiments using a speeded reaction-time (RT) paradigm, under two tracking task conditions: motor and visual. Experiment 1 studied the influence of the tempo of the sequences. RT decreased as tempo became faster. Experiment 2 studied the influence of temporal irregularity. Outlier analyses show that RT decreased with the irregularity of the sequences. A scaled judgment of urgency was also given by subjects. Two underlying processes seem to be involved: experiment 1 highlights expectancy effects (temporal regularities provide a temporal reference frame), whereas experiment 2 emphasizes capture effects (a temporally perturbed event is more salient than a temporally expected one) [Jones, M.R., ''Attention and timing,'' in *Ecological Psyhoacoustics*, edited by J. Neuhoff (Elsevier Acadamic Press, San Diego, 2004) pp. 49–88 (2004)]. Moreover, comparisons between RT measurements and subjective judgments improve the validity of our approach.

11:10

3aPP14. Investigating auditory induction without complete continuity. Justin Aronoff (Neurosci. Program, Univ. of Southern California, HNB 27, Los Angeles, CA 90089, aronoff@usc.edu)

Auditory induction, the process of restoring missing auditory information, plays a crucial role in perception. A number of researchers have examined the characteristics of this phenomenon when there is a signal anchoring both sides of the induced sound (i.e., complete anchoring). However, the characteristics of partially anchored induced sounds (i.e., anchoring on only one side of the induced sound) are largely unknown, although prior research suggests that such partially anchored induction can occur [Wrightson and Warren, J. Acoust. Soc. Am. **69**, S105–S106 (1981)]. To examine the characteristics of this latter type of induced sound, synthetic stop-initial syllables were used in a forced-choice task, allowing easy assessment of the type and amount of induced information. The first 25 ms of each transition was removed, and noise (used to elicit the induced sound) was added to the truncated transition in small segments, replacing the removed transition. Accuracy increased with increasing amounts of noise, indicating that induction was occurring with partial anchoring. However, this induction was strongly limited, extending only 15 ms into the noise. This suggests that complete anchoring plays an important role in extending the induced sound, although it is not necessary for its creation.

11:25

3aPP15. Using psychometric functions to determine auditory perception of the size and speed of a rolling ball. Christophe Stoelinga, Dik Hermes (Technische Universiteit Eindhoven, Den Dolech 2, P.O. Box 513, 5600MB Eindhoven, The Netherlands), and Armin Kohlrausch (Philips Res. Labs. Eindhoven, 5656AA Eindhoven, The Netherlands)

The purpose of the presented experiments is to quantify exactly how precise listeners can judge certain mechanical parameters of a system by listening to the sounds generated by this system. We used sounds generated by a ball rolling over a plate. In the recording of these sounds, the ball diameter, rolling speed, and plate thickness were varied. Subjects were asked to judge each of these three parameters in the sound in two different ways; in a paired comparison, two alternative forced choice, and using absolute magnitude estimation. Both methods were used to calculate psychometric functions which estimate the probability that a listener correctly identifies the difference in one of the physical parameters. Both paired comparison and absolute magnitude estimation lead essentially to the same psychometric functions. For each pair of sounds the percentage of correct responses is plotted against the fraction of their sizes. These datapoints are then fitted to a cumulative normal distribution. The good fit of the results, using this logaritmic scale for the estimated parameter, is in close agreement with Steven's law. This method was used to test whether if discrimination capabilities decrease when two parameters, eg. speed and size, are changed simultaneously.

11:40

3aPP16. Effect of context on identification of environmental sounds. Brian Gygi (Acoust. Res. Inst. of the Austrian Acad. of Sci., Reichratsstrasse 17, Vienna 1010 Austria) and Valeriy Shafiro (Rush Univ. Medical Ctr., Chicago, IL)

This study investigated the role of auditory context in identifying individual environmental sounds. Limited previous research has not demonstrated a facilitatory influence of context on sound identification comparable to the benefits grammatical context provides for identifying speech sounds. Unlike some previous studies, in which context was determined by sound presented sequentially before and after the target environmental sound, in the present experiments the sounds to be identified were mixed in with a background of naturally occurring auditory scenes, such as the ambience of a beach or a street. Thirty-one familiar environmental sounds from previous environmental sounds studies were presented mixed either with congruent scenes (such as hammering at a construction site) or incongruent scenes (such as a horse galloping in a restaurant). The sounds were presented at signal-to-background ratios of −12 and −16 dB and trained listeners identified them using three-letter codes. Results indicate auditory context had a small but significant effect on identification performance. However, contrary to expectations, the sounds were identified more accurately in incongruent scenes (55% correct) rather than congruent ones (46% correct). The differential effects of auditory attention in determining foreground and background sounds will be discussed in connection with the results.

11:55

3aPP17. Relationships of performance accuracy, target call sign sensitivity, and response bias on the Coordinated Response Measure task. Valeriy Shafiro (Commun. Disord. & Sci., Rush Univ. Medical Ctr., 1653 W. Congress Pkwy., Chicago, IL 60612), Brian Gygi (Acoust. Res. Inst., Vienna, Austria), and Hillary Snapp (Rush Univ. Medical Ctr., Chicago, IL 60612)

Coordinated Response Measure tasks evaluate attentional processing by requiring the listener to mark the color-number response contained in a phrase that starts with a specific call sign. However, when a target call sign is present on every trial, it is not known whether subjects must perceive the call sign in order to respond correctly. The present experiment examined the relationships among sensitivity to the target call sign, response bias, and performance accuracy in 6 normal-hearing subjects over 4 testing sessions. The number of talkers (2–4) and talker sex configuration were manipulated in 9 stimulus conditions. In session 1, a target call sign was present on every trial. During sessions 2 and 4, 10% of the trials were foils with no target call sign, while the stimuli in the third session had an equal number of trials with and without the target call sign. Results indicate significant main effects of test session and sex configuration for all 3 experimental parameters. High correlations were observed between sensitivity and accuracy for all sessions. However, a significant improvement in sensitivity was not accompanied by a similar improvement in accuracy, suggesting that better call sign detection does not necessarily lead to better task performance.

WEDNESDAY MORNING, 7 JUNE 2006 ROOM 552B, 7:55 A.M. TO 12:00 NOON

Session 3aSA

Structural Acoustics and Vibration: Ultrasonic Waveguides for Structural Monitoring

Joel M. Garrelick, Chair

Applied Physical Sciences, Inc., 4 Muzzey St., Lexington, MA 02421

Chair's Introduction—7:55

Invited Papers

8:00

3aSA1. Monitoring of bulk materials using ultrasonic waveguides. Michael Lowe, Peter Cawley, and Francesco Simonetti (Dept. of Mech. Eng., Imperial College, London SW7 2AZ, UK)

Much progress has been made in the use of guided waves for the detection of cracks and corrosion in structures such as pipelines, railway lines, and steel reinforcing tendons. The challenge with those ideas is to use waves which are sensitive to the defects and insensitive to any materials in contact with the structure. However, a complementary approach is to identify waves which are particularly sensitive to materials in contact with the waveguide and thus to identify their presence and properties. The authors have worked on a number of interesting applications of this kind, including the characterization of sludge and detection of blockages in pipelines, the *in situ* measurement of the material properties of soil, the monitoring of the cure of adhesives, and the measurement of the properties of complex fluids. The paper will discuss the physical phenomena which determine how the guided waves in bars, pipes, and plates are affected by materials with which they are in contact, and give examples of how these phenomena can be exploited.

8:30

3aSA2. Guided waves for stress corrosion crack detection in pipelines—feature selection and classification. Austin P. Albright, Venugopal K. Varma, Raymond W. Tucker, and Philip R. Bingham (Oak Ridge Natl. Lab., 1 Bethel Valley Rd., Bldg. 3500, Rm. C028, Oak Ridge, TN 37831-6010, albrightap@ornl.gov)

This paper describes a wavelet-based analysis of electromagnetic acoustic transducer (EMAT) signals for in-line inspection of flaws in natural gas pipeline. The main objective of the project has been to implement the use of EMATs for pipe flaw detection, specifically the ability to detect stress corrosion cracks (SCCs) that are undetectable by current techniques. In this approach, two EMATs are used; one is the transmitter, while the second one, located a few inches away from the first, is used to receive the induced signal. Using a four-level wavelet decomposition, the EMAT data are filtered based on frequency. The features used to classify are derived from the coefficients representing each level of the four-level decomposition of the signature. The objective of the project was

to detect SCC with minimal false positive even if smaller SCCs (shallow) are not identified. Although many features could be used, selecting the right features that results in maximum separation between the classes (SCC flaw, other pipe artifacts, and no flaw) was a challenge. This paper describes the process of down-selecting the feature sets and separating the classes. The results using this approach have shown promise.

9:00

3aSA3. Decay curvature due to "losses" into a waveguide from a reverberant body. Richard Weaver (Dept. of Theoretical and Appl. Mech., Univ. of Illinois, 216 Talbot Lab., 104 S. Wright, Urbana, IL 61801)

We study the dissipation of diffuse ultrasonic energy in a reverberant body coupled to an effectively infinite elastic waveguide. A simple model predicts a distribution of complex eigenfrequencies and a corresponding nonexponential (''curved'') dissipation. This behavior was predicted in its broad outlines by Manfred Schoeder and is largely confirmed by measurements. For the case of fully open channels, however, measurements deviate from the simple model to a statistically significant degree. A random matrix supersymmetric calculation is found to accurately model the observed behaviors at all coupling strengths. We discuss implications for modeling losses in large complex systems and, in particular, for modeling decay curvature.

9:30

3aSA4. Waveguide scattering from flanges, ribs, and defects in fluid-loaded, laminated cylinders. Ann Stokes, Charles Corrado (Appl. Physical Sci., 4 Muzzey St., Lexington, MA 02421), and Gregory McDaniel (Boston Univ., Boston, MA)

The vibration of laminated, fluid-loaded cylinders is efficiently predicted with a wave transmission line analysis. The transmission line approach expresses the motion of homogeneous cylinder sections in terms of propagating and evanescent waves. The full set of wave dispersion relationships is computed with a semi-analytic technique, which uses a finite element radial discretization of the laminated shell and interior fluid. External fluid is represented using a state-space representation of the rational function approximation (RFA). Scattering of waves from discontinuities, including flanges, delamination, and pitting, is predicted through coupling a two-port admittance of the structure or defect to the semi-infinite admittance of the cylinder waveguide. Issues regarding the measurement of scattered and reflected waves, and identification of structural wave numbers, are discussed. The use of causality and reciprocity constraints substantially reduces measurement requirements and noise. By exciting and measuring a mixture of structural waves, defects may be classified and located. An example of predicting and measuring scattering of a defect beyond a flange is presented. [Work sponsored by the Office of Naval Research, Code 33.]

10:00–10:15 Break

10:15

3aSA5. Leaky wave excitation on concave surfaces and hyperbolic Fresnel zones. Philip L. Marston (Phys. and Astron. Dept., Washington State Univ., Pullman, WA 99164-2814)

Leaky waves are waves guided by elastic structures having phase velocities exceeding the speed of sound in the surrounding fluid. When analyzing the excitation of leaky waves on smoothly curved elastic structures it is often helpful to determine the Fermat paths. Those paths satisfy Fermat's principle for the combined acoustic propagation in the fluid and guided propagation on the structure. It is also helpful to determine the shape of the Fresnel zones associated with deviations from Fermat paths. That information is directly useful for calculating guided leaky wave amplitudes [P. L. Marston, ''Spatial approximation of leaky wave surface amplitudes for three-dimensional high-frequency scattering: Fresnel patches and applications to edge-excited and regular helical waves on cylinders,'' J. Acoust. Soc. Am. **102**, 1628–1638 (1997)]. The Fresnel zones associated with the external excitation of helical leaky waves on submerged circular pipes or cylinders were previously shown to be elliptical. The resulting leaky wave amplitudes are in agreement with amplitudes derivable by other methods and with experiments. In the present work, the analysis is generalized to excitation of leaky waves by ultrasound incident from the concave side of cylindrical shells or pipes. Even though the resulting Fresnel zones are approximately hyperbolic in shape, ray approximations for the amplitude are applicable when expressed using the Hessian of the phase deviation. [Supported by the Office of Naval Research.]

Contributed Papers

10:45

3aSA6. Selective modal excitation using phase-shifted ultrasound radiation force. Thomas M. Huber (Phys. Dept., Gustavus Adolphus College, 800 College Ave., St. Peter, MN 5608), Mostafa Fatemi, Randall R. Kinnick, and James F. Greenleaf (Mayo Clinic College of Medicine, Rochester, MN 55905)

The ultrasound radiation force can be used as a noncontact method for exciting vibrational modes. In one implementation, an audio-range modulation signal is combined with a 40-kHz carrier to form a double-sideband suppressed-carrier (DSSC) waveform that is emitted from an ultrasound transducer. The frequency of the ultrasound radiation force is the difference between the frequencies of the sidebands, therefore at twice the modulation frequency. We describe a method for selective mode excitation using a transducer pair, each emitting a DSSC signal. The modulation signal for one transducer was phase shifted relative to the other transducer. When the two modulation signals were in phase, the radiation stress field was uniform. For devices studied, this excited transverse modes while suppressing torsional modes. Similarly, if the modulation signals were out of phase, torsional modes were excited while suppressing transverse modes. This technique has been demonstrated for devices including an MEMS mirror, hard-drive suspension, and brass cantilever. For a cantilever, the ratio of the amplitudes of 180-Hz transverse over 1.3-kHz torsional modes was varied by over two orders of magnitude by shifting the phase difference between the two beams over a range of 90 deg. This noncontact technique allows control over the types of vibration excited.

11:00

3aSA7. Ultrasonic nondestructive evaluation in thin-walled concrete for flaw detection. Michael K. Pedrick, Bernhard R. Tittmann (Dept. of Eng. Sci. and Mech., The Penn State Univ., 212 EES Bldg., University Park, PA 16802, mkp136@psu.edu), and Shivprakash Iyer (The Penn State Univ., University Park, PA 16802)

Accurate inspection techniques for today's infrastructure have become an area of great interest. Several ultrasonic techniques for testing and evaluating concrete have been established and show great promise. However, much of the work is concerned with concrete piles or thick-walled specimen (greater than 500 mm). Vast amounts of concrete piping in sewer systems and water mains require testing techniques applicable to thinner-walled systems (less than 100 mm). This work aims to extend ultrasonic inspection techniques to thin-walled systems. Impact-echo and resonant ultrasonic spectroscopy (RUS) techniques are explored. Finite-element models have been developed to describe the propagation characteristics in different specimen. Several specimens have been tested experimentally to determine the effects of steel reinforcements, pipe curvature, and simulated defects. The influence of sensor configuration, transducer characterization, and data capture on measurement accuracy and inspection time is examined. The potential for wavelet analysis is explored for isolation of wave propagation modes and defect characterization.

11:15

3aSA8. Nondestructive inspection of in-service pipelines by guided ultrasonic waves. Hasson M. Tavossi (Dept. of Physical and Environ. Sci., Mesa State College, 1100 North Ave., Grand Junction, CO 81501)

Ultrasonic techniques for inspection of pipelines are investigated, using them as channels for guided ultrasonic waves. The current standard method for pipe inspection requires sending a rotating probe inside the pipe and measuring the wall thickness by pulse-echo technique. The inside wall of the pipe should therefore be accessible, clean, and smooth. These requirements limit the applicability of this method. The inconvenience of the above method can be removed by the guided wave technique that uses the pipe itself as a waveguide. In the guided wave technique the waves are not reflected from the inside surface of the pipe and they do not have the restrictions mentioned above. It is well known that dispersion of ultrasonic guided waves occurs in the wall of the pipe. As the waves propagate along the length of the pipe, different modes travel at different speeds. Selection of an appropriate mode for the transmitted waves is used to detect various flows in the pipelines. These flaws include wall thinning, pitting by corrosion, or cracks. High-frequency ultrasonic transducers in the range of 10 to 20 MHz are used for this purpose. Theoretical results are discussed and compared with the experimental findings on the ultrasonic nondestructive in-service pipe inspections.

11:30

3aSA9. Time-reversal imaging for structural monitoring. Gregoire P. Derveaux (INRIA, domaine de Voluceau, BP105, 78153 Le Chesnay Cedex. France, gregoire.derveaux@inria.fr), George Papanicolaou (Stanford Univ., Standford, CA 94305), and Chrysoula Tsogka (Univ. of Chicago, Chicago, IL 60637)

We investigate time-reversal imaging with distributed sensors in order to locate localized damage in a structure. Given a configuration of transducers, we assume that a full response matrix for the healthy structure is known. It is used as a basis for comparison with the response matrix that is recorded when there is damage. We have carried out a numerical experiment with the wave equation in two dimensions. The healthy structure is a domain containing many scatterers. We want to image two pointlike defects with the help of 12 sensors regularly distributed. Because of the complexity of the environment, the traces have a lot of delay spread and travel time migration does not work well. Instead the traces are back-propagated numerically in the medium, assuming that we have some knowledge of the background. Since the time at which the back-propagated field will focus on the targets is unknown, we compute the Shannon entropy of the image and pick the time where it is minimal. The TV norm proves also a good indicator. Thanks to the superfocusing property of time reversal in a highly scattering medium, this algorithm works well: it gives reliable and stable results.

11:45

3aSA10. High-frequency thermoviscoelastic vibration of functionally graded thin beams. M. Cengiz Dkmeci (Istanbul Tech. Univ., P.K.9, Gmssuyu, 34430 Istanbul, Turkey, dkmeci@itu.edu.tr) and Glay Altay (Bogazii Univ., 34342 Istanbul, Turkey)

The paper presents the system of 1D equations so as to analyze the thermoviscoelastic behavior of an axially functionally graded beam of rectangular cross section at high-frequency vibration. First, the 3D fundamental equations of a hyperbolic thermoviscoelasticity are expressed by a unified variational principle in terms of the Laplace transformed field variables [cf., Dkmeci and Altay, Acta Mechanica **143**, 91–11 (2000)]. Next, the temperature increment and the displacement field as well as the material density and elasticities [e.g., Sankar and Tzeng, AIAA J. **40**, 1228–1232 (2002)] are expressed by the power series expansions in the thickness coordinate of the beam. A variational averaging procedure [e.g., Dkmeci and Altay, Int. J. Non-linear Mech. **37**, 225–243 (2002)], together with the series expansions, is used in deriving the system of 1D beam equations in both differential and fully variational forms. The system of 1D equations governs the extensional, flexural, torsional, and thickness shear and also the coupled vibrations of the beam at high frequency. Special cases are indicated involving the material properties and/or the vibration of the beam. Also, the uniqueness is investigated in solutions of the system of beam equations. [Work supported by TUBA.]

Session 3aSC

Speech Communication: Speech Processing and Special Populations (Poster Session)

Janet L. Slifka, Chair
Massachusetts Inst. of Technology, Research Lab. of Electronics, 55 Vassar St., Cambridge, MA 02139

Contributed Papers

All posters will be on display from 8:45 a.m. to 11:45 a.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:45 a.m. to 10:15 a.m. and contributors of even-numbered papers will be at their posters from 10:15 a.m. to 11:45 a.m.

3aSC1. Recording and environmental effects in sociolinguistic interviews: Implications for sociophonetic analysis. Tyler Kendall (English Dept., Duke Univ., Box 90017, Durham, NC 27708, tsk3@duke.edu)

Sociophonetians are often interested in examining speech that has been recorded in acoustically diverse conditions. Sociolinguistic methodology is centrally concerned with combating the observer's paradox, finding ways to minimize informant's awareness that they are under investigation, and often the requirements needed to ensure high-quality recordings conflict with this theoretical focus. While sociophoneticians are aware of some potential problems when comparing different speakers, with only rare exceptions have they examined the potential problems on both inter- and intraspeaker analyses due to recording and environmental effects. Meanwhile, other fields (especially computer speech recognition) have investigated the acoustic effects of environmental noise and recording equipment and have demonstrated that these factors impact the resulting speech signals in ways that are possibly significant for sociophonetics. This study emulates a number of traditional sociolinguistic interview recording techniques by taking an array of recordings for single interviews. Each interview is simultaneously recorded with a wired lavalier microphone connected to an audiocassette recorder, a unidirectional vocal microphone placed 2 ft from the speaker connected to an audiocassette recorder, and a wireless lavalier to a digital video camera. The results of phonetic analyses of these recordings, along with the implications for sociophonetic analysis, are explored.

3aSC2. Acceptable noise levels with intelligible and unintelligible primary discourse. Susan G. Hickey and Robert E. Moore (Dept. of Speech Pathol. and Audiol., Univ. of South Alabama, 2000 UCOM, Mobile, AL 36688-0002)

Acceptable noise levels (ANL) were assessed for 30 adult, normal-hearing listeners in sound field in three conditions. The three conditions were primary signal of (1) intelligible discourse (Arizona Travelogue); (2) unintelligible backwards discourse (Arizona Travelogue recorded backwards); and (3) unintelligible Chinese discourse. In each condition the primary speaker was male. ANL was determined for each of three conditions by having participants select their most comfortable level (MCL) for the primary discourse. Next, for each condition, speech babble was added and adjusted to find the participant's maximum background noise level (BNL) that was acceptable while listening to the primary discourse. The ANL was determined by subtracting the BNL from the MCL. The group mean ANLs for the intelligible discourse, the backwards discourse, and the Chinese discourse were 7.16, 8.64, and 9.31 dB, respectively. An ANOVA for repeated measures indicated a significant main effect for primary discourse type ($F=5.54$, $p=0.006$). *Posthoc* paired samples t-tests indicated a significant difference between the intelligible and unintelligible backwards discourse condition ($t=-2.47$, $p=0.020$) and the intelligible and unintelligible Chinese discourse condition ($t=-3.17$, $p=0.004$).

3aSC3. Using the reassigned spectrogram to obtain a voiceprint. Sean A. Fulop (Dept. of Linguist., California State Univ. Fresno, 5245 N. Backer Ave. PB92, Fresno, CA 93740-8001) and Kelly Fitz (Washington State Univ., Pullman, WA 99164-2752)

While established methods for imaging the time-frequency content of speech—such as the spectrogram—have frequently been christened ''voiceprinting,'' it is well-known that it and other currently popular imaging techniques cannot identify an individual's voice to more than a suggestive extent. The reassigned spectrogram (also known by other names) is a relatively little-known method [S. A. Fulop and K. Fitz, ''Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications,'' J. Acoust. Soc. Am. **119**, 360–377 (2006)] for imaging the time-frequency spectral information contained in a signal, which is able to show the instantaneous frequencies of signal components as well as the occurrence of impulses with dramatically increased precision compared to the spectrogram (magnitude of the short-time Fourier transform) or any other energy density time-frequency representation. It is shown here that it is possible to obtain a reassigned spectrogram image from a person's voice that appears to be sufficiently individuating and consistent so as to serve as a true voiceprint identifying that person and excluding all other persons to a high degree of confidence. This is achieved by focusing on just a few phonatory pulsations, thereby revealing the vocal-fold vibrational signature unique to each person.

3aSC4. A spike-based computational model for noise robust vowel classification. Ismail Uysal, Harsha Sathyendra, and John G. Harris (Computational Neuro Eng. Lab., The Univ. of Florida, Gainesville, FL 32611)

The human ability to recognize speech drastically outperforms that of commercial ASR systems especially in noisy environments. Presently, there is limited knowledge of the auditory system dynamics, however it is known that coding and processing of information is carried out via action potentials. This research aims to better understand the coding mechanisms along the auditory pathway, while devising a noise robust system for speech recognition. A biologically plausible algorithm for vowel classification is proposed, which solely uses spikes for both the feature extraction and the classification stages. The algorithm uses an improved and adaptive model of the inner-hair cell [Sumner *et al.*, J. Acoust. Soc. Am. **113**, 893–901 (2003)] to generate spike trains at different characteristic frequencies. The synchrony among the hair cells is used as a noise robust means for feature extraction. Detected features are then classified using a

spike-based rank order coder, which uses the spike arrival times to the postsynaptic neuron to encode information [Delorme and Thorpe, Neural Networks. **14**, 795–803 (2001)]. Experiments on a noisy vowel dataset (5 dB SNR) show an average of 15% increase in the recognition rate for the prototype system when compared to a nearest-neighbor classifier employing Mel frequency cepstral coefficients.

3aSC5. Noise robust bandwidth extension of telephone speech for mobile and landline communications. Harsha M. Sathyendra, Ismail Uysal, and John G. Harris (Computational Neuro Eng. Lab., Univ. of Florida, Gainesville, FL 32611)

The present day telephone infrastructure limits the bandwidth of telephone speech to 300–3400 Hz, whereby conversational speech falls mostly within 50–8000 Hz. The ensuing band-limited narrow-band (NB) speech has a considerable loss in sound quality, in terms of naturalness and intelligibility. The purpose of bandwidth extension (BWE) is to utilize the receiver end NB speech and to add relevant frequency information to the missing low and high frequencies. This novel BWE algorithm utilizes the linear source filter model, which separates NB speech into two roughly independent components, the excitation and the spectral envelope. These components are then extended and convolved together to attain reconstructed wideband (RWB) speech. Excitation extension involves harmonic reconstruction for voiced frames and some modulations and filtering steps for unvoiced frames. Spectral envelope extension utilizes a Gaussian mixture model trained via the expectation maximization algorithm, which is involved in the mapping of NB noise robust features to the corresponding RWB features. Noise analysis is used to isolate the best feature set for real world applications. Results from a subjective listening test indicate this algorithm has an improvement in sound quality, while having the lowest known distortion metric among other speaker independent bandwidth extension algorithms.

3aSC6. Acoustic parameters for nasality based on a model of the auditory cortex. Tarun Pruthi and Carol Y. Espy-Wilson (Dept. of Elec. Eng. and Inst. of Systems Res., Univ. of Maryland, College Park, MD 20742, tpruthi@umd.edu)

Due to the longer length of the nasal tract as compared to the oral cavity, the asymmetry of the left and right nasal passages, and the coupling of the nasal tract to the sinuses, nasal sounds generally have more formants than oral sounds. However, all of the formants may not be prominent in the acoustic spectrum due to close-lying antiresonances. The net effect of the nasal peaks can be an increase in frequency modulation. In this work, we study the effectiveness of two new acoustic parameters (APs) that capture this increase in frequency modulation based on the ''scale'' dimension of a model of the auditory cortex [Chi *et al.*, J. Acoust. Soc. Am. **118**(2), 887–906 (2005)]. The new APs, mean center of gravity of scales and mean ratio of high scales to low scales, were used along with a previous set of four APs for the discrimination of nasal consonants and semivowels. Overall accuracy of 93.10% was obtained for this task. This corresponds to an 18.63% reduction in error rate over previously reported results. It remains to be seen if these APs are also useful in detecting nasalization in vowels. [Work supported by NSF Grant No. BCS0236707.]

3aSC7. Automated measurement of nasality over time in vowels. Michael A. Berger, Meghan Clayards, Neil Bardhan, and Joyce McDonough (Dept. Linguist., Univ. of Rochester, Rochester, NY 14627, berger@ling.rochester.edu)

Nasalized vowels have acoustic characteristics that allow the degree of nasality to be quantified from the acoustic signal. These characteristics include decreased first-formant amplitude ($A1$) and increased amplitudes of two extra peaks ($P0$ and $P1$). Chen (1995, 1997, 2000) used the parameters $A1-P1$ and $A1-P0$ (dB) to quantify nasality in vowels. These two measures are complementary in that different vowel types may be more accessible to one or the other. This research attempts to extend Chen's measures in several ways: (1) normalize the two parameters so that they are comparable; (2) integrate them into a single composite measure by a weighted average; (3) measure nasality over the time domain of each vowel token, resulting in nasality contours with high temporal resolution; and (4) automate the process of measuring nasality in vowels, so that large amounts of speech data can be processed rapidly by computer. Nasality measurement in this framework, whether manual or automated, requires a solution to the problem of tracking the three relevant peaks over time: $F1$, $N0$, and $N1$. A deterministic peak-tracking algorithm is proposed, but alternative statistical approaches are also considered. The validity of a peak-based approach for female speech is examined.

3aSC8. Detection of vowels in long-term speech-shaped noise. Chang Liu (Dept. of Commun. Sci. and Disord., Wichita State Univ., Wichita, KS 67260) and David A. Eddins (Univ. of Rochester, Rochester, NY, 14616)

The method of calibration is a critical issue in studies of vowel identification in noise, especially when manipulating speech-to-noise ratio. Common methods focus on stable estimates of the intensity or sound pressure level for specific speech sounds. However, vowels with the same sound pressure level often differ in both loudness, when presented at supra-threshold levels, and audibility, when masked detection thresholds for the vowel stimuli are measured. In the present study, detection thresholds for 12 isolated American-English vowels naturally spoken by three male and three female talkers were measured for young normal-hearing listeners in the presence of a long-term speech shaped (LTSS) noise. The duration of the vowels was equalized to 170 ms. The spectrum of the LTSS noise was identical to the long-term average spectrum of a 12-talker babble. The LTSS noise was present at 70 dB SPL. Thresholds for vowel detection showed distinct pattern as a function of the vowel category that was similar across listeners and talkers. This indicates that certain vowels were more audible than others, regardless of listeners and talkers. To account for the vowel-detection thresholds in noise, various analyses including comparisons of excitation patterns and estimates of specific loudness will be considered.

3aSC9. Enhancement of spectral modulation frequency in acoustics and identification of vowels. David A. Eddins (Dept. of Otolaryngolgy, Univ. of Rochester, Rochester, NY, 14616 and Intl. Ctr. for Hear. and Speech Res., RIT, Rochester, NY 14623) and Chang Liu (Wichita State Univ., Wichita, KS 67260)

The pattern of intensity variations across audio frequency can be represented as a series of spectral modulation frequencies. Selectivity for spectral modulation frequency in the auditory system has been demonstrated by physiological and psychophysical studies, suggesting that auditory perception of complex sounds might be based on spectral envelope channels. The present study investigated the effects of enhancing spectral modulation frequency on vowel acoustics and vowel identification. Twelve naturally spoken American-English vowels by three female talkers served as the standard stimuli. Various spectral modulation frequency regions of the standard vowels combined with a long-term speech-shaped (LTSS) noise were enhanced. Identification performance for the vowels presented in LTSS noise with and without spectral modulation frequency enhancement was estimated for seven young normal-hearing listeners. Vowels were presented at a sensation level of 3 dB in the LTSS noise (70 dB SPL). Results indicated that vowel identification performance was improved on average by 4%–6% over the 12-vowel categories for spectral modulation enhancement between 1.0 and 2.5 cycles/octave. Acoustic analysis suggested that the improved vowel identification as a result of the global spectral modulation enhancement may be due to increases in the spectral contrast of individual formants.

3aSC10. Information conveyed by *f*0 for vowel identification. Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, Edmonton, AB, T6G 2E7, Canada) and Peter F. Assmann (The Univ. of Texas at Dallas, Richardson, TX, 75083)

Our recent experiments with vocoded natural speech, wherein the spectral envelope and fundamental frequency are manipulated independently, have confirmed that some coordination of *f*0 and formant patterns are beneficial to vowel identification by humans. In an effort to model the perceptual dependency more precisely, we have investigated the performance of several alternative pattern recognition models on natural speech samples. This paper reports on several quite distinct methods of exploiting statistical relations between formant frequencies and *f*0 for recognition. Many of these methods yield quite similar results on the classic Peterson and Barney data and on larger, more recently collected data sets. Methods involving indirect normalization whereby the *f*0 of a single token is restricted to the role of estimating the formant frequency average of a speaker's entire vowel system perform well. Indeed, they are often better than a method where the role of *f*0 is unconstrained, thus accommodating inherent pitch differences among vowels. The indirect use of *f*0 also allows for methods of combining *f*0 and formant range information in ways that preliminary results suggest to be more effective for modeling perceptual effects with modified stimuli. More formal evaluation against perceptual data will be presented.

3aSC11. Recurrence quantification analysis of speech. Aaron Schultz (Deptartment of Psych., Univ. of Connecticut, 406 Babbidge Rd., Unit1020, Storrs, CT 06269-1020)

A time-series analysis technique known as recurrence quantification analysis (RQA) holds promise as a new method for investigating the dynamical properties of the acoustic speech signal. RQA was created to deal with complex time series generated by high-dimensional, nonlinear dynamical systems. While RQA has most frequently been used for physiological studies and investigations into nonlinear dynamics, it has not been widely applied to the acoustic analysis of speech. Recurrence quantification analysis is free of many key assumptions involved in Fourier-type analyses such as data set size, stationarity, and linearity, which are routinely violated in the speech stream. This study offers some preliminary findings on the dynamic nature of speech and characteristics of speaker variability obtained by using recurrence quantification analysis. Future directions for the use of RQA in speech perception and production studies, including shadowing and imitation studies, are discussed.

3aSC12. Predicting changes in breathiness resulting from variations in aspiration noise and open quotient using a loudness model. Arturo Camacho (Comput. and Information Sci. and Eng., Univ. of Florida, Gainesville, FL 32611) and Rahul Shrivastav (Univ. of Florida, Gainesville, FL 32611)

Previous studies have reported that increases in high-frequency aspiration noise (AH) and open quotient (OQ) are correlated with an increase in the perception of breathiness in vowels [Klatt and Klatt, J. Acoust. Soc. Am. **87**, 820–857 (1990)]. However, a system based on psychoacoustic loudness models has been shown to predict the level of breathiness in vowels to a greater extent than AH and OQ themselves [Shrivastav and Sapienza, J. Acoust. Soc. Am. **114**, 2217–2224 (2003)]. The goal of this experiment was to verify the significance of AH and OQ in the perception of breathiness and to determine how each of these influences the output of the loudness model, individually and in combination. Ten vowels, five male and five female, were synthesized using a Klatt synthesizer. Each vowel was used to generate three continua varying in terms of their OQ and/or AH. Ten listeners rated breathiness for each continuum using a magnitude estimation task. Results show that normalized partial loudness as well as the loudness of aspiration noise predicted changes in breathiness resulting from an increase in AH. Increase in OQ resulted in very small changes in perceived breathiness, resulting in poor correlation with loudness measures. [Research supported by NIH/R21DC006690.]

3aSC13. Towards a probabilistic system for estimation of acoustic landmarks for speech recognition. Chiyoun Park and Janet Slifka (Speech Commun. Group, RLE, MIT, Cambridge, MA 02139, goodie@mit.edu)

Algorithms for consonant landmark detection, such as Liu [S. Liu, J. Acoust. Soc. Am. **100**, 3417–3430 (1996)], extract cues to specific types of abruptnesses in the acoustics. The abruptnesses indicate occurences of closure and release for obstruent and sonorant consonants, and burst release for stop consonants. In Liu's algorithm, fixed thresholds are used to filter out abruptnesses that are unlikely to be true landmarks. The resulting set of landmarks does not retain any information regarding these filtered-out instances. However, such information may be useful later in the lexical access process, especially given the range of contextual variation in the speech signal. In this work, the landmark detection process is reformulated as a probabilistic system. First, thresholds are lowered to include more candidates, and then a probability value is calculated for each candidate. An *N*-best search is used to pick the most likely sequences of obstruent landmarks based on the calculated probabilities. Experiments with 80 sentences from the TIMIT database detect corresponding landmarks within 40-ms windows of 96% of hand-labeled obstruent landmarks, 98% of burst-release landmarks, and 76% of sonorant landmarks. Applying 5-best search results in 9% deletion and 4% insertion rate. [Work supported by NIH DC02978.]

3aSC14. Evaluation of token-based acoustic measures for classification of irregular phonation in normal speech. Janet Slifka and Kushan Surana (Speech Commun. Group, Res. Lab. of Elec., MIT, Cambridge, MA 02139, slifka@speech.mit.edu)

Quantification of the acoustic characteristics of irregular phonation provides a foundation for automatic detection of regions of irregular phonation in continuous speech. Recent results for automatic classification of regions of phonation as either regular or irregular demonstrate classification rates greater than 90% (false positive $<10\%$) [K. Surana, M.Eng. thesis, MIT, Cambridge, MA 2006]. Similar acoustic cues may be useful in separating subtypes of irregular phonation. Two types of irregular phonation are examined: (1) regions characterized by reduced airflow, assumed to correspond to tightly adducted vocal folds with brief regions of separation, and (2) regions characterized by increased airflow, assumed to correspond to a spread or spreading vocal-fold configuration [J. Slifka, J.Voice (in press)]. Reduced-airflow tokens are extracted, using airflow, audio, and electroglottography signals, from utterance-medial locations, and increased-airflow tokens are all utterance final (20 tokens/speaker, 4 speakers). Surana's cue set was evaluated for the ability to separate these subtypes of irregular phonation. Preliminary results indicate that an energy difference measure (between tokens smoothed at 6 vs 16 ms) and normalized rms amplitude yield statistically separable populations. Limitations on the use of these cues are discussed including intra- and interspeaker variation and the effect of token duration. [Work supported by NIH DC02978.]

3aSC15. Effects of spectral slope on perceived breathiness in vowels. Mario A. Landera and Rahul Shrivastav (Dept. of Commun. Sci. and Disord., Univ. of Florida, P.O. Box 117420, Gainesville, FL 32611)

Previous studies have indicated that breathiness is correlated with changes in spectral slope as well as aspiration noise [Klatt & Klatt, "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857 (1990)]. In a recent study, Shrivastav and Pinero [Shrivastav & Pinero "Effect of aspiration noise and spectral slope on perceived breathiness in vowels," J. Acoust. Soc. Am. **117**, 2622 (2005)] observed that varying spectral slope resulted in minimal changes on the perceived breathiness for synthetic vowels. However, the stimuli tested in this experiment included a relatively narrow range of spectral slope variation. The goal of the present experiment was to verify the role of spectral slope on the perception of breathiness by testing stimuli that had a wider range of variation in spec-

tral slope. Ten voices (5 male and 5 female) were synthesized using a Klatt synthesizer. Each of these was manipulated to generate two continua varying in their spectral slope from −3 to −30 dB/octave. One continuum had a constant cutoff frequency of 500 Hz, while the other continuum had a cutoff frequency between the second harmonic (H2) and the third harmonic (H3). Ten listeners judged the degree of breathiness using a 7-point rating scale. Preliminary results suggest that spectral slope has minimal effect on the perception of breathiness in synthetic vowels. [Research supported by NIH/R21 DC006690.]

3aSC16. Perceptual judgments of breathy voice quality using rating scale, magnitude estimation, and matching tasks. Sona Patel, Rahul Shrivastav (Dept. of Commun. Sci. and Disord., Univ. of Florida, Gainesville FL 32611), and David A. Eddins (Univ. of Rochester, Rochester, NY)

Experiments to study voice quality have typically used rating scales to obtain listener judgments. Recent research has questioned the validity of this data due to poor agreement between listeners [Kreiman *et al.*, J. Acoust. Soc. Am. **87**, 820–857 (1993)]. However, Shrivastav *et al.*, J. Acoust. Soc. Am. **117**, 2622 (2005) showed that such variability was partly related to the experimental methodology and could be minimized by multiple presentations of each stimulus to each listener. The nature of this averaged data, whether ordinal or interval, needs to be determined since it is necessary to develop appropriate perceptual scales for voice quality. The present experiment used three different psychophysical methods to obtain perceptual estimates of voice quality. These include a rating scale (with single and multiple presentations of stimuli to each listener), a direct magnitude estimation task, and a matching task. The experiment sought to determine if these three methods provided equivalent information. Perceptual judgments of breathiness were obtained for ten synthetic vowel continua that varied in terms of their aspiration noise, open quotient, or both. Preliminary results show that perceptual judgments using the magnitude estimation task and rating scale task (with multiple presentations of each stimulus) give very similar information. Results of the matching task will also be discussed. [Research supported by NIH/R21DC006690.]

3aSC17. Acoustic source measures for documenting outcomes of combined modality treatment of adductor spasmodic dysphonia. Erin Silverman, Rahul Shrivastav, and Christine Sapienza (Dept. of Commun. Sci. and Disord., Univ. of Florida, 335 Dauer Hall, Gainesville, FL 32608, ejpear@hotmail.com)

Spasmodic dysphonia (SD) is a voice disorder of neurogenic origin, wherein one or more intrinsic laryngeal muscles exhibit irregular and uncontrollable spasms. These spasms interfere with the production of perceptually normal voice. This paper presents a prospective, randomized, and placebo-controlled study to examine the effects of phonatory retraining (voice therapy) following Botox injections for ADSD. Thirty subjects were randomly assigned to 1 of 3 experimental groups: Botox only, Botox plus voice therapy, and Botox plus sham therapy. The sham therapy group was included in order to examine nonspecific treatment effects which may emerge as a result of close patient-therapist interaction during voice therapy for ADSD. Dependent variables of interest include acoustic source measures as well as voice-related quality of life, subject self-rating of voice, and duration of benefit between initial and repeat Botox injections. Source measures were typed according to National Center for Voice and Speech and signal typed into type 1, 2, and 3 signals. Voice breaks and frequency shifts were also measured. Preliminary results indicate that the combined modality treatment of voice therapy plus Botox has the more significant positive effect on source measures with trend toward more normal phonation or type 1 signal.

3aSC18. Minimizing occlusion: An evaluation of open-fit versus deep-canal fit hearing aids. Kristin A. Vasil and Kathleen M. Cienkowski (Dept. of Commun. Sci., Univ. of Connecticut, 850 Bolton Rd., Storrs, CT 06269, kristin.vasil@uconn.edu)

Individuals with good low-frequency hearing and poor high-frequency hearing may complain that when wearing hearing aids, their voice sounds boomy or hollow [B. C. Grover and M. C. Martin, Audiology **18**, 335–350 (1979)]. This is known as the occlusion effect. This effect occurs due to an improvement of bone-conduction responses when the ear canal is covered or blocked [D. P. Goldstein and C. S. Hayes, J. Speech Hear. Res. **8**, 137–148 (1965)]. The purpose of this study was to determine the degree of measured and self-rated occlusion for hearing aids that attempt to minimize occlusion by utilizing open-fit or deep-canal fit designs. Measured occlusion was calculated from aided and unaided real ear responses obtained while subjects vocalized [i]. Self-ratings were acquired using an occlusion effect scale. Results showed that open-fit and deep-canal fit hearing aids differ in degree of measured and self-rated occlusion. Hearing aids with open-canal characteristics had less measured occlusion, according to real ear measures, than the deep-canal fit device. Self-ratings of occlusion were correlated with objective real ear measures. For deep-fitting devices, optimum insertion depth had to be reached in order for maximum occlusion reduction to be achieved. [Work supported by a grant from Vivatone LLC.]

3aSC19. Acoustical differences in mothers' speech to children with cochlear implants versus normal-hearing children. Rose Bradley (Commun. Disord., Butler Univ., 4600 Sunset Ave., Indianapolis, IN 46208-3485), Tonya R. Bergeson (Indiana Univ. Med. School, Indianapolis, IN 46202), and Carolyn Richie (Butler Univ., Indianapolis, IN 46208-3485)

Infant-directed speech includes higher pitch overall and more exaggerated pitch contours compared to adult-directed speech. Clear speech is characterized by increased duration and expanded vowel space compared to conversational speech. This study investigated the differences between mothers' speech to children with cochlear implants versus normal-hearing children in terms of acoustic characteristics associated with infant-directed speech and clear speech. Ten mothers of children with cochlear implants, 12- months postimplantation, and 25 mothers of normal-hearing, 12-month-old children were recorded during a 5-min play session between mother and child. Pitch range, vowel duration, and vowel space in mothers' speech to children with cochlear implants and normal-hearing children were examined. Results are discussed in terms of differences between groups depending upon the hearing status of the child, and in terms of the acoustic characteristics of infant-directed speech and clear speech directed toward hearing-impaired listeners. [Work supported by NIH-NIDCD Training Grant T32DC00012 to Indiana University and NIH-NIDCD Research Grant R21DC06682 to the Indiana University School of Medicine, HUG grant Butler University.]

3aSC20. Sentence comprehension when formant transitions are present or absent by normal-hearing and hearing-impaired listeners. Jae Hee Lee and Diane Kewley-Port (Dept. of Speech and Hearing Sci., Indiana Univ., Bloomington, IN 47405)

Burkle et al. [''Contribution of consonant versus vowel information to sentence intelligibility by normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **115**, 2601 (2004)] reported that young normal-hearing (YNH) and elderly hearing-impaired (EHI) listeners recognized sentences better with only vowels compared to only consonants in a noise replacement paradigm. Using the same paradigm, this experiment examines whether the 2:1 benefit was due to the type of information presented in whole segments or to the proportion of total sentence duration. Forty-two sentences were processed to present only steady-state or dynamic transition information, rather than whole segmental information. YNH and EHI listened to sentences with 50% and 70% proportion of total sentence du-

ration, and the number of words that listeners correctly repeated in sentences was scored. There was no effect of processing conditions. For sentences with 50% and 70% proportions, YNH recognized 60% and 84% of the words across conditions, while EHI recognized only 16% and 45%. The benefit ratio for the 70% proportion compared to 50% was 0.4 for YNH and 3.7 for EHI. YNH improved linearly with increases in proportional duration regardless of the type of information presented. EHI improved with proportional duration, but less systematically, and were more affected by the type of information. [Supported by NIHDCD-02229.]

3aSC21. Infant's perception of speech in noise: Effect of the number of background talkers. Rochelle S. Newman (Univ. of Maryland, 0100 Lefrak Hall, College Park, MD 20742, rnewman@hesp.umd.edu)

A number of recent studies have investigated infants' abilities to hear speech in noise [Newman & Jusczyk, ''The cocktail party effect in infants,'' Percep. Psychophs. **58**, 1145–1156 (1996); Newman, R. S., ''The cocktail party effect in infants revisited: Listening to one's name in noise,'' Develop. Psychol. **41**, 352–362 (2005); Barker & Newman, ''Listen to your mother! The role of talker familiarity in infant streaming,'' Cognition, **94**, B45–B53 (2004); Hollich, *et al.*, ''Infants' use of synchronized visual information to seperate streams of speech,'' Child Develop. **76**, 598–613 (2005)]. In the current study, we examine how the type of background noise influences infants' ability to understand speech. Infants aged 5 months heard a talker repeat either their own name or another infant's name in the presence of one of 3 types of background noise: multitalker babble, a single background talker, or a single talker reversed in time. With multitalker babble, infants could recognize their name (shown by longer listening to own-name than other-name test trials) only when the target speech was at least 10 dB more intense than the background. However, even at this high SNR, infants fail to recognize their name when the background consists of a single talker. This difficulty is not the result of semantic information or speechlike properties from the single talker; infants likewise fail when the single talker is temporally reversed. Rather, infants appear to be distracted by signals containing the time-varying amplitude patterns of a human voice, and do not appear to benefit from the dips in noise amplitude that occur with a single voice.

3aSC22. Acoustics of clear speech from children with normal hearing and cochlear implants. Andrew J. Lotto (Univ. of Texas at Austin, 1 University Ave. A8000, Austin, TX 78712-0187, andrew@mail.cps.utexas.edu), Dana L. Ide-Helvie, Elizabeth A. McCleary, and Maureen B. Higgins (Boys Town Natl. Res. Hospital, Omaha, NE 68131)

It is common among clinicians to ask children to produce their best speech during intervention. However, it is unclear whether children know how to respond to these directions and whether any contingent changes in their production actually result in better speech. Normal and best speech for children with normal hearing (NH) and cochlear implants (CI) were compared on a variety of acoustic measures. The children (7 to 14 years of age) were asked to read ten simple sentences to train up a fictional automatic speech recognizer for children. They subsequently produced the same sentences with their best and then very best, very clearest speech (ostensibly to test the recognizer). Sentence, pause, and vowel duration were calculated for all sentences as well as vowel intensity and fundamental frequency. Additionally, phonetic distinctiveness was evaluated through vowel space size and voice onset time differences between voiced and voiceless stops. NH children demonstrated patterns of best speech that resemble adult clear speech productions. CI children's speech resembled NH best speech in all elicitation conditions, except that they slowed (their already slower) speaking rate for best speech. The results indicate that clear speech can be produced by children. [Work supported by NIH.]

3aSC23. Vowel production with and without auditory feedback in children with cochlear implants. Sneha V. Bharadwaj and Amanda G. Graves (Callier Ctr. for Commun. Disord., Univ. of Texas at Dallas, 811 Synergy Park Blvd., Richardson, TX 75080, snehab@utdallas.edu)

The effects of auditory feedback on vowel formant frequencies were investigated in the speech produced by three early-implanted (received cochlear implants before the age of 3 years) and three late-implanted children (after 5 years). The syllables /heed/, /hid/, /had/, /whod/, /hood/, and /hod/ were produced in a carrier sentence 12 times, each in both implant processor ON (auditory feedback present) and OFF (feedback absent) conditions. Comparisons of the vowel formant frequencies in processor ON versus OFF conditions revealed primarily second formant frequency ($F2$) shifts for a majority of the vowels, while showing minimal changes in the first formant frequencies ($F1$). In addition, a majority of the $F2$ shifts was towards the middle of the vowel space, suggesting centralization of vowels in the absence of auditory feedback. Further, most of the formant frequency changes were present in the vowels produced by late- compared to early-implanted children. These findings suggest that the late-implanted children may rely on auditory feedback to a greater extent than early-implanted children for speech production. Implications of the short-term effects of auditory feedback on speech production will be discussed in reference to the feedback and feed-forward models for speech production. [Work supported by NIH/NIDCD R03DC007052-01.]

3aSC24. Normalizing pitch variability in intoxicated speech. Ryan Hanke and Thomas Purnell (Dept. of Linguist., Univ. of Wisconsin, Madison, 1168 Van Hise, 1220 Linden Dr., Madison, WI 53706)

Intoxicated speech studies have shown mixed results with respect to prosodic characteristics. Both an overall $F0$ rise [Hollien *et al.*, J. Acoust. Soc. Am. **110**, 3198–3206 (2001)] and $F0$ lowering [Johnson *et al.*, Phonetica **47**, 215–237 (1990)] have been observed under intoxication. These conflicting findings may be due to increased $F0$ fluctuations at various levels of intoxication and use of inconsistent prosodic domains, such as averaging over entire sentences or utterances. If phonetic analysis takes into account prosodic boundaries, then models will account for appropriate prosodic rises and falls [Taylor, J. Acoust. Soc. Am **107**, 1697–1714 (2000)]. This paper reports an analysis of recordings of both intoxicated and control subjects. Speech samples are analyzed at sober and three levels of intoxication. Of particular interest are the prosodic patterns at the often unreported descending limb of intoxication, which displays the widest range of variability in behavior. Use of $F0$ parameter extraction, stylization, and rotation methods over appropriate prosodic domains factors out spurious variation at different intoxication levels. Evaluation of this normalization procedure is reported with respect to both within- and across-subject variation. Implications for the analysis of normal, intoxicated, and disordered speech are discussed.

Session 3aSP

Signal Processing in Acoustics and Underwater Acoustics: Optimum and Robust Signal Processing in Uncertain and Random Environments I

Ning Xiang, Cochair
Rensselaer Polytechnic Inst., Architecture, 110 8th St., Troy, NY 12180

R. Lee Culver, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Chair's Introduction—8:25

Invited Papers

8:30

3aSP1. Overview of signal processing in uncertain and random environments. Leon H. Sibul (Appl. Res. Lab. and Grad. Program in Acoust., Penn State Univ., P.O. Box 30, State College, PA, 16804, lhs2@psu.edu)

This paper is an overview of optimum and/or robust signal-processing approaches for detecting signals in random and uncertain propagation and interference environments, a topic that is intertwined with statistical modeling of the medium and signals. The classical approach to processing of signals that have propagated through randomly fluctuating media is based on the concepts of randomly time-varying impulse responses, spreading functions, and scattering functions. This approach uses second-order statistics and gives insight into performance of active sonar. Implementation of maximum likelihood (ML) detectors for signals that have propagated through random environments requires knowledge of the probability density functions (pdf's) of the random signal and noise parameters. Random signal parameters that are typically affected by medium randomness are amplitude, phase, and arrival time; pulse and Doppler spread; arrival angle bias and spread. Signal pdf's can be derived from ocean acoustic models using the maximum entropy (ME) principle, which exploits what is known but is maximally noncommittal of what is uncertain. The ME method can be used to generate densities that belong to the exponential class, and for this class, ML detectors can be implemented as an estimator-correlator and estimator-noise canceler structure. [Work supported by The Undersea Signal Processing Program Office of Naval Research.]

9:10

3aSP2. Bayesian geoacoustic inversion of time-averaged horizontal-array data. Stan E. Dosso (Univ. of Victoria, Victoria, BC V8W 3P6 Canada) and Dag Tollefsen (Norwegian Defence Res. Establishment, 3191 Horten, Norway)

This paper considers quantifying data errors in Bayesian geoacoustic inversion applied to time-averaged data. Cross-spectral density matrices are formed by averaging spectra from a sequence of time-series segments (data snapshots). Error estimation for snapshot-averaged data has typically assumed either that averaging reduces errors as if they are fully independent between snapshots (an optimistic assumption), or that averaging does not reduce errors at all (a pessimistic assumption). Data errors are quantified here assuming that averaging reduces measurement error (dominated by ambient noise, which can be reasonably assumed independent), but does not reduce theory (modeling) error. This provides a physically reasonable intermediate result between the optimistic and pessimistic assumptions. Bayesian inversion is applied to data collected by FFI with a bottom-mounted horizontal array at a shallow-water site in the Barents Sea. Supporting geophysical measurements (seismic reflection and refraction, bottom-penetrating sonar, gravity core) provide independent information on seabed properties. A towed acoustic source transmitted multiple low-frequency tones at levels comparable to those of a merchant ship. Inversion results in the form of marginal posterior probability distributions are compared for the different approaches to data error estimation, and for data collected for several source ranges and bearings.

9:30

3aSP3. Use of prior information in acoustic inference. Kevin H. Knuth (Dept. of Phys., Univ. at Albany, Albany, NY 12222, kknuth@albany.edu)

Often in scientific and engineering applications, we possess prior information about the specific problem under investigation. In acoustics, this prior information may consist of knowledge about the design and arrangement of our detectors, the behavior of the sound sources, and the laws of physics describing sound propagation through an intervening medium. In this talk I explore the use of prior information for use in Bayesian inference in acoustic applications. This prior information can be either encoded in the signal model, which is represented in the Bayesian formulation by the likelihood function, or in the prior probabilities, which weight the particular values of the signal model parameters. I give examples of the use of such prior information by focusing on the specific problems of source separation and localization. By choosing to focus on the appropriate model parameters one can change a source separation problem into a source localization problem and vice versa, thus changing the very nature of the acoustic inference.

9:50

3aSP4. Bayesian model comparison and selection in energy decay analysis of acoustically coupled spaces. Tomislav Jasa and Ning Xiang (Grad. Program in Architectural Acoust., School of Architecture, Rensselaer Polytechnic Inst. Troy. NY 12180)

The use of Bayesian inference in problems of parameter estimation from noisy data and in uncertain environments has been well discussed in the acoustical signal processing literature. In many acoustical problems where it is uncertain which suitable model among a set of competing ones should be used, the model comparison and selection become crucial prior to the actual parameter estimation. However, tools for model selection in Bayesian inference have received less attention. In problems of model comparison and selection, the Bayesian methodology is most different from orthodox statistical methods. We present several methods for model comparison and selection from a Bayesian viewpoint and demonstrate their application to the problem of estimating the number of decay rates present in acoustically coupled spaces.

10:10–10:25 Break

10:25

3aSP5. Applications of Markov chain Monte Carlo methods in ocean acoustics. David Battle (MIT Ctr. for Ocean Eng., 77 Massachusetts Ave., Cambridge, MA 02139, dbattle@mit.edu)

Markov chain Monte Carlo (MCMC) methods have recently been applied to ocean-acoustic parameter estimation problems, as well as to model selection, wherein different parametrizations can be compared to assess which model most efficiently describes acoustic field measurements at a given signal-to-noise ratio. This talk will deal with some ocean-acoustic applications of Monte Carlo parameter estimation, including approaches to calculating Bayesian evidence integrals for model selection. The talk will conclude with a discussion of known defects in algorithms currently available and some possible future directions for algorithm development.

10:45

3aSP6. Optimal source localization in the presence of interfering sources. Zoi-Heleni Michalopoulou (Dept. of Mathematical Sci., New Jersey Inst. of Technol., Newark, NJ 07102, michalop@njit.edu)

Source localization in shallow-water ocean environments is hindered by the presence of loud interfering sound sources. To remediate these nuisance sources, one needs to back out the true signal from the received mixed acoustic field by estimating a most likely interference scheme. In this work the focus is on resolving problems introduced by an unknown number of interfering signal sources. A statistical model for the sensed acoustic field is estimated using Gibbs sampling in which one obtains through iteration an approximation to the full joint posterior probability description of the model and data. From this joint probability description, the impact of canceling undesirable field components is readily seen on the estimated uncertainty that characterizes the localization process. [Work supported by ONR.]

Contributed Paper

11:05

3aSP7. Using the maximum entropy method to estimate probability density functions for oceanographic and acoustic parameters. H. John Camin, R. Lee Culver, Leon H. Sibul, Jeffrey A. Ballard, and Colin W. Jemmott (Appl. Res. Lab. and Grad. Program in Acoust., P.O. Box 30, State College, PA 16804)

The maximum entropy (MaxEnt) method offers a powerful means of estimating the probability density functions (pdf's) of measured data. Traditionally, pdf's have been estimated by fitting data to particular distributions (e.g., Gaussian, log-normal, Rayleigh, chi-squared, etc.). This approach can lead to misleading results, especially if the data are highly skewed or kurtotic. First published in 1957 by Jaynes [E. T. Jaynes, Phys. Rev., **106**, 620–630 (1957)], MaxEnt has gained wide use in recent years. The method uses Lagrangian multipliers to maximize the *entropy*, a measure of uncertainty, subject to specified constraints. The result is a pdf that Jaynes describes as being the ''least biased estimate possible on the given information; i.e., it is maximally noncommittal with regard to missing information.'' In other words, all possibilities under the given constraints are considered equally. This method is particularly useful in determining prior pdf's based on limited measurements. MaxEnt pdf's, which belong in the exponential class of distributions, can also be used to derive optimum processors in the maximum likelihood sense. This talk will discuss the theory and implementation of the MaxEnt method and emphasize its utility for modeling acoustic propagation in uncertain environments. [Work supported by The Undersea Signal Processing Program, Office of Naval Research.]

Session 3aUW

Underwater Acoustics: Propagation

Kathleen E. Wage, Chair
George Mason Univ., Dept. of Electrical and Computer Engineering, 4400 University Dr., Fairfax, VA 22030

Contributed Papers

7:40

3aUW1. Polynomial chaos expansions for coupled mode amplitude and phase uncertainty in the presence of oceanographic uncertainty. Kevin D. LePage (Naval Res. Lab., Code 7144, 4555 Overlook Ave. SW, Washington, DC 20375)

Polynomial chaos (PC) techniques have recently been introduced to the field of underwater acoustics for modeling the nonlinear transfer of probabilistic measures of the propagation environment into measures of acoustic propagation uncertainty [Finette, J. Acoust. Soc. Am. **117**, 997–1000 (2005)]. Recently we have proposed and implemented a PC method for modeling this process for coupled mode propagation through internal waves (IWs) [LePage, J. Acoust. Soc. Am. **118**, 1903–1904 (2005)]. While showing promise, the approach was implemented for a single empirical orthogonal function (EOF) component of the IW field only and was derived for small PC coefficients. In this paper the restriction of small coefficients is removed for the first two terms of the PC expansion, and results for uncertainty caused by more than one uncorrelated EOF component are also presented. [Work supported by ONR.]

7:55

3aUW2. Stochastic differential equation analysis for ocean acoustic energy scattering by internal waves. Andrey K. Morozov (Woods Hole Oceanograph. Inst., Woods Hole, MA, 02543) and John A. Colosi (Naval Postgrad. School, Monterey, CA 93943)

Stochastic coupled mode theory is used to study acoustical scattering of a narrow-angle, low-frequency acoustical beam. Random index of refraction perturbation fields are described using the Garrett-Munk internal wave spectrum. The stochastic differential equations for the first- and second-order statistical moments of the acoustical field are derived using the Markov approach similar to Dozier, Tappert, and Chernov theories. The main objective of the research is to account for cross-mode correlations and to show that the method can be used for a calculation of the detailed structure of the mean acoustic field intensity distribution. For this purpose a narrow finite-frequency beam was taken as an example, and the intensity of the scattering field was calculated as a function of frequency and sound source depth. The coefficients of the matrix differential equation were obtained analytically, and an efficient numerical computation was obtained. Comparisons to Monte Carlo simulation show that the method can be used for detailed statistical sound field analysis in long-range propagation.

8:10

3aUW3. Mean, variance, and temporal coherence of the 3D field forward propagated through random internal waves in deep-ocean waveguides. Tianrun Chen (MIT, 77 Massachusetts Ave., Cambridge, MA 02139), Purnima Ratilal (Northeastern Univ., Boston, MA 02115), and Nicholas Makris (MIT, Cambridge, MA 02139)

An analytical expression for the temporal coherence of an acoustic signal propagated through a 3D random internal-wave field in a deep-ocean waveguide is derived. The modeled coherence time scale of the acoustic fluctuations at 3000-km range is on the order of 10 min and matches the ATOC experimental measurement. It is found that the 3D scattering effects on the forward field become pronounced when the Fresnel width exceeds the cross-range coherence length of the internal-wave field. The depth-integrated intensity losses due to internal waves first increase then decrease as frequency decreases from 75 to 10 Hz. These results are derived by applying a general modal solution for the statistical moments of an acoustic field propagating through 3D random inhomogeneities [Ratilal and Makris, ''Mean and covariance of the forward field propagated through a stratified ocean waveguide with three-dimensional random inhomogeneities,'' J. Acoust. Soc. Am. **118**, 3532–3559 (2005)] to the acoustic propagation through a random internal-wave field. The mean and variance of the forward field are expressed in terms of the stochastic moments of the inhomogeneous medium's scatter function density. This is derived by applying the Rayleigh-Born approximation to Green's theorem in terms of the medium's compressibility and density fluctuations, which are related to the displacement of internal waves. The statistical properties of the displacement of internal waves are formulated with the Garret-Munk spectrum.

8:25

3aUW4. The vertical structure of shadow-zone arrivals at long range in the ocean. Lora J. Van Uffelen, Peter F. Worcester, Matthew A. Dzieciuch, Daniel L. Rudnick, Bruce D. Cornuelle, and Walter H. Munk (Scripps Inst. of Oceanogr., Univ. of California at San Diego, 9500 Gilman Dr., Mail Code 0238, La Jolla, CA 92093, lvanuffe@ucsd.edu)

Multimegameter-range acoustic data obtained by bottom-mounted SOSUS receivers show significant acoustic energy penetrating several hundred meters into the geometric shadow zones below cusps in the predicted acoustic time fronts, much further than can be accounted for by diffraction [Dushaw *et al.*, IEEE J. Oceanic Eng. **24**, 202–214 (1999)]. Acoustic data from two vertical line array receivers deployed in close proximity in the North Pacific Ocean, together virtually spanning the water column, show similar shadow-zone arrivals for transmissions from broadband 250-Hz sources moored at the sound-channel axis (750 m) and at surface conjugate depth (3000 m) at ranges of 500 and 1000 km. The vertical structure of the receptions will be compared to parabolic equation propagation simulations for sound-speed fields constructed from high-resolution underway conductivity-temperature-depth (UCTD) data in the upper 400 m, widely spaced deep CTD casts, and climatological data. Predictions will be made (i) for the observed field; (ii) for a smoothed field, and (iii) for fields that include sound-speed fines structure due to density variations (internal waves) or due to density-compensated sound-speed variations (spice).

8:40

3aUW5. Propagation in highly range-dependent shallow-water waveguides with elastic sediments and beach interactions. Jon M. Collis, William L. Siegmann (Rensselaer Polytechnic Inst., Troy, NY 12180, collisj@rpi.edu), and Michael D. Collins (Naval Res. Lab., Washington, DC 20375)

Accurate and efficient prediction of propagation over realistic models of elastic sediments has been achieved recently using parabolic equations. A major advance is the variable rotated parabolic equation solution [Outing *et al.*, J. Acoust. Soc. Am. **115**, 2579 (2004)], which allows treatment of variable bathymetry. The bathymetry is modeled as a series of constant slopes, and within each region the parabolic equation is solved in a coor-

dinate system aligned with the interface. This approach has been benchmarked against analytic solutions and laboratory data. The variable rotated parabolic equation has been extended recently to solve problems involving variable topography (when the elastic sediment rises above the surface of the ocean). For these problems the surface boundary is modeled by a series of constant slopes. In this presentation we show how the transition from air-water to air-sediment surface boundary conditions, such as in beach environments, can be handled to high accuracy. In addition, an improved treatment of the boundary conditions at the air-sediment interface is described. Finally, procedures for handling elastic sediment interfaces of nonconstant thickness, along with range-dependent bathymetry, are developed and benchmarked. [Work supported by the Office of Naval Research.]

8:55

3aUW6. Effect of nonlinear internal waves on midfrequency acoustic propagation on the continental shelf. Frank S. Henyey, Dajun Tang, Kevin L. Williams, Ren-Chieh Lien (Appl. Phys. Lab., Univ of Washington, 1013 NE 40th St., Seattle, WA 92105), Kyle M. Becker, R. Lee Culver, Paul C. Gabel, Joy E. Lyons (The Penn State Univ., State College, PA 16804), and Tom C. Weber (The Penn State Univ., State College, PA 16804)

Packets of nonlinear internal waves and strong sound-speed variability due to neutrally buoyant intrusions were measured in detail using a towed conductivity, salinity, pressure (CTD) chain. The measurements were taken in October 2005, on the continental shelf south of Rhode Island. Acoustic propagation through the measured sound-speed field is modeled. The waves in October are qualitatively different from the more familiar summertime nonlinear waves. Further measurements using the CTD chain will be made in the summer of 2006, in association with acoustic propagation measurements.

9:10

3aUW7. Downslope propagation of adiabatic and stepwise coupled modes. Richard B. Evans (Scientific Applications International Corp., 23 Clara Dr., Mystic, CT 06355, richard.b.evans@saic.com)

Downslope propagation, from a shallow source, redirects steeply propagating modes toward the horizontal. Slope conversion is of particular interest for noise sources near the shelf break. Noise modeling procedures may rely on adiabatic modes for efficiency, but their ability to accurately handle downslope conversion is not well tested. The adiabatic approximation is derived from continuously coupled modes by assuming the coupling matrices are the identity matrix and by using an asymptotic approximation to solve the uncoupled ordinary differential equations for the modal coefficients. A simpler interpretation of the twofold adiabatic approximation is obtained in the context of stepwise coupled modes when the continuous differential equations are replaced by a single matrix recursion. The stepwise interpretation of the adiabatic approximation is used to evaluate its validity, for a shelf-break environment, in a side-by-side comparison with stepwise coupled modes. The adiabatic approximation partially accounts for downslope conversion, but vertical directivity, computed in with adiabatic modes and stepwise coupled modes, is not comparable. Parabolic equation calculations, which corroborate the stepwise coupled mode results, provide a practical alternative to adiabatic modes for ambient noise modeling in shelf-break environments.

9:25

3aUW8. Regions where transient signals are influenced between a source and receiver. John Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 240 S. 33rd St., Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu)

At infinite frequency, a medium only influences the waveform of a received signal emitted from a source along one or more infinitesimally thin rays. For transient signals at finite frequencies, an exact method is developed to compute the regions in a medium that significantly influence the received signal for any specified window of signal travel time. Results are (1) in homogeneous media, a significant portion of the signal originates from waves that traverse paths that are approximately integer and half-integer numbers of cycles greater than the straight path between source and receiver. Called ''constructive and destructive paths of influence,'' they do not obey Fermat's principle of path stationarity. (2) For reflection from a flat interface in a homogeneous medium, the received signal is significantly influenced by constructive and destructive paths of influence whose angles of incidence and reflection differ. (3) Near 100 Hz in an oceanic waveguide, the region of influence markedly departs from a ray. Near upper turning points, the influential region is flat for O(10) km instead of O(1) km for a ray. (4) The first Fresnel zone is an inappropriate scale to characterize the region of influence for transient signals near a steep ray in inhomogeneous media.

9:40

3aUW9. A near-field solution for underwater acoustics; a finite-difference solution to the Helmholtz equation. Catherine de Groot-Hedlin (Scripps Inst. of Oceanogr., 9500 Gilman Dr., UCSD, La Jolla, CA 92093)

A finite-difference method is derived for the frequency-domain acoustic wave equation as expressed in cylindrical coordinates. L'Hopital's rule is used to transform the singularity at $r=0$ into a determinate form. The resulting finite difference frequency domain (FDFD) algorithm can be used to solve for sound intensities in complex models that may include high material contrasts and arbitrary bathymetry. Solutions obtained using this method are presented and compared to wide-angle parabolic equation solutions and analytic solutions, where available. Four models are examined: a uniform half-space, a waveguide with a rigid bottom boundary, an upslope wedge with a source in the ocean column, and a downslope wedge with a source below the seafloor. Comparison with analytic results indicates that the FDFD method accurately solves for the acoustic wave-field at all propagation angles and is more accurate than the wide-angle parabolic equation method at low radial distances from the source. The FDFD method is more accurate in situations where high angle scattering occurs near the source.

9:55–10:10 Break

10:10

3aUW10. Bistatic invariance principle for multi-static active geometries. Lisa Zurk, Jorge Quijano (Portland State Univ., 1900 SW Fourth Ave., Ste. 160-17, Portland OR 97207, zurkl@cecs.pdx.edu), and Dan Rouseff (Univ. of Washington, Seattle, Washington)

The invariance principle describes the spatial interference patterns resulting from the coherent addition of propagating normal modes in the ocean channel. The relationship simplifies propagation structure into a scalar invariant which is approximately unity for many shallow-water environments. For passive sonar geometries (i.e., source to receiver) this valuable relationship has been extensively explored and applied to the interpretation of passive lofargrams. In active bistatic geometries, a transmitted pulse travels over two distinct propagation pathways (i.e., source-to-target and target-to-receiver). Assuming a distributed network (i.e., long-range propagation), each of these paths in isolation would produce interference patterns as dictated by the invariance principle. However, for bistatic active applications, the received pulse is a product of both propagation paths and a target-dependent scattering matrix which introduces mode coupling. For broadband waveforms, the question arises as to whether there is an observable interference pattern for a moving source. In this presentation, a derivation of the bistatic invariance principle is presented and discussed. Simulation results for the predicted time-frequency structure in a shallow-water channel are presented for specific sensor geometries and choices of target scattering matrix. Implications for application to multi-static active target tracking are presented in a companion paper.

10:25

3aUW11. Effect of shelf break front on acoustic propagation. James H. Miller, Gopu R. Potty, Kristy A. Moore (Dept. of Ocean Eng., Univ. of Rhode Island, 217 Sheets Bldg., Narragansett Bay Campus, Narragansett, RI 02882), and James F. Lynch (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

The effect of a shelf break front on acoustic propagation is discussed in the context of the Shallow Water Experiment (SW-06) in the New Jersey Shelf. The primary effect is a horizontal analog to the classical horizontal Lloyd's mirror effect, produced by the fact that fronts can totally internally reflect sound incident upon them at low grazing angles. The direct and reflected modal rays add up to produce an interference pattern in range. It has been shown [Lynch *et al.*, ''Consideration of finescale oceanography and 3D acoustics effects for the ESME sound exposure models,'' IEEE JOE (in press)] based on simple calculations that this frontal effect has the potential to increase the maximum insonification level up to 6 dB. Noise measurements analyzed using the SWARM -95 data also provide some evidence to this effect. The present study explores this effect in more detail using a horizontal ray-vertical mode approach. Calculations were made using environmental parameters corresponding to the New Jersey Shelf area, where field measurements will be made in 2006 to explore this effect. [Work supported by Office of Naval Research.]

10:40

3aUW12. Shallow-water waveguide characterization using an autonomous underwater vehicle-towed hydrophone array. Jason D. Holmes, William M. Carey (Boston Univ., 110 Cummington St., Boston, MA 02215), and James F. Lynch (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543)

Sound transmission results are reported for frequencies of 220, 415, 635, 823, 1031, and 1228 Hz in an iso-velocity waveguide 13 m deep with known bottom properties. The experiments were conducted in Nantucket Sound in proximity to the Lynch-Frisk (1982) experiment in an area surveyed prior to the experiments. A fixed source and an autonomous underwater vehicle six-element towed hydrophone array were used. The experimental track consisted of two 2–4-km straight legs with measurement of the bathymetry and currents performed by the vehicle. Transmission results show stable modal-multipath interference effects with predictable phase rate variations. Coherent processing using vehicle estimates of ground speed and position yields results close to theoretical expectations and indicates a constant depth stable tow was realized. Horizontal wavenumber spectra using coherent synthetic aperture Hankel transforms are presented and shown to be calculable and comparable to the previous work. The demonstration of this measurement capability provides for cost effective measurements of the sound field at sea.

10:55

3aUW13. Effect of wind-generated surface waves on vertical coherence of sound propagation in the East China Sea. Jie Yang, Ji-Xun Zhou, and Peter H. Rogers (School of Mech. Eng., Georgia Inst. of Technol., MS 209, Atlanta, GA 30339-0405, gte036y@mail.gatech.edu)

Literature has shown that wind-generated surface waves are an important factor in determining the spatial coherence of long-range sound propagation [Kuperman and Ingenito, J. Acoust. Soc. Am. **113**, 1178–1187 (1977); Zhou *et al.*, J. Acout. Soc. Am. **115**, 2550 (2004)]. During the ASIAEX in the East China Sea, wind speed and the rms waveheight were recorded between 29 May and 7 June along with acoustic measurements of broadband explosive sources. The rms surface waveheight varied between 0.1–0.6 m during the acoustic measurements. This paper analyzes the characteristics of vertical coherence of sound propagation under the influence of these surface waves. A model is built to explain observed characteristics of vertical coherence which incorporates coring analysis in the ASIAEX area. [Work supported by ONR.]

11:10

3aUW14. Propagation of modulated pulses in shallow water. Aijun Song, Mohsen Badiey, (Ocean Acoust. Lab., Univ. of Delaware, Newark, DE 19716), Boris Katsnelson, and Alexander Tshoidze (Voronezh State Univ., Voronezh 394006, Russia)

In this paper propagation of low-frequency pulses, carrying information in shallow water, is studied. Different types of information-carrying pulses are analyzed and compared. These include phase shift keying (PSK) signals, frequency shift keying (FSK) signals, and noncoherent signals with respect to their ability to information transfer. Numerical modeling of pulses propagation is carried out. Basic attention is paid to studying the influence of dispersive characteristics of a shallow water channel (boundary conditions, variation of properties along the track, internal waves, etc.) on the waveguide capacity for different types of pulses and frequencies. Complex space-frequency and time-frequency intensity fluctuations of a propagating signal allow us to consider shallow water as the unique information channel (medium). A model for numerical calculations is constructed on the basis of the SWARM95 experimental data. [Work was supported by ONR and RFBR.]

11:25

3aUW15. The effect of turbulence and acoustic waves on underwater acoustic propagation. Stefanie E. Wojcik, William W. Durgin, and Tatiana A. Andreeva (WPI, 100 Institute Rd., Worcester, MA 01609, stefanie@wpi.edu)

The primary objective of this study was to develop a predictive methodology for received signal variation as a function of ocean perturbations using a ray-based analysis of the effects of internal waves and ocean turbulence on underwater acoustic propagation. In the present formulation the eikonal equations are considered in the form of a second-order, nonlinear ordinary differential equation with harmonic excitation due to an internal wave. The harmonic excitation is taken imperfect, i.e., with a random phase modulation due to Gaussian white noise, accounting for both chaotic and stochastic behavior. Small scale and geostrophic turbulence are represented using the potential theory two-wave-number model for a vortex array. The focus is to numerically study the influence of these fluid velocity and sound speed fluctuations and the role of initial ray angle on underwater acoustic propagation to provide a realistic characterization of acoustic arrivals. Predicted arrival behavior is analyzed using ray trace, time-front, and phase plots for varying initial conditions. Regions of instability are identified using bifurcation and phase diagrams. This work leads to understanding how turbulence and internal waves affect sound transmission and utilizes the statistics of received signals to identify fluid mechanic phenomena that occurred along the sound channel.

11:40

3aUW16. Multiple frequency acoustic propagation through clusters of bubbles. Thomas C. Weber (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824), Anthony P. Lyons, and David L. Bradley (Penn State, State College, PA 16804)

A numerical study has been conducted in order to examine the effect of bubble clustering on acoustic propagation through bubble clouds. Commonly used effective medium theories make the assumption that the positions of bubbles are statistically independent from each other. It is plausible, however, that there are some scenarios in which this assumption does not apply. For example, bubbles may become preferentially concentrated or clustered as they cross fluid flow streamlines due to their density difference from that of the surrounding liquid. In order to help understand the acoustic implications of bubble clustering, propagation through two types of bubble clouds was investigated numerically. In the first type of bubble cloud the positions of the bubbles were statistically independent from each other. In the second type the bubbles were clustered. Both the average attenuation and the variance in the acoustic pressure amplitude were recorded at nine different frequencies for each type of bubble cloud. It was found that clustering has the effect of reducing the average attenu-

ation from that observed in the nonclustered bubble cloud while increasing the variance. This contrasting behavior suggests a method for detecting the presence of bubble clustering using multiple frequency acoustic propagation measurements.

11:55

3aUW17. Geometrical dispersion of broadband sound pulse in a three-dimensional shallow-water environment with a corrugated bottom. Frédéric Sturm (LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 36, avenue Guy de Collongue, FR-69 134 Ecully Cedex, France)

Numerical results of sound propagation in a three-dimensional oceanic waveguide with a sinusoidal (corrugated) bottom are reported. A broadband sound pulse with a center frequency of 25 Hz and a bandwith of 30 Hz is considered. This acoustic problem was studied by others considering a harmonic point source [M. D. Collins and S. A. Ching-Bing, J. Acoust. Soc. Am. **87**(3), 1104–1109 (1990)]. The numerical method used to solve the 4-D acoustic problem is based on a Fourier synthesis of frequency-domain solutions. The calculations in 3-D are carried out using the fully 3-D parabolic equation based model 3DWAPE. To analyze the acoustic problem, we follow closely the methodology used in previous works to investigate the 3-D ASA wedge and the 3-D Gaussian canyon [F. Sturm, J. Acoust. Soc. Am. **117**(3), 1058–1079 (2005)]. Results corresponding to a 25-Hz cw point source are first presented and compared with predictions by other models. Then, the acoustic problem is solved considering the broadband source pulse. The modal structure of the received signals on several distinct vertical arrays is analyzed and clearly exhibits mode arrivals of the propagating signal not predicted by pseudo-3-D or 2-D models.

WEDNESDAY MORNING, 7 JUNE 2006 PROVIDENCE BALLROOM I (WESTIN), 8:30 TO 10:00 A.M.

Meeting of Accredited Standards Committee (ASC) S2 Mechanical Vibration and Shock

R. J. Peppin, Chair S2
5012 Macon Road, Rockville, MD 20852

D. J. Evans, Vice Chair S2
National Institute of Standards and Technology (NIST), 100 Bureau Dr., Stop 8220, Gaithersburg, MD 20899

Accredited Standards Committee S2 on Mechanical Vibration and Shock. Working group chairs will report on the status of various shock and vibration standards currently under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAG for ISO/TC 108, Mechanical vibration and shock, and its five subcommittees, take note – that meeting will be held in conjunction with the Standards Plenary meeting at 10:30 a.m. on Tuesday, 6 June 2006.

Scope of S2: Standards, specifications, methods of measurement and test, and terminology in the field of mechanical vibration and shock, and condition monitoring and diagnostics of machines, including the effects of exposure to mechanical vibration and shock on humans, including those aspects which pertain to biological safety, tolerance and comfort.

WEDNESDAY MORNING, 7 JUNE 2006 PROVIDENCE BALLROOM I (WESTIN), 10:30 A.M. TO 12:00 NOON

Meeting of Accredited Standards Committee (ASC) S3 Bioacoustics

R. F. Burkard, Chair S3
Hearing Research Lab., State Univ. of New York at Buffalo, 215 Parker Hall, Buffalo, NY 14214

C. A. Champlin, Vice Chair S3
Univ. of Texas, Dept. of Communication Sciences & Disorders, CMA 2-200, Austin, TX 78712

Accredited Standards Committee S3 on Bioacoustics. Working group chairs will report on the status of standards under development. Consideration will be given to new standards that might be needed over the next few years. Open discussion of committee reports is encouraged.

People interested in attending the meeting of the TAGs for ISO/TC 43 Acoustics and IEC/TC 29 Electroacoustics, take note – those meetings will be held in conjunction with the Standards Plenary meeting at 10:30 a.m. on Tuesday, 6 June 2006.

Scope of S3: Standards, specifications, methods of measurement and test, and terminology in the fields of psychological and physiological acoustics, including aspects of general acoustics, which pertain to biological safety, tolerance and comfort.

Session 3pAA

Architectural Acoustics: Session in Honor of John Kopec

David L. Moyer, Chair
Riverbank Acoustical Labs., 1512 S. Batavia Ave., Geneva, IL 60134-3300

Chair's Introduction—1:00

Invited Papers

1:05

3pAA1. John Kopec and the papers of Wallace C. Sabine. Leo L. Beranek (975 Memorial Dr., Ste. 804, Cambridge, MA 02138, beranekleo@ieee.org)

In April 1975, the Illinois Institute of Technology called the author saying that some notebooks had been found that appeared to be those of Wallace Clement Sabine. It was thought previous to this find that Sabine had destroyed all of his notebooks and papers because none was found when Professor Theodore Lyman was collecting Sabine's papers for publication. Those notebooks were studied, a paper was published [J. Acoust. Soc. Am. **61**, 355–365 (1977)], and they were deposited in the Harvard University Archives. Imagine the surprise when John Kopec called in early 1979, saying that at another location at the Riverbank Acoustical Laboratory, a pile of folders that contained some of Sabine's consulting experiences had been discovered. John said that he had copied Sabine's faded pages into typescript, and he invited the author to join with him in studying them and perhaps publishing a paper jointly with him. Discussed here is our work together and our publication of the results [J. Acoust. Soc. Am. **69**, 367–382 (1981)]. Through this effort readers can judge Sabine's breadth of activity and his depth of understanding in the field of applied acoustics.

1:35

3pAA2. John W. Kopec and *The Sabines at Riverbank*. William Cavanaugh (Cavanaugh Tocci Assoc., Inc., 327F Boston Post Rd., Sudbury, MA 01776)

I first became acquainted with John Kopec at a meeting of the Greater Boston Chapter of ASA in the late 1970s. He told a delightful story of his discovery of papers of Wallace Sabine, long thought to be lost, at the IITRI Riverbank Acoustical Laboratories in Geneva, IL, where John had been an employee since 1961. John was working with Leo Beranek to have these papers donated to the Harvard Rare Books Library for safekeeping [L. L. Beranek, "The Notebooks of Wallace C. Sabine," J. Acoust. Soc. Am. **61**(3), (1977) L. L. Beranek and J. W. Kopec, "Wallace C. Sabine, acoustical consultant," J. Acoust. Soc. Am. **59**(1), (1981)]. I quickly learned that John had been collecting historical information on not only Wallace Sabine, but his cousin Paul Sabine, who was called upon to direct the laboratory in 1919 after Wallace's untimely death, and Paul's son Hale who also became director of the laboratory after Paul. It became obvious to all of us who heard John's tales of Riverbank that he had to put these into print. Hence, at the urging of his abundant friends and colleagues, the Acoustical Society published *The Sabines at Riverbank* in 1997. This book is John's legacy to all of us interested in the science of architectural acoustics and the part played by the Sabines at Riverbank.

Contributed Paper

2:05

3pAA3. John W. Kopec at Riverbank. David Moyer (Riverbank Acoust. Labs. 1512 Batavia Ave., Geneva, IL 60134)

John Kopec's tenure at Riverbank as manager, supervisor, and chief liaison engineer during an active time of architectural and noise control related testing by manufacturers as well as trade associations made for some interesting challenges and client interactions. The author hopes to collect material from a broad range of interested parties to contribute to both aspects. A brief review of some of the 20-plus publications and presentation attributed to Kopec will also be provided.

Session 3pAB

Animal Bioacoustics: Animal Psychophysics

Peter M. Scheifele, Chair
Univ. of Connecticut, Dept. of Animal Science, 850 Bolton Rd., Storrs, CT 06269-1085

Contributed Papers

1:00

3pAB1. Impaired frequency resolution in canaries with hereditary high-frequency hearing loss. Amanda Lauer, Robert Dooling (Dept. of Psych., Univ. of Maryland, College Park, MD 20742), and Marjorie Leek (Natl. Ctr. for Rehabilitative Auditory Res., Portland, OR 97207)

Belgian Waterslager canaries (BWS), bred for a distinct low-pitched song, have an inherited high-frequency hearing loss related to hair cell abnormalities. Many hair cells are missing or damaged, and many of the remaining hair cells have abnormal stereocilia bundles. To investigate how the hair cell abnormalities affect frequency tuning along the BWS basilar papilla, we measured psychophysical tuning curves (PTCs) in BWS and normal-hearing non-BWS canaries using operant conditioning procedures. Non-BWS canaries showed relatively symmetrical tuning curves that became more narrowly tuned with increasing center frequency. Despite near-normal hearing at low frequencies, PTCs in BWS were elevated for all center frequencies tested. The sharpness of tuning was normal in BWS for PTCs with a center frequency of 1000 Hz and broader than normal at higher frequencies. The shapes of the PTCs were abnormal in BWS, particularly at higher frequencies. The abnormal tuning properties in the BWS are consistent with changes in the impedance characteristics of the basilar papilla due to structural damage. [Work supported by NIH DC01372 and DC04664.]

1:15

3pAB2. Spatial release from aerial masking in a harbor seal. Marla M. Holt and Ronald J. Schusterman (UCSC Long Marine Lab., 100 Shaffer Rd., Santa Cruz, CA 95060)

Spatial release from masking (SRM) occurs when a signal and masker are spatially separated, resulting in improvement of signal detection relative to when they are spatially coincident. Harbor seals feed in the water but haul out on land for a variety of activities. There have been no SRM investigations conducted on harbor seals in air. In this study, SRM was measured at 1, 8, and 16 kHz in a harbor seal (*Phoca vitulina*) who had to detect an aerial tone in the presence of an octave band of white noise centered at the tone frequency. While the masker always occurred in front of the subject (0 deg), the tone occurred at 0, 45, or 90 deg in the horizontal plane. Absolute thresholds were also measured at these angles to account for differences in hearing sensitivity based on source azimuth. Current data show that the largest masking level differences (MLDs) of 7 dB occurred at 1 kHz when the signal was projected at 90 deg. MLDs reported here on an amphibious carnivore are larger than those measured under water, are consistent with measurements collected on terrestrial animals, and have important implications for noise effects on free-ranging animals.

1:30

3pAB3. A comparison of behavioral and auditory brainstem response measurements of absolute and masked auditory thresholds in three species of birds. Isabelle C. Noirot (Dept of Psyc., Univ. of Maryland, College Park, MD 20742, and Univ. of Liege, Bat B6c, 4000 Liege, Belgium, inoirot@psyc.umd.edu), Elizabeth F. Brittan-Powell, Robert J. Dooling, and Lindsay Montgomery (Univ. of Maryland, College Park, MD 20742)

Auditory brainstem responses (ABRs) were used to measure auditory thresholds for pure tone in quiet and in noise for three species of birds: budgerigars (*Melopsittacus undulatus*), canaries (*Serinus canarius*), and zebra finches (*Taeniopygia guttata*). ABRs were recorded in response to 5-ms pure tones at frequencies between 500 and 8000 Hz. Results showed that the shape of ABR audiograms generally paralleled the shape of behavioral audiograms but ABR thresholds were generally 20 to 40 dB higher than those obtained behaviorally. ABR thresholds in noise were also measured at frequencies between 1000 and 4000 Hz. Three different levels of broadband white noise were used and critical ratios (CR; signal-to-noise ratio at masked threshold) were calculated. In general, CRs were independent of the level of the masking noise, in good correspondence with behavioral data. While CRs across a wide variety of vertebrates show a 3-dB/octave increase, CRs in these birds were similar or better at 2860 Hz than 2000 Hz, and worse at 4000 Hz. To the extent that these relationships between behavioral and ABR thresholds are generalizable across species, ABR measurements can be used to calculate audiograms and CRs in species that cannot be tested behaviorally. [Work supported by DC00198, DC04664 to R.J.D.]

1:45

3pAB4. Comparative analyses of canine hearing using event-related potentials. Peter M. Scheifele, Michael Darre, Michelle G. Pinto, Frank E. Musiek (Univ. of Connecticut, Unit 4040, 3636 Horsebarn Hill Rd. Ext., Storrs, CT 06269-4040), and John Preece (NYU, New York, NY 10010)

The hearing of three dogs was tested using auditory brainstem response (ABR) and distortion product oto-acoustic emissions (DPOAE) tests. The three subject dogs included one with known normal hearing, one congenitally deaf dog, and one dog with presbycusis. ABRs were run using a 100-s click stimulus at a repetition rate of 19.7 Hz at 90, 60, and 40 dB nHL. Based on the combined analysis of ABR and DPOAE, differences in canine hearing of dogs with presbycusis and congenital (Waardenburg syndrome) are evident. These results continue to be compared with those of other dogs of known otology and audiology in an effort to develop canine norms for ABR and to readily detect these pathologies in companion animals to assist veterinarians in the diagnosis and treatment of hearing loss in dogs.

2:00

3pAB5. Auditory feedback of vocal production in budgerigars using earphones. Michael Osmanski and Robert Dooling (Dept. of Psych., Univ. of Maryland, Biol.-Psych. Bldg., College Park, MD, 20742, mosmanski@psyc.umd.edu)

There is considerable interest in understanding the mechanisms whereby auditory feedback (AF) enables a bird to learn and to produce a stable vocal repertoire. In the free field, budgerigars (*Melopsittacus undulatus*) show evidence of a Lombard effect, an increase in the intensity of their contact calls in the presence of background noise. They also adjust the frequency of their calls to compensate for pitch-altered AF. Here, we extend these findings using earphones to more precisely determine the role of AF in vocal production. Budgerigars were fitted with small head-mounted earphones and were trained to produce contact calls for food reinforcement in the presence of altered AF; alterations included noise and changes in call amplitude, frequency, and spectrotemporal structure (e.g., reversed calls). Results show that budgerigars produce compensatory changes to perturbations in AF. For example, these birds increase the amplitude of their vocal productions in the presence of noise (Lombard effect). They also reduce the level of vocal output when provided with amplified feedback of their own vocalizations. [Work supported by NIH Grant DC-00198 to R.J.D., NIH grant DC-04664 to R.J.D., and NIH Grant DC-006766 to MSO.]

WEDNESDAY AFTERNOON, 7 JUNE 2006 BALLROOM D, 1:30 TO 2:35 P.M.

Session 3pID

Interdisciplinary: Hot Topics in Acoustics

David R. Dowling, Chair
Univ. of Michigan, Dept. of Mechanical Eng., 2019 Automotive Lab. 2133, 1231 Beal Ave., Ann Arbor, MI 48109-2133

Chair's Introduction—1:30

Invited Papers

1:35

3pID1. Vector sensors—over 40 and still hot. Paul C. Hines (Defence R&D Canada-Atlantic, P.O. Box 1012, Dartmouth, NS, B2Y 3Z7, Canada, paul.hines@drdc-rddc.gc.ca)

Acoustic vector sensors and gradient arrays have been in use in underwater acoustics for more than 4 decades, so one may wonder why they are featured in a session entitled ''Hot Topics.'' The reason lies in the recent resurgence in their use. This, in turn, is due primarily to major advances in engineering and signal processing that have been applied to these devices. Historically, theoretical gains have been difficult to achieve with these sensors due to their susceptibility to uncorrelated noise. That is to say, the very process of making a localized measurement of the vector acoustic field lowers the signal-to-noise ratio, relative to a simple pressure measurement. However, with today's advances in design, manufacturing, and digital signal processing, high-quality performance can be achieved in a very small package size. Moreover, the current interest in these devices is not limited to underwater acoustic applications; rather, it extends across a number of technical areas within the acoustics community. This presentation will begin with a brief introduction to the theoretical foundation and history of these devices. Then, some current applications in both underwater and airborne acoustics will be highlighted.

1:55

3pID2. Extracting coherent information from cross correlations of random wave fields. Karim G. Sabra (Marine Phys. Lab, Scripps Inst. of Oceanogr., U.C. San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0238, ksabra@mpl.ucsd.edu)

The random nature of noise and scattered fields tends to suggest limited utility. Indeed, seismic or acoustic fields from random sources or scatterers are often considered to be incoherent. Nevertheless, it has been shown that it is possible to extract information from random noise fields of a type that would typically be associated with a coherent process. An estimate of the Green's function (or impulse response) between two points can be obtained from the cross correlation of ambient noise recorded at these two points. Recent theoretical and experimental studies in ultrasonics, underwater acoustics, and seismology have investigated this technique in various environments and frequency ranges. These results provide a means for passive imaging using only the ambient noise field, without the use of active sources. The coherent wave-fronts emerge from a correlation process that accumulates contributions over time from noise sources whose propagation paths pass through both receivers. Results will be presented from experiments using ambient noise cross correlations for the following applications: (1) localization and synchronization of unconnected acoustic receivers; (2) passive tomography of the seismoacoustic environment; and (3) studies of marine and airborne structures embedded in turbulent flow.

2:15

3pID3. Demonstrations and experiments: The life-blood of acoustics education. Uwe J. Hansen (Dept. of Phys., Indiana State Univ., Terre Haute, IN 47809)

You hear: You forget; You see: You remember; You do: You understand. Acoustics represents a science discipline with very broad appeal. It is essential, both for the well being of the field, as well as the influence of science at large, to generate enthusiasm for science in general and acoustics in particular in our young people. Thus educational outreach is vital. The effectiveness of that

outreach is closely related to the direct involvement of the students. A number of examples of demonstrations and hands-on experiments will be shown. They include acoustic levitation, mode studies, Dan Russell's web site, and video animations of modal analysis studies in several musical instruments.

WEDNESDAY AFTERNOON, 7 JUNE 2006 ROOM 557, 1:00 TO 3:00 P.M.

Session 3pNS

Noise, Architectural Acoustics, and Speech Communication: Fifty Years of Speech Privacy II

Gregory C. Tocci, Chair
Cavanaugh Tocci Associates, 327F Boston Post Rd., Sudbury, MA 01776

Contributed Papers

1:00

3pNS1. Measurements of sound leaks or "hot spots" and their effect on the architectural speech security of closed rooms. Bradford N. Gover and John S. Bradley (Inst. for Res. in Construction, Natl. Res. Council, Ottawa, ON, K1A 0R6, Canada)

A new measurement procedure has been developed that accurately indicates the degree of speech security (speech privacy) of a closed office or meeting room. The procedure first determines the attenuation from an average sound-pressure field within a source room to single receiving points outside the room. These average-to-spot attenuations are used to predict transmitted speech levels outside of meeting rooms, which along with the background noise levels are used to derive a reliable indicator of the audibility or intelligibility of speech at the receiving points. A key aspect of the approach is that a localized weak spot in an otherwise highly insulating partition can be identified and assessed. Investigations were carried out as to the severity and detectability of different types of sound leaks intentionally introduced into an otherwise ''good'' wall separating two reverberation chambers. Steady-state and impulse response measurements of received pressure were made at a distance of 0.25 m from the wall. In addition, a highly directional beamforming microphone array was used in an effort to quickly locate the position of potential leaks. An overview of the measurement procedures will be presented, including discussion of the severity of several types of ''hot spots,'' such as holes, penetrations, and electrical boxes.

1:15

3pNS2. Speech privacy in design: A cost-privacy analysis. Tim Lavallee (LPES, Inc., 14053 Lawnes Creek Rd., Smithfield, VA 23430) and Kenneth Good (Acoust. Privacy Enterprises, LLC, Mount Joy, PA 17552)

With the development of new standards in the health care industry, design professionals will soon have additional construction criteria to ensure speech privacy in new facilities. There are currently sound transmission class (STC) ratings required for partitions in a healthcare setting outlined in the Guidelines for Design and Construction of Hospital and Healthcare Facilities. These currently implemented criteria have traditionally focused on patient personal privacy and not speech privacy. This effort will be a general analysis to determine the shortcomings of these standards with respect to speech privacy and a cost-privacy analysis for commonly used wall constructions. Wall constructions included in the analysis will be suitable for a healthcare setting. Commonly referenced published, modeled, and measured transmission loss across partitions will be shown. The privacy indices will be evaluated for each configuration. The talk will highlight that, in most cases, no extraordinary measures will be needed to ensure reasonable safeguards have been implemented.

1:30

3pNS3. The performance of a newly developed mobile hands-free context-aware communications system in a hospital environment. Joshua D. Atkins (Dept. of Elec. Eng., Johns Hopkins Univ., 3400 North Charles St., Baltimore, MD 21218, joshatkins@jhu.edu), Michael J. Sammon, and Lynne S. Brotman (Avaya Labs Res., Basking Ridge, NJ 07920)

This paper characterizes the effect of hospital noise conditions on a mobile communications system developed for a research trial in a Johns Hopkins Hospital nursing unit. The system provides an alternative to telephones and overhead paging in hopes of reducing the amount of noise on the hospital unit. Benefits of this system over current similar solutions include a completely hands free approach using a computerized voice agent and a design that addresses privacy concerns to comply with HIPAA requirements. The nurses were happy with the system and bypassed the phones and overhead paging in preference of their headsets. However, during the system trial, the effect of noise on the automatic speech recognition (ASR) engine from sources like alarms, medical equipment, and wind noise from mobile users was an issue. Utterance tests were run using a variety of different Bluetooth headsets to measure the impact of various noise sources, S/N, and noise reduction algorithms on ASR performance. Common noise reduction methods were not helpful for the type of noise encountered, suggesting that solutions such as noise robust speech systems and noise reduction with adaptive microphone arrays currently being researched will be of great importance to future systems of this kind.

1:45

3pNS4. The role of sound masking treatment in the value engineering of office designs. Thomas Horrall and Carl Rosenberg (Acentech, 33 Moulton St., Cambridge, MA 02138)

For over 50 years sound masking systems have often been included in the design and planning of office spaces because the owner understood that they are required to provide adequate speech privacy, or because he believed they can create a more pleasing and private acoustic environment. However, installations have rarely been considered on the basis of project value engineering or return on investment. In this paper, we argue that sound masking systems should in fact be considered not only as a stand-alone cost component of the office design, but evaluated along with ceiling tile, private office wall construction, and workstation design for their mutual impact on the overall project budget, instead of solely on their contribution to acoustical performance. With over 70 years of combined experience dealing with speech privacy issues and masking systems of all kinds, we provide typical project data to demonstrate that contemporary systems often can result in substantial overall project cost savings.

2:00–3:00
Panel Discussion

Session 3pSP

Signal Processing in Acoustics and Underwater Acoustics: Optimum and Robust Signal Processing in Uncertain and Random Environments II

Ning Xiang, Chair
Rensselaer Polytechnic Inst., Architecture, 110 8th St., Troy, NY 12180

Chair's Introduction—1:00

Contributed Papers

1:05

3pSP1. Comparison of extended Kalman filtering with split-aperture processing for angle of arrival estimation in multibeam echo-sounding. Daniel S. Brogan and Christian P. de Moustier (Ctr. for Coastal and Ocean Mapping, Univ. of New Hampshire, 24 Colovos Rd., Durham, NH 03824-3525, daniel.brogan@unh.edu)

Echo angle of arrival estimates may be obtained from the FFT beamformed output of a multibeam sonar system at each successive time sample. Whether tracking disparate targets in the water column or tracking a mostly continuous surface such as the seafloor, the sequence of angles of arrival versus time is inherently nonlinear and nonstationary, and lends itself to processing with an extended Kalman filter. This filter provides both an approximation of the optimum (in a least squares sense) angles of arrival estimation and a measure of the associated errors. This technique is applied to quadrature-sampled field data collected with the 26-element hydrophone array of a 160-kHz multibeam sonar system. Results are compared to arrival angle estimates obtained with a conventional split-aperture processor. [Work supported by NOAA Grant 111833.]

1:20

3pSP2. Universal properties of the channel Q function: Applications to underwater acoustic communications. T. C. Yang (Naval Res. Lab., Code 7120, Washington, DC 20375)

Channel Q function is, for a time-invariant channel, the sum of autocorrelation functions of the channel impulse response functions over multiple receivers [T. C. Yang, IEEE J. Oceanic Eng. **28**, 229–245 (2003)]. It can be used to characterize the performance (output SNR) of acoustic communications using either passive-phase conjugation or decision feedback equalizer [T. C. Yang, IEEE J. Oceanic Eng. **29**, 472–487 (2004)]. Channel Q function also determines the channel capacity for a single source and multiple receivers (SIMO). We analyze in this paper the properties of the channel Q function and show that it possesses universal and stable properties for underwater acoustic waveguides, irrespective of the channel sound-speed profiles, source and receiver range/depths. This property is examined with real data collected in various oceans with significant temporal variations in the propagation conditions and/or source-receiver range. Applications to underwater acoustic communications are discussed. [Work supported by the Office of Naval Research.]

1:35

3pSP3. Error performance of equalizers based on sparse and dynamic channel estimation. Weichang Li and James C. Preisig (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

For broadband communications in the very shallow water environment, it has been shown that both model-based dynamic channel estimation and sparse channel estimation are capable of reducing the signal prediction error by several dB [W. Li and J. C. Preisig, J. Acoust. Soc. Am. **118**, 2014 (2005)]. Dynamic channel estimation is based on explicit modeling of the rapid channel variations; sparse channel estimation exploits the channel structure formed by the resolved multipath arrivals. In this paper, the error performance of the equalizer based on these channel estimators are analyzed, demonstrated through experimental results and compared with the performance of equalizers that do not explicitly model and exploit the channel dynamics and structure. [Work supported by ONR Ocean Acoustics.]

1:50

3pSP4. Adaptive predictive-error optimal gridding. Lisa A. Pflug and Donald R. DelBalzo (Planning Systems, Inc., 40201 Hwy. 190 E., Slidell, LA 70461, lpflug@plansys.com)

Transmission loss computations in littoral areas often require a dense spatial and azimuthal grid to achieve acceptable accuracy and detail. The computational cost of accurate predictions in complex areas can be unacceptable for some naval applications. EAGLE (Efficient Adaptive Gridder for Littoral Environments) was developed to reduce the computational requirements by adapting the fundamental calculation grid to the natural environmental structure. EAGLE produces a nonuniform grid structure within which the underlying data have approximately constant spatial uncertainty, or iso-deviance. The EAGLE point-selection approach is based on a predictive-error metric and least-squares polynomial curve fitting with arbitrary order. This research indicates the degree to which EAGLE grids are related to the underlying coherence properties of the natural environment. The results show significant efficiency enhancements compared to uniform grids of the same accuracy, when the environmental coherence properties are used to control the polynomial order. [Work sponsored by NAVAIR.]

2:05

3pSP5. Quantifying ocean acoustic environmental sensitivity. Stan E. Dosso (School of Earth and Ocean Sci., Univ. of Victoria, Victoria BC, Canada), Peter Giles, Gary H. Brooke (General Dynam. Canada, Dartmouth NS, Canada), Diana F. McCammon (McCammon Acoust. Consulting, Waterville NS, Canada), Sean Pecknold, and Paul C. Hines (Defence Res. and Development Canada—Altantic, Dartmouth NS Canada)

This paper examines the sensitivity of acoustic propagation data required for sonar performance predictions to physical parameters of the ocean environment. Sensitivity is quantified here by an appropriate measure of the relative uncertainty of the acoustic data due to realistic uncertainties in environmental parameters. Within a linear approximation, Gaussian-distributed environmental parameter uncertainties lead directly to Gaussian data uncertainties which scale linearly with the parameter uncertainties. However, the actual data uncertainties, which can be esti-

mated numerically via Monte Carlo methods, do not necessarily possess these desirable linear properties. The applicability of the linearized approximation and the relative magnitude of acoustic sensitivities are examined for realistic uncertainties in seabed geoacoustic parameters and oceanographic features of the water-column sound-speed profile. Sensitivities are considered as a function of source and receiver depth, range, and source frequency for various shallow-water environments. [Funding provided by DRDC-Atlantic Rapid Environmental Assessment Project.]

2:20

3pSP6. Approximating acoustic field uncertainty in underwater sound channels. Kevin R. James and David R. Dowling (Dept. of Mech. Eng., Univ. of Michigan, 2019 Lay Auto Lab., 1231 Beal Ave., Ann Arbor, MI 48109, drd@umich.edu)

Precise acoustic field prediction relies on accurate knowledge of many environmental parameters. Uncertainty in any of the inputs to a propagation routine leads to uncertainty in the predicted field. The most accurate but time-consuming method of assessing the resulting field uncertainty is Monte Carlo simulation, which requires calculating the field with many possible combinations of inputs. This presentation covers a method to approximately determine the uncertainty in the predicted field, using only one additional field calculation for each uncertain parameter, when the input uncertainties are relatively small. The method relies on identifying how variations in uncertain input parameters can be mapped into spatial displacements in the original field calculation. Simultaneous uncertainty in multiple input parameters is handled using a net effective spatial displacement. In addition to expected values and standard deviations, the method allows for the generation of complete probability density functions for the uncertain predicted field. Calculated field amplitude results for both ideal and more realistic range-independent sound channels having small depth and sound speed uncertainty are shown and compared to Monte Carlo simulations at several source-receiver ranges. [Work sponsored by the Office of Naval Research, Code 321OA.]

2:35

3pSP7. Adaptive sampling based on the waveguide invariant for determining environmental provinces. James H. Leclere and George B. Smith (Code 7183, Bldg. 1005, Naval Res. Lab., Stennis Space Ctr., MS 39529, gsmith@nrlssc.navy.mil)

The waveguide invariant (related to the striations seen in range-frequency plots of sonar data) can be used as a tool to determine provinces in adaptive sampling schemes used to plan the placement of environmental measurements. The quantity from which the waveguide invariant is derived is used to demark spatial regions over which the environment can be considered to be range and bearing independent from an acoustic point of view. This allows significant speedup in acoustic propagation calculations. It also provides guidance for the placement of environmental measurements—particularly water column sound speed profile measurements. Results from the application of this technique to simulated data and to data from the Sea of Japan are shown and are compared to results of another method based on the adiabatic condition. [This work was supported by ONR with technical guidance provided by the Naval Research Laboratory, Stennis Space Center, MS.]

Plenary Session, Annual Meeting, and Awards Ceremony

William A. Yost, President
Acoustical Society of America

Welcome Remarks

Annual Meeting of the Membership

Presentation of Certificates to New Fellows

Anders Askenfelt
Alexander U. Case
Torsten Dau
Carol Espy-Wilson
J. Gregory McDaniel
Sheryl Gracewski
Lee A. Miller
Bertel Møhl
Shrikanth Narayanan
Simon D. Richards
Charles M. Salter
Ralph A. Stephen
Dajun Tang
Gail ter Haar

Announcement of Awards

Lawrence A. Crum, Student Council Mentoring Award

John K. Horne, Medwin Prize in Acoustical Oceanography

Rajka Smiljanic, American Speech-Language Hearing Foundation Research Grant in Speech Science

Presentation of Acoustical Society Awards

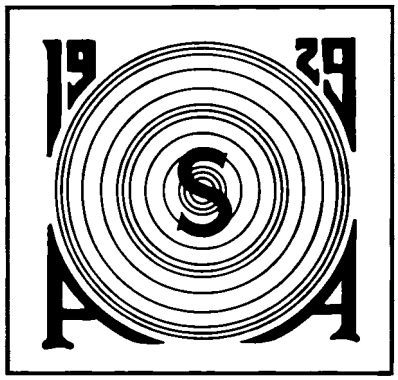
R. Bruce Lindsay Award to Purnima Ratilal

Helmholtz-Rayleigh Interdisciplinary Silver Medal to Mathias Fink

Gold Medal to James E. West

ACOUSTICAL SOCIETY OF AMERICA

R. BRUCE LINDSAY AWARD



Purnima Ratilal

2006

The R. Bruce Lindsay Award (formerly the Biennial Award) is presented in the Spring to a member of the Society who is under 35 years of age on 1 January of the year of the Award and who, during a period of two or more years immediately preceding the award, has been active in the affairs of the Society and has contributed substantially, through published papers, to the advancement of theoretical or applied acoustics, or both. The award was presented biennially until 1986. It is now an annual award.

PREVIOUS RECIPIENTS

Richard H. Bolt	1942	William E. Cooper	1986
Leo L. Beranek	1944	Ilene J. Busch-Vishniac	1987
Vincent Salmon	1946	Gilles A. Daigle	1988
Isadore Rudnick	1948	Mark F. Hamilton	1989
J. C. R. Licklider	1950	Thomas J. Hofler	1990
Osman K. Mawardi	1952	Yves H. Berthelot	1991
Uno Ingard	1954	Joseph M. Cuschieri	1991
Ernest Yeager	1956	Anthony A. Atchley	1992
Ira J. Hirsh	1956	Michael D. Collins	1993
Bruce P. Bogert	1958	Robert P. Carlyon	1994
Ira Dyer	1960	Beverly A. Wright	1995
Alan Powell	1962	Victor W. Sparrow	1996
Tony F. W. Embleton	1964	D. Keith Wilson	1997
David M. Green	1966	Robert L. Clark	1998
Emmanuel P. Papadakis	1968	Paul E. Barbone	1999
Logan E. Hargrove	1970	Robin O. Cleveland	2000
Robert D. Finch	1972	Andrew J. Oxenham	2001
Lawrence R. Rabiner	1974	James J. Finneran	2002
Robert E. Apfel	1976	Thomas J. Royston	2002
Henry E. Bass	1978	Dani Byrd	2003
Peter H. Rogers	1980	Michael R. Bailey	2004
Ralph N. Baer	1982	Lily M. Wang	2005
Peter N. Mikhalevsky	1984		

CITATION FOR PURNIMA RATILAL

. . . for contributions to the theory of wave propagation and scattering through a waveguide, and to the acoustic remote sensing of marine life

PROVIDENCE, RHODE ISLAND • 7 JUNE 2006

''Dancing helps against seasickness,'' said Purnima Ratilal, this year's recipient of the R. Bruce Lindsay Award, as she demonstrated on the helicopter deck, to the beat of the waves in a raging mid-spring squall at the edge of the continental shelf, a day's sail from the nearest landfall. Not all of us were in a condition to follow. At sea, exhausted after a long day's work, we'd sometimes discover her dancing on the fore deck in the twilight, hidden behind the bridge, full of energy, . . . the professor who dances with the ocean.

Purnima was born in Singapore, the youngest of five children, to her parents Ratilal and Pushpaben. She graduated with honors from the Physics Department of the National University of Singapore, where she is fondly remembered by her bachelor's thesis adviser, the astrophysicist Chia Tzu Tit, as one of his best undergraduate students. After graduation, she became a research physicist at Singapore's DSO National Laboratories, where she quickly established herself as a rising young star. She performed at a level well beyond her years by designing and directing a series of offshore oceanographic experiments in the seas near Singapore. From what can be gathered now, this bit of ocean apparently served as a personal laboratory for Purnima to develop her remarkable intuition about the physics of remote sensing. She did this by investigating such phenomena as transmission through a fluctuating medium, as well as the causes and effects of reverberation and ambient noise in a waveguide. It also began her career in characterizing the marine environment with sound, here using matched field inversion and headwave analysis to determine local oceanographic and seabed properties. After presentations at a number of international conferences, including the Office of Naval Research (ONR) Ocean Acoustic Inverse-Methods Workshop where her work found its way to the top, she was heavily recruited by many leading graduate schools, including the Massachusetts Institute of Technology (MIT), which was fortunate enough to sign her on for study.

As a graduate student, Purnima flew through her initial doctoral qualifying exams, easily passing them after only four months at MIT. Almost all students wait 16 months. Unlike most students, she always seemed to pick professors with the toughest reputations from around the institute to be on her exam and thesis committees. Her enthusiasm for learning seemed to short-circuit the usual fear. In the second part of her qualifying exams, years later, she was grilled mercilessly by Jin Au Kong, author of the well-known volumes ''Scattering of Electromagnetic Waves,'' and leading theoretical hydrodynamicist Dick Yue, the Associate Dean of the School of Engineering, who often joked that ''acoustics is just a subset of hydrodynamics.'' Kong left the exam very impressed, but exhausted after two hours, the blackboard densely covered with Purnima's latest derivations in acoustic waveguide-scattering theory. Purnima, on the other hand, was still energetic and hoping for more fun. Yue supplied this ''fun,'' taking another couple of hours from a busy schedule. Recently he commented that her research and thesis work is of the caliber we get only every few years at MIT. As a graduate student Purnima also engaged in many philanthropic undertakings with extreme generosity, serving at homeless shelters, regularly contributing with nothing more than her meager graduate-student stipend to various charities. She pulled many other students, undergraduate and graduate alike, out of the fire by giving them hours and hours of extra tutoring and guidance. She did the same for her research group time and time again by jumping in and tackling the most difficult outstanding problems with amazing creativity, efficiency, sophistication and common sense.

Purnima's most important theoretical contributions began early in her career as a graduate student when she derived the Extinction Theorem for object scattering in a waveguide, and so generalized the classical free-space Extinction Theorem, also known as the optical theorem or Forward Scatter Theorem, one of the most fundamental results in scattering theory. (The free-space version was derived in various forms by such well-known investigators as Rayleigh, Heisenberg and Van du Hulst.) She showed extinction in a waveguide, or removal of power from the forward propagated field by the object, to be far more complicated than in free-space, and to involve intuitive mode-coupling effects arising from subtle interference structures inherent in a waveguide.

Later as an ONR Post-Doctoral Fellow at MIT, and as Assistant Professor of Electrical Engineering and Computer Science at Northeastern University, she returned to what became one of her

favorite topics, forward scattering. In what ASA Associate Editor Bill Siegmann has called a ''theoretical tour de force that for the first time provides mean and covariance expressions for three-dimensional scattering effects from virtually any mechanism of interest in the ocean,'' she derived compact analytic expressions for these field moments after multiple forward scattering in a waveguide, and showed when three-dimensional effects become important, without the need for laborious Monte-Carlo simulations. Michael Collins of the Naval Research Laboratory in Washington, DC, and a previous recipient of the Lindsay Award, whose propagation models are in standard use by the US Fleet, notes that ''Achieving this goal has long been the Holy Grail in this area, and Purnima has done it.'' For the essential inspiration, she went back deeply, citing Rayleigh's 1899 derivation explaining why the sky is blue and the sunset red. As many of us know, it was R. Bruce Lindsay who wrote the introduction to Lord Rayleigh's classic book, which is well referenced in her papers.

In her recent experimental work, appearing both in our Society's journal as well as the journal *Science*, Purnima co-invented a method for instantaneously detecting and imaging fish populations over continental-shelf-scale areas (thousands of square kilometers) and then continuously monitoring these populations with Ocean Acoustic Waveguide Remote Sensing (OAWRS). There she used OAWRS to make a number of fundamental new discoveries about animal group behavior by revealing the instantaneous horizontal structural characteristics, volatile short term behavior and propagation of information in very large fish shoals, containing tens of millions of fish and stretching for tens of kilometers. Colleagues in a wide variety of disciplines expect the impact of her work to the Census of Marine Life, the study of marine ecology, and the management of marine fisheries to be pronounced.

With the same series of experiments, Purnima also played an enormous role in helping to solve one of the Navy's primary problems in active sonar operations by showing that the presence of strong inexplicable ''clutter'' returns in continental shelf environments are often caused by dynamically moving fish groups rather than geology. Jeff Simmen, the ONR Program Manager in Underwater Acoustics at the time, noted this was ''a very important and unexpected finding,'' since the original goal of the experimental program was to seek geologic correlates.

Purnima has made scores of presentations at ASA meetings and has been awarded Best Student Paper Awards on three occasions. She presented one of the Young Investigator Keynote Addresses at the ASA 75th Anniversary Celebration in New York. Purnima is the author or co-author of several papers published in the *Journal of the Acoustical Society of America* and is a member of the Technical Committee on Underwater Acoustics. With Purnima, the best is always yet to come. We will no doubt see many fascinating new discoveries from her and the vibrant research group she has put together at Northeastern University.

NICHOLAS C. MAKRIS

ACOUSTICAL SOCIETY OF AMERICA

Silver Medal in

Biomedical Ultrasound/Bioresponse to Vibration
and
Acoustical Oceanography

Mathias Fink

2006

The Silver Medal is presented to individuals, without age limitation, for contributions to the advancement of science, engineering, or human welfare through the application of acoustic principles, or through research accomplishment in acoustics.

ENCOMIUM FOR MATHIAS FINK

. . . for contributions to the understanding of time reversal acoustics

PROVIDENCE, RHODE ISLAND • 7 JUNE 2006

The common threads of Mathias Fink's activities have been his exceptional diversity and energy. He has made seminal contributions in physical and medical acoustics, wave propagation in simple or complex solids or fluids that often employed analogies between acoustics, optics, quantum mechanics, and has also applied basic physics to accomplish self-adaptive signal processing. All this while being a devoted teacher of a generation of acousticians and also playing an important role in the development of a new, acoustic-based industry.

Mathias A. Fink was born in Grenoble (France) in October 1945. He received the B.S degree in Mathematics from the University of Paris (1967) and M.S. (1968) and Ph.D. (1970) degrees in Solid State Physics, at the Groupe de Physique du Solide de l'Ecole Normale Superieure in Paris. His Ph.D. research was in the area of electron-phonon interaction in semiconductors. Because he was a strong diver and interested in Underwater Archeology, Mathias decided to pursue a research career in Underwater Acoustics. In1972, he joined Pierre Alais at the Pierre et Marie Curie University (Paris). However, with no significant funding in his chosen field, he began working in the new field of Ultrasonic Medical Imaging. In 1978 he received the Doctorat es-Sciences from Paris University for his research in the area of ultrasonic focusing that included building the first real-time ultrasonic imaging system with electronic focusing capability. He developed various coded aperture techniques to achieve high-resolution imaging based on broadband diffraction theory. His water skills, though, were put to use years later when he rescued a colleague from a sinking boat off the coast of Corsica during one of the Cargese Summer Physics Schools.

In 1983 he joined Strasbourg University as a full Professor in Physics, where he conducted research in ultrasonic medical imaging and non-destructive testing. After a sabbatical at the University of California, Irvine, he joined Paris University (Denis Diderot) in 1984 and in 1990, he founded the now world famous Laboratoire Ondes et Acoustique at the Ecole Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI). He was elected a Senior Member of the Institut Universitaire de France in 1994; in1996 he was awarded the Grand Prix, by the New Economist and the Centre National de la Recherche Scientifque (CNRS) for his collaboration with SNECMA (French aircraft engine society) in nondestructive testing. He was elected a Member of the Académie des Technologies (National Academy of Engineering of France) in 2002 and at the French Académie of Science in 2003.

Mathias is most well known for pioneering the development of the Ultrasonic Time Reversal Mirror (TRM) from his basic physics experiments to the development of ultrasonic medical and nondestructive testing techniques. He first demonstrated the high degree of focus one obtains with time reversal (TR) and, more interestingly, showed (with Derode and Roux) that multiscatter media reveal the TR focal properties thereby providing significant opportunities for acoustic and wireless communication through complex environments. His research included an explanation of this phenomenon in terms of basic statistical mechanical quantities such as the mean free scattering path. On a basic research level I mention only three (of many) extremely elegant experiments with associated theory. He studied with Carsten Draeger single channel TR in a chaotic cavity finding an analogy with Anderson localization. With student Roux, he studied the basic nature of time reversal in reciprocity breaking media such as a fluid with vortices showing a beautiful analogy with the electromagnetic Bohm-Aharonov effect. In another elegant set of TR based vibration experiments, he and his student, Julien DeRosny showed how one can go past the diffraction limit by adding the missing evanescent spectrum using an ''anti-sink'' source function. In signal processing, his work with Claire Prada led to the ''DORT'' method—a decomposition of the Time Reversal Operator that has a fundamental relation with nonlinear adaptive processing methods. Further this procedure reveals the properties of elastic objects and other complex targets from scattering measurements. Getting more applied, the *Economist* magazine recently reported on a startup company based on his work in developing a TR-based keyboard without moving parts. Always dancing between applications and basic science, two other interesting research areas in which he has worked with students are on the application of a Time Reversal Mirror to sonoluminescence for the purpose of creating a more highly compressed bubble and to room acoustics using Lamb waves to focus sound from one room to another.

Much of Mathias' work has been in medical ultrasonics. ''Supersonic Imagine,'' which is also the name of a startup company, is an incredibly clever method to image (the rigidity of) human

tissue. Since the shear speed of human tissue is of the order of only a few m/s, he developed (with Mickael Tanter) a way to acoustically create a focus spot in tissue that could move faster than the shear speed. The resulting Mach cone is imaged ultrasonically and the large shear speed variations associated with the rigidity of a tumor, for example, are identified from variations in the Mach cone apex-angle. Further, his group has made important progress in TR based methods to place an acoustic focus past the highly absorbing skull into the brain. Experiments are now being conducted together with medical facilities in Paris. Another example of his work in medical ultrasonics is the study of nonlinear TR and the preshock conditions that still permit TR focusing; this work is leading to a compact lithotripsy system based on the ''acoustic bazooka'' he developed in his lab. Mathias mentioned to me that this latter device might also be used an acoustic paint remover or ''sand blaster'' type device.

In France, Mathias is very active in the politics and the popularization of science. I remember seeing him on the cover of *Le Monde* during one of my first visits to his lab: it was a picture of an ominous looking bald man in a black background with the article entitled ''Master of Time.'' He is constantly embroiled in the politics of not only promoting scientific research in France, but also in promoting an entrepreneurial friendly environment to enhance technology-based startups. Very recently, the French Minister of Industry organized a press conference that featured Mathias explaining his successful efforts in convincing some French scientists to return from the U.S. to start up the new French company Supersonic Imagine.

I have attended many of his lectures and even co-lectured with him during some tutorials; he is not an easy act to follow. He is in great demand as a lecturer and typically circumnavigates the globe more than once a year. But the true measure of his effectiveness is his students—many of whom I have the pleasure of knowing and some of whom I have been lucky enough to have as post docs. Together they make up a rather spectacular collection of scientists—too many to mention here.

An avid experimentalist, in his early days Mathias used to bring equipment home for various fluid experiments—once placing sensors (hooked up to an oscilloscope) in milk and excitedly making, what he thought, a world shattering discovery on the change of the milk after three days as seen through his electro-hydro-acoustic measurements. In this instance, however, his wife, Minelle, who has not only tolerated his jet-set tendencies, has also kept Mathias grounded in reality while he operated at the elevated heights of the most erudite science, pointed out that the milk had simply curdled. His two daughters, Emmanuelle and Julia, also provide a very lively family life; and one of them also provided a concrete example that time *also* goes forward: his new grandson Antoine.

But Fink is not only well known within the acoustics community; his more popular articles on Time Reversal Mirrors in *Scientific American* and *Physics Today* have also served as a beacon of the excitement of contemporary acoustics to those outside the field. With over twenty-five patents and an extensive publication record, dozens of Ph.D. students, Mathias has been a major, interdisciplinary contributor to both basic and applied acoustics. He is truly deserving of the Helmhlotz-Rayleigh Interdisciplinary Silver Medal.

W.A. KUPERMAN

GOLD MEDAL
of the
Acoustical Society of America

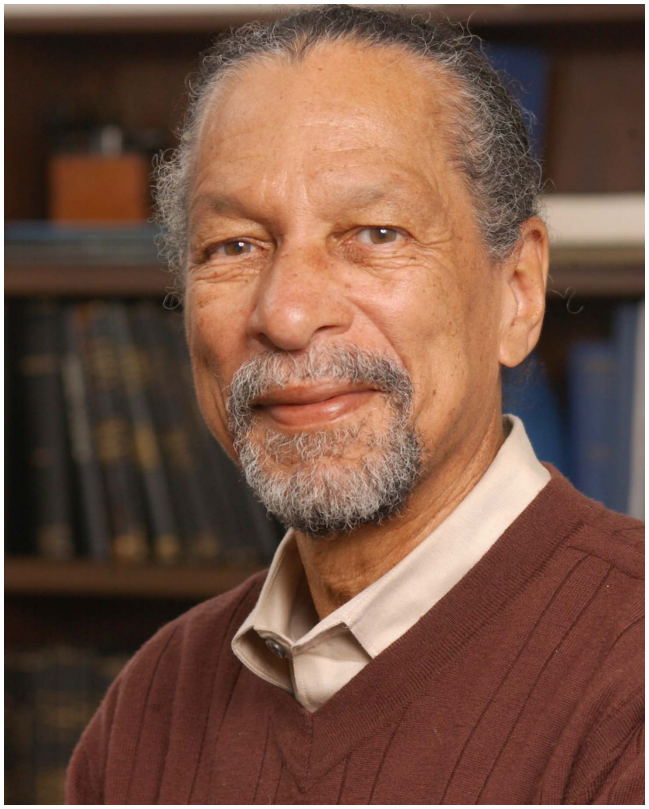
James E. West

2006

The Gold Medal is presented in the spring to a member of the Society, without age limitation, for contributions to acoustics. The first Gold Medal was presented in 1954 on the occasion of the Society's Twenty-Fifth Anniversary Celebration and biennially until 1981. It is now an annual award.

PREVIOUS RECIPIENTS

Wallace Waterfall	1954	Arthur H. Benade	1988
Floyd A. Firestone	1955	Richard K. Cook	1988
Harvey Fletcher	1957	Lothar W. Cremer	1989
Edward C. Wente	1959	Eugen J. Skudrzyk	1990
Georg von Békésy	1961	Manfred R. Schroeder	1991
R. Bruce Lindsay	1963	Ira J. Hirsh	1992
Hallowell Davis	1965	David T. Blackstock	1993
Vern O. Knudsen	1967	David M. Green	1994
Frederick V. Hunt	1969	Kenneth N. Stevens	1995
Warren P. Mason	1971	Ira Dyer	1996
Philip M. Morse	1973	K. Uno Ingard	1997
Leo L. Beranek	1975	Floyd Dunn	1998
Raymond W. B. Stephens	1977	Henning E. von Gierke	1999
Richard H. Bolt	1979	Murray Strasberg	2000
Harry F. Olson	1981	Herman Medwin	2001
Isadore Rudnick	1982	Robert E. Apfel	2002
Martin Greenspan	1983	Tony F. W. Embleton	2002
Robert T. Beyer	1984	Richard H. Lyon	2003
Laurence Batchelder	1985	Chester M. McKinney	2004
James L. Flanagan	1986	Allan D. Pierce	2005
Cyril M. Harris	1987		



CITATION FOR JAMES E. WEST

. . . for development of polymer electret transducers, and for leadership in acoustics and the Society

PROVIDENCE, RHODE ISLAND • 7 JUNE 2006

James Edward West is a consummate acoustician, inventor and an extraordinary human being. In his lifetime of professional activities in acoustics, he has created a large number of distinctly unique microphones and loudspeakers and made important contributions in architectural acoustics and noise control. He has also mentored a number of current ASA members and dedicated enormous time and effort to encouraging students to consider careers in science.

Born on February 10, 1931 in rural Virginia, Jim West is the son of Matilda and Samuel West. Matilda West was a school teacher and during World War II was employed at Langley Air Force Base working with wind tunnel data until she lost her job because she was a leader in the National Association for the Advancement of Colored People, a group that Senator Joseph McCarthy had labeled as communist. Samuel West worked, at various points, as a funeral home owner, B & O Railroad porter, and insurance salesman.

Jim's interest in electrical phenomena began at an early age when he found an old radio and plugged it in to see if it still worked. His lab skills not yet perfectly honed, the resulting electrical shock immobilized him until his brother knocked him over. From that moment he was hooked. In the 1940s Jim's cousin, Edward B. Allen, a school teacher, worked part time wiring homes and Jim, aged 12, went to work for him crawling under houses and carrying the wire from one place to another. Thus began his lifelong habit of tinkering and seeking to understand how things work —in his own words, ''A screwdriver and pliers were very dangerous. Anything that had screws in it, I could open up.''

After graduating from high school, Jim enrolled at Hampton University as a premedical student —honoring the desire of his parents that he pursue one of the careers open to black people of that time (lawyer, doctor, teacher, and preacher). But before completing his degree, Jim was drafted to serve in the Korean War, where he was awarded a Purple Heart. Upon returning to college Jim switched to Temple University and solid-state physics.

Jim's contributions to acoustics largely define the advances in transduction in the last 50 years. He, along with Gerhard Sessler, invented the modern electret microphone found in virtually all telephones and voice recorders, and now produced worldwide at a rate well in excess of 2 billion units annually. He also was involved in the production of adaptive close-talking microphones, first- and second-order differential microphones, blood pressure monitoring transducers, wideband telephone receivers, and microphone arrays of various sorts. For his work on electroacoustic devices, he has been inducted into the National Inventors Hall of Fame and the National Academy of Engineering and honored with too many awards to name. Jim's name is on 47 US and over 200 foreign patents. He has authored over 100 scientific articles.

In addition to his work on transducers, Jim has made important contributions to architectural acoustics. He defined the ''seat effect'' in a 1964 paper and was the first to publish the importance of early reflection arrivals in the subjective quality assessment of concert halls (in an ASA abstract). Jim was also instrumental in the measurement of the performance of a number of halls, including Philharmonic Hall.

More recently, Jim has focused his attention on new piezoelectric materials, new multi-channel tele-collaboration techniques, and noise control in hospitals. The work on piezoelectric material uses previously unexploited mechanisms for generating dipole moments in the design of polymer composites in which the mechanical performance is determined by one polymer and the electroacoustic coupling by a second. This ability to separately control mechanical and electrical performance will make it possible to custom design piezoelectric materials with specified performance measures. Multi-channel tele-collaboration is a joint project with Fred Juang at Georgia Institute of Technology that aims at making more effective use of stereo hearing in tele-collaboration venues. In the hospital noise area, Jim has worked to characterize the noise at Johns Hopkins Hospital and to demonstrate noise interventions that will work long term. This work has generated an enormous positive response from medical caregivers and the lay public.

Jim has also been very active in the Acoustical Society of America. He regularly delivers talks at our biennial meetings and participates in the engineering acoustics, noise, and architectural acoustics technical committee meetings. He has been a member of the Acoustical Society of America (ASA) since 1962 and has served as chair and member of many committees and in several elected positions including Member of the Executive Council (1989-92), President-Elect

(1997-98) and President (1998-99). Jim was instrumental in the establishment of the Society's minority fellowship and the educational programs for local high-school students that are conducted at Society meetings. During his tenure as President of the Society, the ASA held its first meeting outside of North America jointly with the European Acoustics Association. Jim worked to further broaden our reach by fostering a lasting relationship with South American and Mexican acousticians, culminating in the joint meeting in Cancun for which he served as Cochair. The Acoustical Society is clearly core to Jim's administrative work, and his effectiveness in this as in other arenas is defined by his focus on building personal relationships.

To define Jim by his copious and important work in acoustics is to miss his most important aspect–his warmth and generosity of spirit. As anyone who has interacted with Jim already knows, he enjoys forging deep relationships with people and encouraging their success. Jim is at his best when interacting with students at all levels and helping them understand how to learn by trying and how to challenge themselves. He was one of the founders of the AT&T Corporate Research Fellowship Program, the most significant corporate funding program for minority PhD students in the world. He is also one of the premier advisors for a new program in which high school junior and senior women spend a term working in a lab at the Johns Hopkins University campus. This program, called Wise Women, is now serving as a model for academic institutions across the country.

James Edward West has been a sterling colleague, collaborator, friend, and mentor to many of us. He is also a loving spouse to Marlene and a father to four children.

ILENE J. BUSCH-VISHNIAC
GARY W. ELKO
GERHARD M. SESSLER

Session 4aAAa

Architectural Acoustics, Musical Acoustics, and Engineering Acoustics: Composed Spaces Loudspeaker Concert I (Performance Session)

Alexander U. Case, Chair
Fermata Audio + Acoustics, P.O. Box 1161, Portsmouth, NH 03802

Works of music and sound art will be presented as a concert in this session. No lecture papers will be presented. A repeat performance of the compositions will be held in Session 4pAAb from 1:00 p.m. to 6:00 p.m. Attendees from all disciplines are invited to attend.

Invited Papers

8:00

4aAAa1. "Flowers and Wreaths": Radio theater and surround. Ean White (15 Channel Ctr. St., #514, Boston, MA 02210-3426, ean@incendiaryarts.org)

Originally produced in late 1992, ''Flowers and Wreaths'' is a setting for poems by Jacques Prvert (*paroles* Prvert) in French and in English translations by Lawrence Ferlinghetti and me. Prvert's success in France after World War II was largely due to his unabashed use of street language and his staunch antiwar, antirational political stance. The performance featured much original instrumentation: an amplified skull-resonator dental floss violone, a 35-ft.-long infrasonic electric monochord, a naval signal lamp, mechanized moving audience seating, the amplified electrocution of vegetables, unconventional speaker placement, etc. Mari Novotny-Jones and I continued to perform the piece through 1993. This new, surround recording is of only the title poem in the suite, ''Flowers and Wreaths,'' with a few elements which were originally used with other texts. It was a relatively quiet passage in an otherwise fairly raucous performance but was also a kind of spiritual center for the piece. The point of view for the listener during this extracted section is that of a corpse laying in a ditch by the side of a road. My intent is to explore the possibility of new, surround broadcast technology and its suitability for a resurgence of radio theater.

8:15

4aAAa2. "Passing Landscapes." James Moses (Music Dept., Brown Univ., Orwig Music Bldg., 1 Young Orchid Ave., Providence, RI, 02806)

This work was created through experimentation with surround recording and manipulation techniques. A variety of coincident and spaced microphone recording techniques are used to create realistic and immersive recordings of everyday environments. Processing and manipulation are then applied to these recordings to create musical material that plays against the recordings and guides us on a contextualized journey through the landscapes many of us pass though everyday.

8:30

4aAAa3. Vis-a-vis (2002), Katherine Bergeron, voice. Joseph Rovan (Music Dept., Brown Univ., 1 Young Orchard Ave., Providence, RI 02912, rovan@brown.edu)

What is the cause, and the cost, of insight? What does it mean to see the world face to face? These questions were raised by the young Rainer Maria Rilke in a passage from his Notebooks of Malte Laurids Brigge, a journal he kept in 1904 while serving as secretary to the sculptor Rodin. They also lie at the heart of Vis-a-vis, a multimedia work for voice and computer music soundscape, which takes Rilkes words as a dramatic point of departure. Recalling an encounter with a nameless woman in the streets of Paris, Rilke's text is a meditation on the nature of faces. It reads like a drama, a story in five virtual scenes told from multiple points of view and in several voices, alternately droll, philosophical, ironic, pathetic. In this new 5.1 mix the multivocal and multilingual aspects of Vis-a-vis are highlighted through the use of acousmatic space; the aim is to suggest the shifting psychological relationship between the woman who sings and the nameless woman who is the subject of Rilkes narration.

9:00

4aAAa4. Convergence. Andrew Carballeira (Driftwood Audio, 311 Concord Ave., Lexington, MA 02421)

In the words of Mahatma Gandhi, ''We must become the change we wish to see in the world.'' Exploitation, racism, violence, and militarism are all the rage in our world today. Persons of conscience opposed to this intoxicating madness are charged with envisioning something better. This piece is one in a series of attempts to draw a ray from our current predicament to a world free of war and racism. It opens with highly disjointed, atonal interactions physically separated in space. As the piece evolves, the voices begin to move together and speak more consonantly with each other: an aural simulation of the world uniting around the principles of peace, justice, and equality.

9:30

4aAAa5. Untitled Number 3. Andrew Clifford, Eli D. Cohn, Robert M. Mallory, and Alexander U. Case (Sound Recording Technol., Dept. of Music, UMass Lowell, 35 Wilder St., Ste. 3, Lowell, MA 01854, alex_case@uml.edu)

Untitled Number 3 was written and performed by Sparrows Swarm & Sing in 2005 and 2006. The work is a postrock instrumental for cello, violin, percussion, distorted electric guitar, rock drums, voices, handmade noisemakers, and digital signal processors. Composition is a collaborative process that occurs through both live improvisation among the six-piece ensemble and deliberate experimentation, rehearsal, and overdubbing.

9:45

4aAAa6. Moving Space. Pauline Oliveros (Rensselaer Polytechnic Inst., 110 Federal St., Troy, NY 12180, paulineo@deeplistening.org)

Moving Space (2006) is a composition for 5.1 surround sound system. Sustained sounds morph in quality as space changes and space changes as sounds move. The piece uses vector based amplitude panning (VBAP) and Monoverb mono implementation of the Schroeder/Moorer reverb model (mono version of freeverb~) to achieve these changes. Sounds are improvised and processed by the expanded instrument system (EIS), a MAX program designed by the composer consisting of multiple delays and processing algorithms.

10:00

4aAAa7. Rather the flight of the bird passing . . . . Thomas Plsek (Berklee College of Music, MS-1140 Brass, 1140 Boyslton St., Boston, MA 02215, tplsek@berklee.edu)

Rather the flight of the bird passing . . . is a five-channel composition created to produce a surreal sound environment that could only exist in a recorded, multichannel version. The solo trombonist in the piece (the composer) performs and is recorded by 5 human mobile microphone stands, i.e., people holding microphones and moving around the recording space searching for sounds. The sounds and the movements are influenced by a poem by Alberto Caeiro (1889–1915), a poet who did not actually physically exist. He was a heteronym created by the great Portuguese poet Fernando Pessoa (1888–1935). Pessoa created numerous heteronyms, each with their own styles and subject matter. The line of the poem, ''What's past is nothing and remembering is not seeing,'' is central to the practices used by all the performers to create the work.

10:30

4aAAa8. "Thaw/Twist at Bailey's." Stephan Moore (Troy, NY 12180, stephan@oddnoise.com)

A field recording of the 2004 spring thaw at Thomas Bailey's farmhouse in Richland, MI is used to explore the phenomena of temporal/spatial convergence and dispersal, as it can be used to manipulate and disorient our perception of location. Originally composed for a site-specific 16-channel array of custom-built speakers, this piece has been adapted to standard 5.1 surround sound. Duration: approximately 9 min.

11:00

4aAAa9. Two surround compositions: *Crossing Boundaries* and *The Puzzle Revealed*. Eric Chasalow (Music Dept., Brandeis Univ., MS051, Waltham, MA 02454-9110)

Crossing Boundaries (composed in 2000), was commissioned by Bates College to celebrate the millennium. The text aphorisms and reminiscences make a counterpoint of different spaces and moments to give different senses of the passage of historical and musical time. Musically, the piece moves from one large moment to the next one, usually reinforcing the phrasing of the text. Throughout, there is a shifting perception of just how fast we are moving. Musical sources include a Hebrew chant (which maintains a ghostly presence throughout the piece), Beethoven's piano music, improvised jazz, analog synthesized electronic music, and distorted electric guitar. *The Puzzle Revealed* (2006) is a maquette for an evening-long intermedia work, *The Puzzle Master*, a retelling of the Icarus myth by poet F. D. Reeve. This piece was conceived to specifically exploit the BeoLab 5 loudspeakers in 5.1 configuration. Material in the piece derives from vocal sources and explores morphing techniques that may become prominent in the larger work. The envelopment provided by full-range 5.1 playback is used to test ideas about the range of possibilities for spectral fusion or disintegration as a dynamic feature of different sound objects.

11:30

4aAAa10. Power chords, polyphonic trains, and plain old phase. David Moulton (Moulton Labs., 39 Ames Rd., Groton, MA 01450, dmoulton@moultonlabs.com)

Power chords, polyphonic trains, and plain old phase is a 36-min work of music composed directly for a six-channel array of loudspeakers (reduced to 5 channels for this performance). This piece draws on the work of minimalist sculptors such as Donald Judd and Michael Heizer, and explores the evolving nature of slowly shifting phase, reverberation, and the sonic continuum, using extremely simple sonic materials.

12:00

4aAAa11. "Anastasis." John Mallia (New England Conservatory of Music, 290 Huntington Ave., Boston, MA 02115)

"Anastasis" is a five-channel composition realized at the Institut de Musique Electroacoustique de Bourges (IMEB). The work is programmatic in its representation of Christ's "Descent into Limbo," a frequently treated subject in Medieval art. The majority of sounds used are derived from concrete sources, some of which carry conceptual weight. For example, the recurring sound of wax being scraped away from the metal surface of the candle trays at Bourges Cathedral St. Etienne: the residue of prayer . . . its removal, a daily chore (or performance ritual) captured in sound. Also, the emergence of a field recording of windmills at the work's close: transformation, by man, of nature's breath into energy as a storm blows in on the mountainside. The piece itself is a stormy, transformative descent. Sounds of friction and resistance represent the interpenetration of spheres of existence. Long intense strokes of sound such as the dragging of wood along wood, glass on glass, and the grinding and scraping of metals are fractured by abrupt, pointed declamations. Highly spatialized, bristling electronic textures occur as interruptions of the long descent and represent the accumulation of a mysterious, electrical energy in the empty tomb above; static glimpses of that middle sphere.

12:30

4aAAa12. Thirty-minute concert: Human-space-factory and Rojo. Hans Tutschku (Harvard Univ., Music Bldg., Cambridge, MA 02138, tutschku@fas.harvard.edu)

Human space-factory: The sound sources for this composition are derived from sung and spoken human sounds and from recordings in factories. Piano sounds serve as a pitched connection between the two sound worlds. The eight-channel space is used not only to place and to move the sounds, but to form a link between the sound families; by changing their position in space, the sound qualities transfer from one source to another. The three parts of the composition describe different densities: human sounds in an energetic space. The eight-channel electroacoustic composition is presented in a 5.1 mix; production: studio Musiques et Recherches Ohain; duration: 11.49 min; year: 1999; dedicated to Annette vande Gorne. Rojo plays with the imagination of simultaneous musical activities throughout the world. Recorded memories from several trips combine and exchange their qualities. Sometimes they change into huge sound masses, and occasionally their original environment is preserved. The composition creates an imaginary ritual, sourrounding the listeners, in which many sources meet and play together. The eight-channel electroacoustic composition is presented in a 5.1 mix; production: KlangProjekteWeimar (2004); commissioned by Radio Berlin Brandenburg; duration: 15.06 min; year: 2004; dedicated to Jonty Harrison.

THURSDAY MORNING, 8 JUNE 2006 BALLROOM B/C, 8:00 A.M. TO 12:00 NOON

Session 4aAAb

Architectural Acoustics: Spaces of Worship—Another Quarter Century of Experience (1984-Present) I (Poster Session)

Rose Mary Su, Chair
Acentech, 33 Moulton St., Cambridge, MA 02138

Contributed Papers

All posters will be on display from 8:00 a.m. to 12:00 noon. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 8:00 a.m. to 10:00 a.m. and contributors of even-numbered papers will be at their posters from 10:00 a.m. to 12:00 noon.

4aAAb1. The Holy Haram in Makkah, Saudi Arabia. Wasim A. Orfali and Wolfgang Ahnert (Inst. of Tech. Acoust., Tech. Univ. Berlin, Einsteinufer 25, 10587 Berlin, Germany, orfaliwasim@hotmail.com)

4aAAb2. Prophets Mohammed Mosques, Medinah, Saudi Arabia. Wasim A. Orfali and Wolfgang Ahnert (Inst. of Techn. Acoust., Tech. Univ. Berlin, Einsteinufer 25, 10587 Berlin, Germany, orfaliwasim@hotmail.com)

4aAAb3. Beth Israel Memorial Chapel, Houston, TX. Charles M. Salter (Charles M. Salter Associates, Inc., 130 Sutter St., Ste. 500, San Francisco, CA 94123)

4aAAb4. Abundant Life Christian Fellowship, Mountain View, CA. Charles M. Salter, P. E. and Jason R. Duty (Charles M. Salter Associates, 130 Sutter St., Ste. 500, San Francisco, CA 94123)

4aAAb5. St. Michael's Episcopal Church, Raleigh, NC. Frederick C. Schafer (F. C. Schafer Consulting, LLC, P.O. Box 726, Concord, NC 28026), Joseph F. Bridger, and Mathew M. George (Stewart Acoustical Consultants, Raleigh, NC)

4aAAb6. Our Lady of Lourdes Catholic School, Raleigh, NC. Frederick C. Schafer (F. C. Schafer Consulting, LLC, P.O. Box 726, Concord, NC 28026), Joseph F. Bridger, and Mathew M. George (Stewart Acoustical Consultants, Raleigh, NC 27607)

4aAAb7. St. Paul's Catholic Church, New Bern, NC. Frederick C. Schafer (F. C. Schafer Consulting, LLC, P.O. Box 726, Concord, NC 28026), Joseph F. Bridger, and Mathew M. George (Stewart Acoustical Consultants, Raleigh, NC 27607)

4aAAb8. St. Patrick's Episcopal Mission, Mooresville, NC. Frederick C. Schafer (F. C. Schafer Consulting, LLC, P.O. Box 726, Concord, NC 28026), Joseph F. Bridger, and Mathew M. George (Stewart Acoustical Consultants, Raleigh, NC 27607)

4aAAb9. University City United Methodist Church, Charlotte, NC. Frederick C. Schafer (F. C. Schafer Consulting, LLC, P.O. Box 726, Concord, NC 28026) and Joseph F. Bridger (Stewart Acoustical Consultants, Raleigh, NC 27607)

4aAAb10. St. Mary's Catholic Church, Johnson City, TN. Frederick C. Schafer (F. C. Schafer Consulting, LLC., P.O. Box 726, Concord, NC 28026), Joseph F. Bridger, and Noral D. Stewart (Stewart Acoust. Consultants, Raleigh, NC 27607)

4aAAb11. Acoustical design of the Tokyo Church of Christ. Satoru Ikeda (Nagata Acoust. Inc. Hongo-Segawa Bldg. 3F, 2-35-10 Hongo Bunkyo-ku, Tokyo 113-0033, Japan)

4aAAb12. Acoustical design of the Reinanzaka Church. Satoru Ikeda, Hideo Nakamura, and Minoru Nagata (Nagata Acoust. Inc. Hong-Segawa Bldg. 3F, 2-35-10 Hongo Bunkyo-ku, Tokyo 113-0033, Japan)

4aAAb13. Acoustical design of Shinanomachi Church. Masaya Uchida and Toshiko Fukuchi (Nagata Acoust. Inc., Hongo-Segawa Bldg. 3F, 2-35-10 Hongo Bunkyo-ku, Tokyo, Japan, uchida@nagata.co.jp)

4aAAb14. Acoustical design of Catholic Hatsudai Church. Masaya Uchida and Satoru Ikeda (Nagata Acoust. Inc., Hongo-Segawa Bldg. 3F, 2-35-10 Hongo Bunkyo-ku, Tokyo, Japan, uchida@nagata.co.jp)

4aAAb15. Multi-Faith Spiritual Center at Northeastern University, Boston, MA. Benjamin Markham and Douglas Sturz (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

4aAAb16. Sisters Chapel at Spelman College, Atlanta, GA. Benjamin Markham (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

4aAAb17. Vineyard Christian Fellowship, Cambridge, MA. Benjamin Davenny and Benjamin Markham (Acentech, Inc., 33 Moulton St., Cambridge, MA 02138)

4aAAb18. Shah Alam Mosque, Selangor, Malaysia. Benjamin Markham (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

4aAAb19. The Chapel of St. Peter and St. Paul, St. Paul's School, Concord, NH. Benjamin Markham and Michael Pincus (Acentech Inc., 33 Moulton St., Cambridge, MA 02138)

4aAAb20. Padre Pio Pilgrims' Church, Italy. Helmut A. Mueller, Gerhard Mueller, and Juergen Reinhold (Mller-BBM GmbH, Robert-Koch-Str. 11, 82152 Planegg, Germany)

4aAAb21. Saint Agnes Mission Church, Naples, FL. Gary Siebein, Hyun Paek, and Joshua Fisher (Siebein Associates, Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607, office@siebeinacoustic.com)

4aAAb22. Hyde Park United Methodist Church, Tampa FL. Gary Siebein, Martin Gold, Hyeongseok Kim, and Hyun Paek (Siebein Associates, Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607, office@siebeinacoustic.com)

4aAAb23. Palm Harbor United Methodist Church, Palm Harbor, FL. Gary Siebein, Hyun Paek, and Mark LoRang (Siebein Associates, Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607, office@siebeinacoustic.com)

4aAAb24. Grace Lutheran Church, Naples FL. Gary Siebein, Hyun Paek, and Joshua Fisher (Siebein Associates, Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607, office@siebeinacoustic.com)

4aAAb25. Calvary Chapel, West Melbourne, FL. Gary Siebein, Hyun Paek, and Mark LoRang (Siebein Associates, Inc., 625 NW 60th St., Ste. C, Gainesville, FL 32607, office@siebeinacoustic.com)

4aAAb26. Lakewood Church, Houston, TX. Richard Schrag (Russ Berger Design Group, 4006 Belt Line Rd., Ste. 160, Addison, TX 75001)

4aAAb27. Immaculate Conception Church, Boston, MA. K. Anthony Hoover and Christopher Storch (Cavanaugh Tocci Associates, Inc., 327 F Boston Post Rd., Sudbury, MA 02478)

4aAAb28. Arlington Street Church, Boston, MA. William J. Cavanaugh, Matthew J. Moore, and Christopher Storch (Cavanaugh Tocci Associates, Inc., 327 F Boston Post Rd., Sudbury, MA 02478)

4aAAb29. St. Dominic's Chapel, Providence College. Lincoln B. Berry, Matthew J. Moore, and Christopher Storch (Cavanaugh Tocci Associates, Inc., 327 F Boston Post Rd., Sudbury, MA 02478)

4aAAb30. Trinity Church, Boston, MA. Lincoln B. Berry, Matthew J. Moore, Christopher Storch (Cavanaugh Tocci Associates, Inc., 327 F Boston Post Rd., Sudbury, MA 02478), and Scott D. Pfeiffer (Kirkegaard Associates, Chicago, IL 60607)

4aAAb31. Temple Israel, Boston, MA. William J. Cavanaugh and Christopher Storch (Cavanaugh Tocci Associates, Inc., 327 F Boston Post Rd., Sudbury, MA 02478)

4aAAb32. Phillips Church, Phillips Exeter Academy, Exeter, NH. Timothy J. Foulkes and Christopher Storch (Cavanaugh Tocci Associates, Inc., 327 F Boston Post Rd., Sudbury, MA 02478)

4aAAb33. Old South Meeting House, Boston, MA. William J. Cavanaugh and Christopher Storch (Cavanaugh Tocci Associates, Inc., 327 F Boston Post Rd., Sudbury, MA 02478)

4aAAb34. Victory Christian Center, Charlotte, NC. D. Wayne Lee (Lee Sound Design, 1474 Stephens Dr., Atlanta, GA 30329, wlee@leesound.com)

4aAAb35. Saint Brigid Catholic Church, Alpharetta, GA. D. Wayne Lee (Lee Sound Design, 1474 Stephens Dr., Atlanta, GA, 30329, wlee@leesound.com)

4aAAb36. Al-Farooq Masjid Islamic Mosque, Atlanta, GA. D. Wayne Lee (Lee Sound Design, 1474 Stephens Dr., Atlanta, GA 30329, wlee@leesound.com)

4aAAb37. Christ Church of the Carolinas, Columbia, SC. Paul Henderson (Design 2020, 7786 Windsor Forest Pl., Harrisburg, NC 28075)

4aAAb38. Saint Nektarios Greek Orthodox Church, Charlotte, NC. Paul Henderson (Design 2020, 7786 Windsor Forest Pl., Harrisburg, NC 28075)

4aAAb39. The Chapel at Bethany Bible College, New Brunswick, Canada. Paul Henderson (Design 2020, 7786 Windsor Forest Pl., Harrisburg, NC 28075) The Chapel at Bethany Bible College, New Brunswick, Canada

4aAAb40. Saint Paul the Apostle Catholic Church, Juneau, AK. Julie A. Wiebusch (The Greenbusch Group, Inc., 1900 West Nickerson St., Ste. 201, Seattle, WA 98119)

4aAAb41. First Chinese Baptist Church, Los Angeles, CA. Bruce Walker (Channel Islands Acoustics, 676 West Highland Dr., Camarillo, CA 93010) and Neil Shaw (Menlo Science, Topanga, CA 90290)

4aAAb42. The Journey of Faith Christian Church, Manhattan Beach, CA. Neil A. Shaw (Menlo Sci. Acoust., Inc., P.O. Box 1610, Topanga, CA 90290-1610, menlo@ieee.org) and Douglas D. JoyceIII (Onyx Architects, Pasadena, CA 91101)

4aAAb43. Saint Mark Catholic Church. Dana Hougland and Kelly Stumpf (Shen Milsom Wike, 7000 South Yosemite St., Ste. 260, Englewood, CO 80112, kstumpf@smwinc.com)

4aAAb44. Buckley Air Force Base Chapel CO. Dana Hougland and Kelly Stumpf (Shen Milsom Wilke, 7000 South Yosemite St., Ste. 260, Englewood, CO 80112, kstumpf@smwinc.com)

4aAAb45. Our Lady of the Angels, Los Angeles, CA. Dennis Paoletti (SMW, 33 New Montgomery St., San Francisco, CA 94105)

4aAAb46. Christ the Light Cathedral, Oakland, CA. Dennis Paoletti (SMW, 33 New Montgomery St., San Francisco, CA 94105)

4aAAb47. Blessed Sacrament Cathedral, Sacramento, CA. Dennis Paoletti (SMW, 33 New Montgomery St., San Mateo, CA 94402)

4aAAb48. Holy Trinity Parish Church, Beaverton, OR. Russ Altermatt and Todd Matthias (Altermatt Associates, Inc., 522 SW Fifth Ave., Ste. 1200, Portland, OR 97204, raltermatt@altermatt.com)

4aAAb49. St. Joseph's Catholic Church, Wenatchee, WA. Russ Altermatt and Todd Matthias (Altermatt Associates, Inc., 522 SW Fifth Ave., Ste. 1200, Portland, OR 97204, raltermatt@altermatt.com)

4aAAb50. St. Bridgets Episcopal, Bellevue, WA. Michael Yantis (Yantis Acoustical Design, 220 Olive Way, Ste., 1400, Seattle, WA 98101, michaely@yantis.com)

4aAAb51. Bothell United Methodist, Bothell, WA. Michael Yantis (Yantis Acoustical Design, 220 Olive Way, Ste. 1400, Seattle, WA 98101, michaely@yantis.com)

4aAAb52. Saint Martin's Episcopal Church, Houston, TX. Ewart A. Wetherill (28 Cove Rd., Alameda, CA 94502) and John R. Prohs (Shen Milsom & Wilke, San Francisco, CA 94105)

4aAAb53. Saint Andrew's Episcopal Church, Amarillo, TX. Ewart A. Wetherill (28 Cove Rd., Alameda, CA 94502) and John R. Prohs (Shen Milsom & Wilke, San Francisco, CA 94105)

Session 4aAO

Acoustical Oceanography and Animal Bioacoustics: Exploitation of Sound by Marine Mammals I

Whitlow W. L. Au, Cochair
Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

David L. Bradley, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Chair's Introduction—8:25

Invited Papers

8:30

4aAO1. How toothed whales echolocate to find and capture prey. Peter L. Tyack, Mark Johnson (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543), Peter T. Madsen (Univ. of Aarhus, Aarhus C, Denmark), Walter M. X. Zimmer (NATO Undersea Res. Ctr., I-19138 La Spezia, Italy), and Natacha Aguilar de Soto (La Laguna Univ., Tenerife, Canary Islands, Spain)

Sperm and beaked whales are deep-diving toothed whales whose echolocation behavior has been studied using tags that record sound and whale orientation. Combining orientation data from tagged clicking whales with acoustic localization data from a towed hydrophone array allowed measurement of the three-dimensional beam pattern of regular clicks from sperm whales. These data confirm Mohl's bent horn hypothesis for sound production in sperm whales. During deep foraging dives, beaked whales produce several thousand regular clicks with interclick intervals (ICI) between 0.2 and 0.4 s. These search clicks have a distinctive FM upsweep from 25 to 55 kHz and a pulse length of 200–250 μs. The process of echolocating on prey can be divided into search, approach, and capture phases. The approach phase starts when a series of echoes from prey is detected. The capture phase starts when the target is about 2–5 m away, with a terminal buzz with click rates to ICIs of about 0.01 s. Tagged whales do not adjust their ICI to match range to target, nor do they adjust click level to maintain a constant echo level, but rather have a bimodal echolocation pattern using two kinds of clicks each with a typical ICI.

8:50

4aAO2. Exploitation of sound during predator-prey interactions: Killer whales and herring. Lee A. Miller (Inst. of Biol., Univ. of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark), Malene Simon (Greenland Inst. of Natural Resources, DK-3900 Nuuk, Greenland), Fernando Ugarte (Ministry of Fisheries & Hunting, DK-3900 Nuuk, Greenland), and Magnus Wahlberg (Aarhus Univ., 8000 Aarhus C, Denmark)

The majority of killer whales (*Orcinus orca*) in Norway and Iceland feed cooperatively on schools of herring (*Clupea harengus*). The whales swim towards the school and, at short distances, perform underwater tail slaps that produce thudlike sounds containing impulses resembling cavitation pulses in duration, frequencies, and amplitude. Herring are debilitated by the underwater tail slaps, probably due to a combination of factors like sound pressure gradients, particle movements, hydrodynamic effects, and physical contact with the tail or other fish. Following the underwater tail slaps, killer whales can easily capture individual flaccid herring [Simon *et al.*, J. Exp. Biol. **208**, 2459–2466 (2005); a video will be shown]. Icelandic killer whales have another strategy not shown by their Norwegian relatives. A whale can emit a 3-s, 680-Hz call that ends 1 s before the tail slap. The frequency of the call falls within the herring audiogram, but outside that of killer whales. This call seems suited for herding the herring into tighter groups, making it possible to debilitate more fish [Simon *et al.*, Behav. Ecol. Sociobiol. (submitted)]. Herring are not defenseless. The school can produce a flatulent bubble net that could hinder detection by killer whale biosonar [Wahlberg Westerberg, Aquatic Living Resources **16**, 271–275 (2003)].

9:10

4aAO3. Hyper-directionality in clicks from the sperm whale (*Physeter macrocephalus*). Bertel Moehl, Magnus Wahlberg, and Anders Heerfordt (Inst. of Biological Sci., Zoophysiology, Bygn. 1131, C. F. Moellers Alle 133, DK-8000 Aarhus C, Denmark, bertel.moehl@biology.au.dk)

Directivity is an important determinant of animal sonar performance but difficult to assess for large, free-ranging species. Recent field studies of sperm whales indicate usual clicks to be quite directional [directivity index: 27 dB at a transmitting cross section to wavelength ratio ($2a/wl$) of 8], with indications of even narrower beams. To obtain better measures of directivity, a fiber-optically linked, linear hydrophone array with up to ten receivers, spaced 50 m apart, has been designed. The array can be deployed to depths of 1 km and incorporates a transmitting transducer. The array was deployed off Andenes, Norway, in July 2003, 2004, and 2005. On five occasions, transmissions of short broadband pulses (peaking at 14 kHz, SL 135 dB *re*: 1 Pa) apparently induced sperm whales at

ranges of about 2 km to train their sound beams towards the transmitter and lock them there. This implies beam-steering from the whales' auditory system, having a high auditory directionality, and detection capability of signals at 65 dB *re*: 1 Pa. It resulted in a high yield of on-axis recordings with maximum directivity indices of 39 dB, incompatible with plane piston radiation at $2a/wl$ ratio of 8.

9:30

4aAO4. The Humpback Whale Acoustic Research Collaboration studies on the use of sound by humpback whales. Douglas H. Cato (Defence Sci. & Tech Org., and Univ. of Sydney, P.O. Box 44, Pyrmont, NSW 2009 Australia), Michael J. Noad (Univ. of Queensland, St. Lucia, QLD 4072 Australia), Dale Stokes (Scripps Inst. of Oceanogr., La Jolla, CA 92093), Rebecca A. Dunlop, Joshua N. Smith (Univ. of Queensland, St. Lucia, QLD 4072 Australia), Patrick J.O. Miller (Univ. St. Andrews, Scotland, KY168LB, UK), Nicoletta Biassoni (Woods Hole Oceanogr. Inst., Woods Hole, MA 02543), Eric T. Kniest (Univ. of Newcastle, Callaghan, NSW 2038, Australia), John Noad (100 Monash Rd., Tarragindi, Queensland 4121, Australia), Aaron M. Thode, and Grant B. Deane (Scripps Inst. of Oceanogr., La Jolla, CA 92093)

Humpback whales produce two different types of vocalizations with apparent difference functions: songs and social sounds. There is also the potential to use the sounds of their environment, such as surf noise during migration along coasts, though this has not been demonstrated. The Humpback Whale Acoustic Research Collaboration (HARC) is a series of well-controlled, multiplatform experiments designed to improve our understanding of how humpback whales use sound, both vocalizations and sounds of their environment. There have been two major experiments with humpback whales migrating along the east coast of Australia. Whales passing within 10 km were tracked acoustically and visually during daylight hours in conjunction with behavioral observations, and DTAGs were used on some whales to record their 3D movements and the sound field to which they were exposed. There was also playback of social sounds and full characterization of the acoustics of the environment, including the ambient noise exposure. Preliminary results of interactions between singing and nonsinging whales, the use of song as a spacing mechanism, responses to playback, and context of social sounds will be discussed. [Research supported by ONR and DSTO.]

Contributed Papers

9:50

4aAO5. Do humpback whales exploit bioacoustic absorption spectroscopy to detect and classify fish schools? Orest Diachok (Appl. Phys. Lab., Johns Hopkins Univ., Laurel, MD 20723)

Humpback whales employ bubble nets for catching fish. In the Gulf of Alaska the method involves acoustic communication between whales: one whale, situated below the school, transmits sounds and drives the fish toward the surface; a second whale, situated above the school, deploys a bubble net (as wide as 30 m), which encircles the school as it approaches the surface. Before and during this process one (or more) humpback emits signals, called trumpet sounds, which consists of a series of cw tones, which are harmonics of ~300 Hz. The target species in the Gulf of Alaska is adult, 22-cm-long herring. These signals cover the resonance frequency of the target species swim bladders, which is about 1.5 kHz at 30 m. Resonance frequencies of individual fish are determined by the radius and depth of the swim bladder; significantly lower bubble cloud resonance frequencies of schools are determined by the void ratio and the dimensions of the school. One of the functions of these vocalizations may be detection and classification: the listening whale, near the surface, may be capable of interpreting the high absorption losses at resonance frequencies [Diachok, O. ''Effects of absorptivity due to fish on transmission loss in shallow water,'' J. Acoust. Soc. Am. **105**, 2105–2128 (1999)] to infer fish length, species, and size of school.

10:05–10:20 Break

10:20

4aAO6. The role of the spectral characteristics of discrete calls in killer whale (*Orcinus orca*) communication. Filipa Samarra (CREEM, The Observatory, Buchanan Gardens, Univ. of St. Andrews, St. Andrews, Fife KY16 9LZ, UK, fips2@st-andrews.ac.uk) and Patrick Miller (Univ. of St. Andrews, St. Andrews, Fife KY16 8LB, UK)

Many studies of killer whale calling have focused on group-specific time-frequency contours. This study investigated variation in the spectral structure of calls recorded from identified individuals and sought to identify factors that influence call spectral structure. The energy level of the first two harmonics was measured at each 0.01-s interval and linear mixed effects models used to test the sources of variation in the relative level of the first and second harmonics. Within a call, spectral structure changes according to the frequency of the fundamental ($p<0.001$), but not time since call start. Between calls, variation seems related to body size. Adult males had a stronger second harmonic, whereas for females the opposite tended to be true. Size class was a significant factor in the model ($p=0.0108$), while duration of calls or orientation and range of the signaller were not. Further analysis indicates different energy patterns between males and females up to the fifth harmonic. These results are consistent with a source-filter model of sound production, with formants arising from filtering effects of the sound-production apparatus. However, more research is clearly needed on the sound production mechanism of this species.

10:35

4aAO7. Whistle characteristics of bottlenose dolphins (*Tursiops truncatus*) are conserved independently of whistle quality. Carmen Bazúa-Durán (Lab. Acústica, Facultad de Ciencias, UNAM, Circuito Exterior s/n, Ciudad Universitaria, 04510 México, D.F., Mexico, bazua@servidor.unam.mx)

Bottlenose dolphins use their acoustic sense (whistles and pulses) to communicate and inspect their environment. One of these environments is a coastal lagoon in the southern Gulf of Mexico, the Laguna de Términos, where bottlenose dolphins inhabit all year round in herds varying from 1–120 individuals, with a yearly population of about 600 dolphins. Several characteristics of dolphin whistles were measured: quality (whistles with a signal-to-noise ratio, SNR, greater or less than 3 dB), repertoire (by classifying contours as either constant, ascending, descending, concave, convex, or multiple), and time-frequency parameters (duration, average frequency, medium frequency, and frequency range) to investigate how the overall whistle structure varied. Results show that regardless of their quality (SNR), whistles were not significantly different in their parameters and repertoire. Additionally, in this area dolphins produce more whistles/min/dolphin during traveling than during social interactions and feeding (whistle production was lowest during resting). Dolphins are presumed to use whistles to maintain group cohesion and to broadcast identity, thus being important in determining dolphin social structure. The results of this work support these hypotheses, suggesting that faint (whispered and/or far range) or strong (loud and/or close range) whistles broadcast, overall, similar information. [Work supported by CONACyT-Campeche, PAPIIT-UNAM, and ICMyL-UNAM.]

10:50

4aAO8. Differences in whistles of bottlenose dolphins between two populations in southeastern Australia. Michelle Lemon (Marine Mammal Res. Group, Macquarie Univ., NSW, 2109, Australia), Douglas H. Cato (Defence Sci. and Technol. Organisation, Pyrmont, NSW, 2009, Australia), and Robert Harcourt (Macquarie Univ., NSW, 2109, Australia)

The acoustic repertoires of two populations of Indo-Pacific bottlenose dolphins, *Tursiops aduncus*, have been studied in comparison with surface behaviors in southeast Australia. These populations are largely genetically differentiated and geographically isolated, one being in Port Stephens and the other in Jervis Bay about 300 km to the south. Ten variables were measured from the fundamental frequency of 200 whistle contours from each population, recorded in various behavioral contexts. Principal component analysis revealed significant differences between the populations. The Port Stephens dolphin population have less complex whistles of a higher frequency than those in Jervis Bay. Although significant differences were observed in phonation rates between behaviors at either site, there was no interaction between site and behavior. The genetic difference and degree of isolation of the populations suggest that cultural drift may play a part in the differences in whistles between sites. Ambient noise measured during observations showed significant differences between sites, but the measurements did not fully sample the range of ambient noise and differences are small compared to the expected temporal variation.

11:05

4aAO9. Verified passive acoustic detection of beaked whales (*Mesoplodon densirostris*) using distributed bottom-mounted hydrophones in the tongue of the ocean, Bahamas. David Moretti, Ronald Morrissey, Nancy DiMarzio, and Jessica Ward (Naval Undersea Warfare Ctr. (NUWC), Code 71, Bldg. 1351, Newport, RI 02841)

Passive detection of beaked whales has become increasingly important as at least two species, Blainville's beaked whales (*Mesoplodon densirostris*) and Cuvier's beaked whales (*Ziphius cavorostris*) have stranded in events associated with sonar. Passive acoustic detection and localization algorithms develop as part of the Office of Naval Research (ONR) Marine Mammal Monitoring on Navy Range (M3R) program will be presented along with results from a series of tests that demonstrate the efficacy of passive acoustic detection of vocalizing *M. densirostris*. Ninety-two bottom-mounted sensors at the Atlantic Undersea Test and Evaluation Center (AUTEC) were monitored for vocalizations. M3R systems were used to detect and localize vocalizing animals. Based on these real-time data, trained observers were vectored to the animals and verified the species. Focal follows coordinated with passive acoustics were completed. Observer photos along with raw sensor, detection and localization data were collected.

THURSDAY MORNING, 8 JUNE 2006 BALLROOM A, 8:00 A.M. TO 12:00 NOON

Session 4aBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Celebration Session for Edwin Carstensen I

Lawrence A. Crum, Cochair
Univ. of Washington, Applied Physics Lab., 1013 NE 40th St., Seattle, WA 98105-6698

Diane Dalecki, Cochair
Univ. of Rochester, Dept. of Biomedical Engineering, Rochester, NY 14627

Chair's Introduction—8:00

Invited Papers

8:05

4aBB1. Tissue-ultrasound interactions: Absorption to imaging artifacts. Frederick Kremkau (Ctr. for Medical Ultrasound, Wake Forest Univ. School of Medicine, Winston-Salem, NC 27157-1039, fkremkau@wfubmc.edu)

Attenuation determines fundamental limitations in sonography. It directly limits the imaging depth (penetration) and indirectly limits the detail, contrast, and temporal resolutions by limiting the frequency. These resolutions improve with increasing frequency but penetration requirements determine the upper limit on frequency in each imaging situation because attenuation (and therefore penetration) is frequency dependent. Attenuation in soft tissue is dominated by absorption (typically about 90%). Absorption occurs primarily at the molecular level. Sonographic artifacts occur when one of the pulse-echo imaging assumptions is violated. These assumptions are that sound travels in straight lines, that echoes originate only from objects located on the beam axis, that the amplitude of returning echoes is related directly to the reflecting or scattering properties of distant objects, and that the distance to these objects is proportional to the round-trip travel time (13 μs/cm of depth). If scattering strength of an object is high but its absorption is low, a hyperechoic presentation will occur with distal enhancement. A more common occurrence is hyperechoic with distal shadow indicating high echogenicity and attenuation, therefore high absorption. Examples of several imaging artifacts will be shown with discussion of the propagation assumptions violated.

8:25

4aBB2. Phased arrays for high-intensity focused ultrasound therapy. Leon A. Frizzell (Bioacoustics Res. Lab., Univ. of Illinois, 405 N. Mathews Ave., Urbana, IL 61801)

Several groups have examined the design of ultrasound phased array probes for transrectal treatment of the prostate. An effective design must use the entire aperture of the array and maintain relatively low grating lobe intensities, ideally while minimizing the number of elements and associated phasing circuitry. A new approach to the design of a cylindrical array was developed. The cylindrical aperture of the array was divided into six columns of elements oriented along its length. Because of the symmetry when focus formation and scanning are limited to a plane oriented along the length of the array and perpendicular to it, the six columns are really three column pairs such that each column in a pair has the same phasing applied to its elements. The unique feature of this new design is that a different center-to-center spacing (different widths) for the elements was applied to each column pair. This results in a different location for the grating lobes for each of the three column pairs. This leads to much better performance when looking at grating lobe intensity levels compared to the intensity at the focus. [Work supported by NIH and NCSA.]

8:45

4aBB3. Thresholds for sound-induced lung hemorrhage for frequencies from 100 Hz to 1 MHz. Diane Dalecki, Sally Z. Child, and Carol H. Raeman (Dept. of Biomed. Eng., Hopeman 309, Box 270168, and the Rochester Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627)

Edwin L. Carstensen has made outstanding and wide-ranging contributions to the field of biomedical ultrasound. His many achievements span the areas of bioeffects of ultrasound, acoustic cavitation, lithotripsy, thermal and mechanical mechanisms, and nonlinear acoustics. In 1990, Carstensen first reported that pulsed ultrasound at diagnostic exposure conditions could produce mammalian lung hemorrhage [Child *et al.*, Ultrasound Med. and Biol. **16**, 817–825 (1990)]. Recent work from our lab has quantified the thresholds for murine lung hemorrhage over a range of acoustic frequencies from approximately 100 Hz to 1 MHz. Various exposure systems were used to generate acoustic fields over this broad frequency range in the laboratory. Through several different investigations, we have shown that murine lung responds to low-frequency underwater sound as a resonant structure. The resonance frequency of adult murine lung is approximately 325 Hz, and the pressure threshold for lung hemorrhage is lowest at the resonance frequency. The threshold increases for frequencies above lung resonance. The equation $P_{th}=0.01f^{0.64}$, where P_{th} is the threshold pressure in MPa and f is the acoustic exposure frequency in kHz, approximates our experimental lung threshold data, for long pulse durations, over the 2.5–1000-kHz range.

9:05

4aBB4. Nonlinear acoustics in E. L. Carstensen's career. David T. Blackstock (Appl. Res. Labs and Dept. of Mech. Eng., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

HIFU and harmonic imaging are hallmarks of present day biomedical ultrasound. But it wasn't always this way. At one time linear theory ruled supreme. Although by the early 1970s nonlinear acoustics had made its way into many areas of acoustics, e.g., physical acoustics, underwater sound, and aeroacoustics, biomedical ultrasound remained a safe haven for small-signalists. Ed Carstensen changed all that. By the mid-1970s he realized that linear theory could not account for certain phenomena he observed. At the 1978 Allerton Conference his invited talk entitled ''Nonlinear Aspects of Ultrasonic Absorption'' may have been the first public disclosure that nonlinear effects are important in biomedical ultrasound. His first archival work appeared as two papers in 1980; Tom Muir was a coauthor for both. During the next decade Ed and a series of colleagues showed that absorption could easily be dominated by nonlinear propagation effects. In turn the increased absorption causes increased heating, an important practical application. Ed's pioneering work paved the way for many of the well-known applications today.

9:25

4aBB5. Medical imaging using nonlinear ultrasound and the role of Edwin Carstensen. Kevin J. Parker (Univ. of Rochester, Dept. of Elec. and Comput. Eng., P.O. Box 270126, Rochester, NY 14627)

In the 1960s, with the development of weak shock theory by Blackstock and other advances, nonlinear acoustics found growing importance and applications in the atmosphere and underwater. By comparison, during this time frame the field of medical ultrasound remained largely focused on linear mechanisms. Three major subfields within medical ultrasound eventually developed major nonlinear theory and applications: cavitation, lithotripsy, and imaging. Edwin Carstensen's collaborative research and directorship of the Rochester Center for Biomedical Ultrasound played an important role in these developments. This talk focuses on the development of nonlinear acoustics to clinical imaging, tracing benchmark developments from the 1960s to the early 2000s, under the guiding influence of Ed Carstensen.

9:45

4aBB6. Stress, strain, and flow produced by a vibrator in, or on the surface of, a soft solid. Wesley L. Nyborg and Harold M. Frost (Phys. Dept., Univ. of Vermont, Burlington, VT 05405)

The field of biomedical ultrasound is greatly indebted to Edwin Carstensen for important contributions, without number, that he and his associates have made to understanding of the subject. These contributions have dealt not only with linear and nonlinear propagation of ultrasound in biological materials, but also with effects produced by ultrasound through various thermal and nonthermal mechanisms. Such understanding is important for advancing benefits and minimizing risks in applications of ultrasound. Many therapeutic applications discussed in the literature utilize focused beams of megahertz frequency, while others employ lower frequencies. In this talk, some findings will be presented from experiments in which a vibrating source of frequency in the range 20–90 kHz is brought into contact with the surface of a soft viscoelastic solid and information is obtained on resulting fields of strain. Under some

conditions the mechanical properties of the solid are altered by an exposure, in an apparent change of internal structure, and the change is maintained (recorded) until erased by a later sonication. The role of radiation force and other mechanisms in producing such effects will be discussed, as well as their possible relevance to ultrasonic angioplasty and other applications.

10:05–10:20 Break

10:20

4aBB7. Statistical estimation of ultrasonic propagation path parameters for aberration correction. Robert C. Waag and Jeffrey P. Astheimer (Dept. of Elec. and Comput. Eng., Univ. of Rochester, Rochester, NY 14627)

Parameters in a linear filter model for ultrasonic propagation are found using statistical estimation. The model employs an inhomogeneous-medium Green's function that is decomposed into a homogeneous-transmission term and a path-dependent aberration term. Power and cross-power spectra of random-medium scattering are estimated over the frequency band of the transmit-receive system by using closely situated scattering volumes. The frequency-domain magnitude of the aberration is obtained from a normalization of the power spectrum. The corresponding phase is reconstructed from cross-power spectra of subaperture signals at adjacent receive positions by a recursion. The subapertures constrain the receive sensitivity pattern to eliminate measurement system phase contributions. The recursion uses a Laplacian-based algorithm to obtain phase from phase differences. Pulse-echo waveforms were acquired from a point reflector and a tissuelike scattering phantom through a tissue-mimicking aberration path from neighboring volumes having essentially the same aberration path. Propagation path aberration parameters calculated from the measurements of random scattering through the aberration phantom agree with corresponding parameters calculated for the same aberrator and array position by using echoes from the point reflector. The results indicate the approach describes, in addition to time shifts, waveform amplitude and shape changes produced by propagation through distributed aberration under realistic conditions.

10:40

4aBB8. Thermal dose and the probability of adverse effects from biomedical ultrasound. Charles C. Church (Natl. Ctr. for Physical Acoust., Univ. of Mississippi, University, MS 38677)

Absorption of biomedical ultrasound heats tissue. Moderate heating to a few degrees above normal physiological temperatures can perturb biological systems, e.g., by altering normal metabolic processes, while brief, intense heating can kill cells, tissues, or entire organisms. Of particular concern for diagnostic ultrasound is thermally induced teratogenesis, which has been demonstrated in many animal species (chicken, mouse, rat, guinea pig, monkey) and apparently also occurs in humans. The nature and interpretation of such studies will be discussed. In therapeutic ultrasound, e.g., HIFU, the interest is in assuring the destruction of diseased tissue while sparing adjacent, healthy tissue. The concept of thermal dose will be described and used to quantify the probability and extent of biological effects expected from both diagnostic and therapeutic exposures. Additional factors relevant to thermal effects will also be considered, including the nature of the acoustic field *in situ*, as well as the state of perfusion and the variation in thermal sensitivity of the exposed tissue. It is concluded that adverse thermal events following diagnostic exposure are usually extremely rare, although there may be some reason for concern in a few special cases, while HIFU may induce significant collateral damage to proliferating cells pre- and postfocal zone.

Contributed Papers

11:00

4aBB9. The relation between lung damage induced by acoustic excitation and the subharmonic response of bubbles. Jonathan Young, Sheryl M. Gracewski (Dept. of Mech. Eng. and the Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627, grace@me.rochester.edu), Stephen McAleavy, and Diane Dalecki (Univ. of Rochester, Rochester, NY 14627)

Edwin L. Carstensen has made significant contributions to determining thresholds for and understanding mechanisms of mammalian lung hemorrhage caused by both ultrasound and lithotripter shock waves. The work presented here extends this knowledge base to investigate mechanisms of lung damage caused by acoustic excitation at the lung resonance frequency. This situation is relevant to divers exposed to high-amplitude, low-frequency sonar. Based on preliminary experiments, it was hypothesized that the threshold for lung hemorrhage at resonance occurs when lung surface oscillations at the subharmonic frequency of order one-half become unstable, producing significantly larger lung displacements. A balloon model was developed to approximate the behavior of small mammalian lung. The theoretical model for surface mode oscillations of a bubble was modified to account for the initial tension in the balloon membrane and used to investigate the stability of subharmonic oscillations. Reasonable agreement between the theoretical model and experimental measurements of the balloon surface displacements were obtained. In addition, a sharp pressure threshold for the presence of subharmonic oscillations was observed in the displacement data for the balloon model. The relevance of these simulations and measurements to murine lung hemorrhage will be discussed.

11:15

4aBB10. Some relationships among phase cancellation at a phase-sensitive receiving transducer, phase aberration in diagnostic medical imaging, and nonlinear imaging. James G. Miller, Kirk D. Wallace, and Mark R. Holland (Dept. of Phys., Washington Univ., St. Louis, MO, 63130, james.g.miller@wustl.edu)

A 1975 J. Acoust. Soc. Am. article by Marcus and Carstensen [Marcus & Carstensen, ''Problems with absorption measurements of inhomogeneous solids,'' J. Acoust. Soc. Am. **52**, 1334–1335 (1975)] reported perhaps the earliest study of potential artifacts in medical ultrasonics arising from phase-cancellation effects at the face of a phase-sensitive receiving transducer. This paper recognized the insights offered in a 1966 paper by Southgate [Southgate, ''Use of power-sensitive detector in pulse-attenuation measurements,'' J. Acoust Soc. Am. **39**, 480–483 (1966)] proposing a phase-insensitive receiver based on the acousto-electric effect to reduce phase-cancellation effects. Also in the 1970s, the Thurston lab at Duke and our lab in St. Louis examined consequences of propagating an ultrasonic field through the intrinsically inhomogeneous medium of soft tissue. Subsequently, several laboratories developed and implemented sophisticated algorithms applied to signals from receiving transducer arrays to provide some improvement. One of the most significant strides in reducing aberration artifacts came about serendipitously with harmonic mode (nonlinear) imaging. We will present experimental studies from our laboratory that examine approaches for reducing the profound effects arising from phase cancellation at the face of a phase-sensitive receiving transducer on measurements of ultrasonic attenuation and backscatter. Measurements utilizing phase-insensitive receiving transducers based on

the acoustoelectric effect and measurement using two-dimensional arrays of pointlike receivers to characterize tissue-mimicking materials and intrinsically anisotropic hearts will be discussed. [Work supported by NIH R37-HL40302 and R01-HL72761.]

11:30

4aBB11. Ultrasound hyperthermia in clotted blood and *ex vivo* cranial bone. Volodymyr M. Nahirnyak (Dept. of Phys., Univ. of Cincinnati, 231 Albert Sabin Way, Cincinnati, OH 45221-0511) and Christy K. Holland (Univ. of Cincinnati, Cincinnati, OH 45267-0586)

Ultrasound thermal effects have been hypothesized to contribute moderately to ultrasound-assisted thrombolysis. Transcranial ultrasound adjuvant to thrombolytic treatment of ischemic stroke requires a detailed investigation of the heating produced by ultrasound exposure to the skull bone, brain, and blood clots. Thermocouple measurements of *ex vivo* cranial bone and blood clot heating were compared to theoretical predictions of ultrasound hyperthermia, which assumed adiabatic conditions and provided an estimate for the worst-case scenario. The *in vitro* experiments were conducted at two temperatures, 20 °C and 37 °C. *Ex vivo* human skull, and porcine or human blood clots were exposed to 120 kHz, 1 MHz, or 3.5 MHz pulsed ultrasound with an 80% duty cycle and an acoustic spatial peak intensity of 0.5 W/cm^2. The thermal elevation in human cranial temporal bone was below 1 °C for all three frequencies. The ultrasound hyperthermia in porcine and human clots increased with an increase in center frequency, but did not exceed (0.33 ± 0.06) °C. These results provide the information necessary to quantify the contribution of a thermal mechanism for ultrasound-enhanced thrombolysis.

11:45

4aBB12. Acoustic radiation force imaging of prostate: Initial results. Stephen McAleavey, Menoj Menon (BME Dept., Univ. of Rochester, 308 Hopeman, Rochester, NY), and Deborah Rubens (Univ. of Rochester Med. Ctr., Rochester, NY)

Dr. Edwin Carstensen has contributed significantly to the understanding of the mechanical forces produced by acoustic waves in biological tissues, including work on cavitation and tactile perception of acoustic radiation force. This work presents some initial results in using radiation force for remote palpation of prostate tissue. Acoustic radiation force impulse (ARFI) images of excised human prostates are presented. Fresh prostates obtained immediately after surgery were scanned in an isotonic saline bath at room temperature. Scanning was performed with a Siemens Antares scanner and VF10-5 linear array. Tissue displacement was induced by pushing pulses of 30 to 75/μ s duration at 6.67 MHz and an I_{sppa} on the order of 1 kW. Observed peak displacements in healthy tissue were in the range of 6–12/μ m. Most prostates were scanned free-hand. One specimen was scanned with a transducer mounted to a single axis stage and imaged at 1-mm steps. The resulting images are presented with matching histology images and show good correspondence. [Work supported by the Wallace H. Coulter Foundation Early Career Award for Translational Research program.]

THURSDAY MORNING, 8 JUNE 2006 ROOM 552B, 8:00 TO 11:30 A.M.

Session 4aEA

Engineering Acoustics: Microphones and Loudspeakers

Daniel M. Warren, Chair
Knowles Electronics, 726 Lancaster Ln., Geneva, IL 60134-3433

Contributed Papers

8:00

4aEA1. Acoustic noise in microphones from the thermal boundary layer resistance in enclosures. Stephen C. Thompson (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804) and Janice L. LoPresti (Knowles Electronics LLC, 1151 Maplewood Dr., Itasca, IL 60143)

Many types of acoustic sensors include a small enclosure as an integral part of the device. For example, a condenser microphone has a volume enclosed behind its diaphragm so that the acoustic pressure drives only the front of the diaphragm. In other sensors the enclosure may exist mainly to protect the sensitive components or to provide electrostatic shielding to the high impedance parts of the electrical circuitry. Whatever their reason, these volumes can play a significant role in the performance of the sensor. The adiabatic compliance of the enclosure is commonly included in device models. However, these enclosures are often small enough that the thermal boundary layer thickness can consume a substantial fraction of the enclosed volume at low frequencies. Thermal effects modify both the real and imaginary parts of the cavity impedance in ways that may be important in determining the sensitivity and internal noise of the device. This paper examines these effects. One interesting consequence is a thermal noise source that has a power spectral density proportional to 1/*f* throughout a substantial frequency band.

8:15

4aEA2. A ladder network impedance model for lossy wave phenomena. Daniel M. Warren and Janice L. LoPresti (Knowles Electronics, 1151 Maplewood Dr., Itasca, IL 60143, daniel.warren@knowles.com)

The ratio of Bessel functions $J_1(z)/J_0(z)$ often occurs in the calculation of impedance of lossy components in electroacoustic devices. In particular, the electric impedance of hearing aid receivers or loudspeakers has a term proportional to a Bessel function ratio of this form when losses due to eddy currents are taken into account. This so-called *semi-inductance* has been notoriously difficult to model in simple lumped-parameter circuit calculation tools such as SPICE. This paper presents a simple ladder network derived from the Bessel function ratio expressed as a continuous fraction. The same transformation has been used to model the acoustic impedance of small cavities with losses at the thermal boundary layer.

8:30

4aEA3. Contribution of cross talk to the uncertainty of microphone calibrations by the electrostatic actuator method. Allan J. Zuckerwar (Analytical Services and Mater., 107 Resrearch Dr., Hampton, VA 23666, ajzuckerwar@yahoo.com) and Richard M. Faison (Simco Electronics, Newport News, VA 23606)

Cross talk in electrostatic actuator calibrations is defined as the response of a test microphone to the actuator excitation voltage when the

microphone membrane is frozen, i.e., when the actuator polarization voltage is zero—an unavoidable nuisance effect that contributes to the calibration uncertainty. Two models for the cross talk are considered. In the first, the actuator excitation voltage is coupled to the microphone output through the stray capacitance between the actuator electrode and the microphone backplate; in the second, it is coupled through the ground resistance, which provides a common element between the input and output loops. In the stray capacitance model, at frequencies below the membrane resonance, the cross talk is independent of frequency, but in the ground resistance model it varies inversely with frequency. The frequency dependence of the cross talk thus serves as a discriminator between the two models. Measurements on condenser microphones of various sizes yield results that favor the stray capacitance model. The ratio of cross-talk signal to actuator pressure signal (at 94 dB *re*: 20 mPa) is found experimentally to be −74 dB for a 1-in. condenser microphone and nearly the same for 1/2-in. and 1/4-in. microphones.

8:45

4aEA4. Microphone bandwidth extension by digital correction of low-frequency roll-off. Timothy Marston and Thomas Gabrielson (Grad. Program in Acoust., Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802, tmm357@psu.edu)

When the low-frequency roll-off of a condenser microphone needs to be extended, the standard method of achieving extra bandwidth is to place a shunt capacitor in parallel with the input capacitance of the preamplifier. The result increases the bandwidth in the desired region, but sacrifices dynamic range across the entire bandwidth. It is possible to digitally extend the useful bandwidth of a microphone and only compromise dynamic range in the extended region. A technique for doing this has been developed and applied to microphones used to record sonic boom data. Microphones that have an insufficiently low-frequency response to capture the characteristic *N*-wave shape of sonic booms are corrected using the developed methods, and the results are compared to data captured by microphones with adequate low-frequency response. During the development of the filtering technique, it was found that phase response is critical in the correction process and its importance is demonstrated. [Research funded by the Applied Research Laboratory Educational and Foundational Fund.]

9:00

4aEA5. Infrasonic pistonphone calibration of acoustic transducers. Timothy Marston and Thomas Gabrielson (Grad. Program in Acoust., Penn State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802, tmm357@psu.edu)

A simple infrasonic pistonphone was developed for the low-frequency calibration of acoustic sensors. The pistonphone, which utilizes a variable speed dc motor and a crank-driven piston, has a frequency range of 20 mHz to 20 Hz. Digital techniques have been implemented to correct for the inherent distortion in the crank-driven piston design, enabling simpler mechanical construction without significant loss in calibration accuracy. As in similar design geometries [H. E. Dahlke and John J. Van Houten, J. Acoust. Soc. Am. **37**, 1209 (1965)], the transition from an adiabatic to an isothermal pressure-volume relation is readily observed within the specified frequency range and results in a 0.5-dB reduction in calibration signal amplitude. A digital model for this transition is compared to data acquired using the pistonphone, and possibilities for increasing the calibration accuracy by use of this digital model are presented. [Research funded by the Applied Research Laboratory Educational and Foundational Fund.]

9:15

4aEA6. Multimicrophone method for measurement of acoustic absorption. Charles H. Oppenheimer (DSA, 4900 SW Griffith Dr., Ste. 216, Beaverton, OR 97005)

A multimicrophone method for measurement of the acoustic absorption of small material samples using impedance tubes is presented. The multimicrophone method may be implemented with actual or synthetic microphone arrays, the latter involving successive measurements and a roving microphone, and is an alternative to the two-microphone method of ASTM E 1050. The effects of absorption measurement inaccuracy due to microphone response (magnitude and phase) bias and measurement noise are determined by simulation. Simulation results indicate that the multimicrophone method is more robust against microphone bias and noise and is therefore more accurate than the standard method, especially as microphone counts increase, while requiring minimal additional effort and cost.

9:30

4aEA7. Impulse response of a spherical microphone array (eigenmike). Colin Barnhill, James West, and Ilene Busch-Vishniac (Johns Hopkins Univ., 3100 Wyman Park Dr., Rm. 157, Baltimore, MD 21211, shaolin_cb@yahoo.com)

Spherical microphone arrays are capable of complete three-dimensional sound field capture and robust noise cancellation [J. Meyer & G.W. Elko, ''A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield,'' in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 2002, Vol. II, pp. 1781–1784]. The spherical array tested has 24 omni-directional pressure microphones arranged on a rigid sphere. The sound pressure from each sensor is decomposed into spherical harmonics (standard eigen analysis) which results in rotationally symmetric output along any axis (three dimensionally) through the center of the sphere. An accurate impulse response of this system is important to sound localization and tracking methods. Time history can also be used to enhance the prediction model for source movement. The impulse measurements on this array were made in a semi-anechoic chamber and are in good agreement with expected results. The three main characteristics of the impulse response that will be discussed are the delay times, amplitude differences, and frequency response.

9:45

4aEA8. Development of microelectromechanical systems capacitive microphone for high-frequency applications. Michael Pedersen (Novusonic Corp., P.O. Box 183, Ashton, MD 20861, info@novusonic.com)

In this paper the development is described of a new high-frequency capacitive microphone based on the Knowles SiSonic MEMS microphone technology. The intended application for these microphones is aeroacoustic testing, which requires a transducer bandwidth of 80–100 kHz and the ability to measure sound pressures to 150 dB SPL without excessive distortion. The MEMS microphone structure was modified to meet these requirements and still be manufactured using the Knowles SiSonic production process. In addition, the microphone package was modified to eliminate the front port resonance, at approximately 12 kHz, associated with the standard SiSonic device. Specifically, the microphone package was modified with a lid that has lower acoustic inertance, thereby increasing the resonance frequency of the Helmholtz resonator formed by the front port and the microphone interior cavity. Details of the design process and the theoretical behavior of the microphones are presented, and it is shown that a microphone sensitivity of 0.5–1 mV/Pa can be achieved for a 100-kHz bandwidth. Preliminary results from fabricated microphones indicate a sensitivity around 0.5 mV/Pa and an A-weighted noise level of approximately 61.5-dB SPL. A more detailed analysis of the microphone noise spectrum and frequency response is presented.

10:00–10:15 Break

10:15

4aEA9. Multiphysics analysis of a loudspeaker. Marcellin Zahui (Univ. of North Dakota, Grand Forks, ND 58202)

A coupled physics analysis of a loudspeaker is presented. This finite-element analysis of a theoretical loudspeaker is performed using a commercial finite-element package installed on a Window-based personal

computer. The objective is to use multiphysics modeling to analyze the loudspeaker coupled fields (electric, magnetic, structural, fluid, and acoustic) that are solved in a single step. All the components of the electro-acoustic transducer, i.e., the rear plate, the magnet, the voice coil, the pole piece, the spider, the frame, the cone, and the dust cap are incorporated in a single finite-element model. The system is then analyzed using an electrical input circuit also modeled using finite element to evaluate the acoustic radiation from the cone. Even though the solution is obtained in a single analysis, each physics field is separately validated with known and well-established data. First the magnetic field is validated, then the resulting structural displacement of the cone, and finally the fluid field, followed by the acoustic field. The work is an attempt to provide a global design and optimization environment for electro-acoustic transducers and other acoustic multiphysics systems such as active noise control.

10:30

4aEA10. Acoustic impedance matching for axial wave radiation from cylindrical polyvinylidene fluoride film air ultrasonic transducer mounted inside of concentric hollow cylinder. Mionoru Toda (Measurement Specialties Inc., 460 E. Swedesford Rd., Ste. 3005, Wayne, PA)

It is well known that radial vibration of cylindrical polyvinylidene fluoride (PVDF) film excites acoustic waves towards radial direction in 360-deg range. When the length of cylindrical film is equal to or less than half the wavelength, the excited wave towards the axial direction becomes stronger than the radial wave. In this work, a cylindrical PVDF transducer was concentrically mounted inside of a hollow cylinder, and it was observed that the smaller the gap between film surface and inner wall surface of the pipe, the more the axial wave was excited. In order to explain this phenomenon, radiation impedance for axial wave excitation was calculated, and it was found that (1) radiation impedance increases with decreasing gap; (2) energy flow from transducer to axial wave increases with decreasing gap and becomes maximum at impedance matched condition; and (3) for the radiation impedance larger than optimum matched condition, the output decreases with increasing radiation impedance (decreasing gap) and the resonance frequency of cylindrical transducer shifts to the higher side due to loading of imaginary part of too high radiation impedance. All these results were successfully observed by experiments at 85-kHz burst mode. This principle can be applied to a new class of ultrasonic air transducer.

10:45

4aEA11. Microelectromechanical-systems-based parametric transmitting array in air—Application to the ultrasonic ranging transducer with high directionality. Haksue Lee and Wonkyu Moon (Dept. of Mech. Eng., POSTECH, Pohang, Kyungbuk 790-784, South Korea, wkmoon@postech.ac.kr)

Reduction of the transducer size decreases the efficiency of the secondary wave generation in the parametric transmitting array. To overcome this problem, MEMS-based transducer radiating high intensity sound is required in the parametric array. A high-performance PZT MEMS ultrasonic transducer composed of many small membrane-type radiator units was designed and fabricated by bulk micro-machining, and the secondary wave generated by the transducer was observed. This experiment proves that a MEMS-based parametric transmitting array is possible and applicable to some devices. A new ranging ultrasonic transducer with high directionality as an application was proposed because ultrasonic transducers generally need to be designed for high directionality to enhance the spatial resolution in the pulse-echo measurement in air. The 40-kHz difference tone, the measuring signal, was generated using 100- and 140-kHz primary tones by the PZT MEMS ultrasonic transducer. The difference tone had a half-power beam width of around 5 deg and a sound pressure level of over 85 dB at the peak. For the real application, higher sound pressure level should be achieved, and this will be included in the future work. [Research supported by MIC.]

11:00

4aEA12. A stepped plate transducer as an ultrasonic loudspeaker. Dong Hoon Yi, Hak-sue Lee, and Wonkyu Moon (POSTECH, San 31, Hyoja-dong, Nam Gu, Pohang, Gyungbuk, Korea, maneuvre@postech.ac.kr)

Research on the ultrasonic loudspeaker has been under way since Mary Beth Bennett and Blackstock verified the existence of parametric array phenomenon in air experimentally. Although several transducers that make use of piezoelectric films have been proposed, an ultrasonic loudspeaker system is not developed enough to be practical due to its high cost and high power consumption. In the 1970s, Gallego-Juarez *et al.* developed an ultrasonic transducer exploiting flexural modes of circular radiating plate, which shows highly efficient and directional performance. The feasibility of this type of transducer as ultrasonic loudspeakers, however, has not been investigated. Thus, a transducer with a stepped radiating plate of 283 mm diameter and a resonance at 40 kHz was designed, manufactured, and tested. In order to maximize the pressure radiated, piezoelectric ceramics were placed at the nodal point of the structure. The features of the transducer manufactured are over 130 dB of maximum primary wave at resonance on the acoustic axis, half-power beamwidth of primary ultrasound less than 10 deg, 40 to 60 dB of secondary waves from 1.5 to 20 kHz at 1m, and half-power beamwidth of secondary sound less than 10 deg.

11:15

4aEA13. Electrostatic negative stiffness and coupling factor of micromachined capacitive ultrasonic transducers. Li-Feng Ge (School of Electron. Sci. and Technol., Anhui Univ., Hefei 230039, China)

A theoretical model for micromachined capacitive or electrostatic ultrasonic transducers (mCUT or cMUT) was presented [L.-F. Ge, Theoretical and Computational Acoustic 2001, edited by E.-C. Shang *et al.*, *World Sci.*, 459–469, 2002; *Proc. 18th Intl. Conf. Acoustics, Vol. III*, 2255–2258, 2004]. The model takes into account the bending stiffness of diaphragm tension in diaphragm, compressibility of air within air-gaps, the effects of the metal electrode coated on diaphragm, and a negative stiffness caused by electrostatic attraction, providing a comprehensive and accurate description of the behavior of mCUTs. The electrostatic negative stiffness is derived from the electrostatic force based on the parallel plate capacitor assumption and is expressed as $K_{neg} = -\varepsilon_0 V_0^2 / x_0^3$, where ε_0 is the permittivity of air, V_0 the bias voltage, and x_0 the effective separation between two electrodes. It corresponds to a negative capacitance C_{neg} or negative impedance $\mathbf{Z_{neg}}$ at the mechanical port in the equivalent electrical circuit for mCUTs. Thus, the electromechanical coupling factor $\mathbf{K}$ can be defined by $\mathbf{K}^2 = (-\mathbf{Z_{neg}})/\mathbf{Z_{MO}}$, which is a ratio of the electrostatic negative impedance to the open-circuit mechanical impedance of the mCUT. The definition may give a new view for the coupling factor of electromechanical transducers. [Work supported by NSFC (60374044).]

Session 4aEDa

Education in Acoustics: Medical Education

Daniel R. Raichel, Chair
Eilar Associates, 2727 Moore Ln., Fort Collins, CO 80526

Chair's Introduction—8:00

Invited Papers

8:05

4aEDa1. Ultrasound imaging instruction for veterinary students at Colorado State University. Robert H. Wrigley (Veterinary Medical Ctr., Colorado State Univ., Fort Collins, CO 80523)

Didactic ultrasonography instruction to the veterinary students begins with the physical principles of sonography, machine selection, and operating controls. Subsequently, image artifacts and normal sonograms of the heart and abdominal organs are presented. Sonograms of pregnancy and abnormal sonograms of common diseases are discussed. In the final-year practicum course, small groups of students (2–4) spend a week in the echocardiography lab and the soft-tissue imaging lab. Here, students acquire hands-on sonography skills and become proficient in sonography of small animals and horses. An Internet-based interactive instruction module on the equine limb was developed to enable veterinary students to better learn ultrasonography of tendons/ligaments of the metacarpus. Interactivity was achieved by creating clickable zones on the tendons and ligaments. Students are challenged to apply the knowledge gained studying normal ultrasound anatomy to clinical patients affected with tendon/ligament injuries. Students are asked to identify (by clicking on) the location of the sonographic lesions. The program provides immediate feedback on their ability to recognize the site of the injury. In addition, having the program web-based allows students to review the instructional material on a time schedule that best fits their needs and from remote locations.

8:35

4aEDa2. Education in medical acoustics. Tyrone Porter (Dept. of Biomed. Eng., Univ. of Cincinnati, 231 Albert Sabin Way, 6166 Medical Sci. Bldg., Cincinnati, OH 45267-0586)

The field of medical acoustics has grown tremendously in the last 30 years. Sophisticated ultrasound scanners are utilized to image a developing fetus in the womb or blood flow in the vessels. High-pressure acoustic pulses travel through muscle, fat, and capillary beds to pulverize kidney stones with minimal damage to the intervening tissue. Gas-based microbubbles stabilized with polymeric or liposomal shells are injected into the vasculature to enhance the contrast in ultrasound images or to deliver life-saving drugs. Due to the interdisciplinary nature of medical acoustics, educating the next generation of clinicians, engineers, and researchers within the field has become more challenging. Students must gain some understanding of a variety of concepts, including the derivation of the wave equation, the effect of biological media on the propagation of ultrasound (attenuation, scatter, etc.), signal-processing techniques for converting scattered sound into images, and the effect of viscoelastic shells on the forced oscillation of microbubbles. Armed with this knowledge, medical acousticians will continue to develop innovative diagnostic and therapeutic applications for ultrasound.

9:05

4aEDa3. Teaching ultrasound instrumentation in medical and dental schools. Maureen Stone (Dept. of Biomed. Sci. and Dept. of Orthodontics, Univ. of Maryland Dental School, 666 W. Baltimore St., Baltimore, MD 21201, mstone@umaryland.edu)

This presentation will describe teaching strategies for applying ultrasound, a commonly used instrument in medical schools, to studying a novel application, the tongue. These methods have been used with students in otolaryngology and dentistry as well as linguistics, engineering, and speech and hearing. Four areas are considered important for students to know. (1) Introduction to ultrasound principles and use of commercial machines. Students are taught basic principles of sound propagation including focused versus nondirectional sound waves, scan types, and tissue properties. Commercial machine features are explained, such as gain control, depth setting, and calibration. (2) Data collection for research versus clinical purposes. Students are taught procedural requirements for research studies, including the importance of choice of task, positioning of transducer, and head stabilization to allow normalization across subjects. (3) Reduction and analysis of the high-dimensional ultrasound images. Data conditioning includes understanding how to interpret images, identify structures of interest, determine artifacts, find and use references, and extract tongue surface contours from the images. (4) Interpretation of data. It is easy to see tongue differences, but harder to capture them numerically. Students learn tools to reduce contour dimensionality, average and compare contours and contour sequences. [Research supported by R01 DC01758 NIDCD.]

9:35

4aEDa4. Use of ultrasound as educational tool in medicine. E. Carr Everbach (Eng. Dept., Swarthmore College, Swarthmore, PA 19081)

Ultrasound is an imaging and therapy technology that is becoming increasingly important in modern medicine. Though only a fraction of physicians nationwide uses ultrasound in their practices, most medical students and many technicians receive an overview of the principles of ultrasound as part of their training. Often this overview is the only introduction to acoustics, and can be used as an educational tool to raise broader issues. This talk will describe the basic education in biomedical ultrasound obtained by most physicians and healthcare workers, and will discuss what is left out that should be part of medical education.

10:05–10:20 Break

10:20

4aEDa5. Teaching acoustics in the medical and related fields. Daniel R. Raichel (CUNY Grad. Ctr. and Eilar Assoc., 2727 Moore Ln., Fort Collins, CO 80526)

Acoustics is playing an ever-increasing role in the medical and related fields. Besides the more obvious applications such as otology, audiometry, stethoscope usage, effects of sound on hearing, and psychological well-being, more modern applications include ultrasound diagnostics and therapeutics, fetal monitoring, surgical applications, removal of dental plaque, musical therapy, etc. This gives rise to the question: how much acoustical theory needs to be taught to future practitioners, many of whom were not science or engineering majors during their undergraduate years? A motor vehicle operator, for example, does not need to understand the fundamentals of internal combustion and mechanics of power transmission in order to develop the skill to handle a vehicle; he/she needs to learn applicable motor vehicle laws. In the same vein, a prospective medical practitioner does not need to learn solutions to the wave equation and understand the basis of how frequency analyzers function, but he/she needs to develop the skills to use applicable acoustic technology, interpret the results with understanding of their ramifications, and be aware of technical limitations such as permissible ultrasound dosages. Acoustic education requirements in a number of medical subfields are discussed in terms of depth and breadth of coverage.

THURSDAY MORNING, 8 JUNE 2006 ROOM 553A, 11:00 A.M. TO 12:00 NOON

Session 4aEDb

Education in Acoustics: Topics in Education in Acoustics

Courtney B. Burroughs, Chair
1241 Smithfield Street, State College, PA 16801

Contributed Papers

11:00

4aEDb1. Acoustic education and the Bolognan process in Russia. Nikolai Zagrai and Vladislav Zakharevich (Taganrog State Univ. of Radio-Eng., GSP-17A, 44 Nekrasovsky St., Taganrog, 347928, Russia)

Acoustic education in Russia is traditionally based on a solid natural-sciences curriculum with core classes in higher mathematics, physics, continuum mechanics, and electronics. However, parallel to the so-called specialist qualification encompassing a standard 5-years engineering training, a new education trajectory of bachelor (4 years) and master (6 years) qualifications has emerged as a result of the international Bolognan Agreement recently signed by the Russian Federation. Students willing to pursue the master qualification may enroll in this program after completing either bachelor or specialist training. Aimed at making Russian educational programs more easily understood, we are implementing a smooth transition from traditional academic hours used for weighting program content and value in the Russian educational system to internationally recognized credit hours. Nowadays, mainstream technical universities include individual creative thinking approach in their modulus-rating systems with the multilevel assessment. The latter is related to direct rating educational technologies. In general, the content of Russian education programs in acoustic is equivalent to similar foreign programs.

11:15

4aEDb2. Optical table with embedded active vibration dampers (smart table) as an educational tool in structural acoustics and vibration. Vyacheslav M. Ryaboy (Newport Corp., 1791 Deere Ave., Irvine, CA 92606, vyacheslav.ryaboy@newport.com)

The actively damped optical table contains two sensor-actuator assemblies connected to a digital feedback controller. The controller is programed to damp resonant flexural vibrations of the table. The controller can auto-tune to accommodate changes in the payload. It has a software interface that allows one to switch the damping on and off, to regulate the amount of active damping manually, to analyze the signals from embedded vibration sensors, and to display the results in both time and frequency domains. For an advanced user it is possible to completely reprogram the controller. Besides its main purpose of vibration control, the smart table can be used for laboratory demonstrations of various phenomena in structural acoustics without additional equipment or with minimal additional technical means. Among these effects: structural response in the frequency domain (resonance peaks versus forced operational peaks, damped versus undamped system); time response of damped versus undamped system; tonal sound emission in response to random excitation (damped versus

undamped system); open-loop response of the control system and mechanical impedance of the structure versus closed-loop response. More experiments and demonstrations can be devised as educators and students become more familiar with the system.

11:30

4aEDb3. Development of an acoustics program at Morgan State University. M. Roman Serbyn (Dept. of Phys. and Eng. Phys., Morgan State Univ., 1700 E. Cold Spring Ln., Baltimore, MD 21251, rserbyn@jewel.morgan.edu)

This report describes the gradual implementation of one educator's plan to introduce acoustics into the curriculum of a medium-size university. Rooted in the physics department, the program from its inception has emphasized the interdisciplinary nature of the discipline. The first course, consisting of a series of lectures within an existing format of ''Great Issues,'' was offered to seniors of the university in 1998. Subsequently one graduate and two undergraduate courses were developed and approved. The undergraduate course, Acoustics and You, has been particularly popular because of the diversity of topics. The graduate course is a typical presentation of the fundamentals of acoustics with discussion of selected applications. Acoustical research has been limited to the author's areas of interest and subject to time and funding constraints. Recent collaboration with the departments of biology and psychology have been very beneficial to the growth of the program. The narrative will be illustrated with course syllabi, student data, as well as lessons learned, and plans for the future.

11:45

4aEDb4. A Portable device to illustrate two methods for vibration reduction namely tuned dynamic absorber and isolation mounting. Samuel Koretas and M. G. Prasad (Dept. of Mech. Eng., Stevens Inst. of Technol., Hoboken, NJ 07030, mprasad@stevens.edu)

Mechanical vibrations can be harmful to structures as well as to humans. It is important to understand the principles of design that are used in developing various methods for reducing vibration. There are several tools at our disposal to reduce the vibrations transmitted from a vibration machine to the building structure. Two commonly available methods of vibration control are using passive vibration isolation mountings and tuned mass absorbers. This paper presents the design and working of a portable device to illustrate both methods of vibration reduction. We can examine the performance of both mounting and absorber using this physical device. The analytical and experimental work shows that such a device can effectively illustrate the principles involved in the design of mountings and absorbers. Also, the device can be used to show the influence of various physical parameters on performance. The presentation will also include the discussion on the role of such devices in enhancing the understanding of design principles in acoustics, noise, and vibration.

THURSDAY MORNING, 8 JUNE 2006 BALLROOM E, 9:00 TO 10:45 A.M.

Session 4aMU

Muscial Acoustics: Topics in Musical Acoustics

Peter J. Hoekje, Chair
Baldwin-Wallace College, Dept. Physics and Astronomy, 275 Eastland Rd., Berea, OH 44017

Contributed Papers

9:00

4aMU1. Real-time measurement/viewing of vocal-tract influence during wind instrument performance. Gary P. Scavone (Music Technol., Schulich School of Music, McGill Univ., 555 Sherbrooke St. West, Montreal, QC, H3A 1E3 Canada)

Since the early 1980s, there have been a number of investigations into the role and influence of a player's vocal tract on the sound production of wind instruments. While the underlying acoustic principles are relatively well understood, a general lack of agreement remains within both the music acoustics and performance communities with regard to the importance of this mechanism during playing conditions. Disparities arise in part because subtle manipulations of the oral cavity can affect the response of the instrument without necessarily producing effects audible to an observer. A real-time measurement system is demonstrated that provides a visual comparison of the relative strengths of the ''upstream'' windway and ''downstream'' air column impedances under playing conditions. The system assumes continuity of volume flow on either side of the ''reed,'' which leads to a direct proportionality between the upstream and downstream pressures and impedances. Playing experiments clearly demonstrate many instances in which vocal-tract manipulations can cause impedance peak magnitudes in the mouth cavity to exceed those in the downstream air column. In addition to providing visual ''proof'' of such manipulations, the system is expected to offer a pedagogical tool to help performers better learn how to make use of this control mechanism.

9:15

4aMU2. Amplitude fluctuations within musical instrument sounds because of impulse reflections. Rolf Bader (Inst. of Musicology, Univ. of Hamburg, Neue Rabenstrasse 13, 20354 Hamburg, R_Bader@t-online.de)

Musical instrument sounds nearly always show amplitude fluctuations in their overtones within the quasisteady state of the sound.These fluctuations make the sound more lively and are found in blown instruments like the saxophone or clarinet, in string instruments like the guitar or the violin, or in percussion instruments. The reason for this can be found within a feature common to all musical instruments, their impulse-driven behavior. Not only do the blown instruments work with impulses; the string forces of plucked, strucked, or bowed instruments of the top plate are also an impulse train. So, when looking at the system during the steady state not like one of standing waves but of impulses being reflected at different inhomogeneties within the structure, these reflections can be shown to be the reason for the amplitude fluctuations. The mechanism is caused by the fact that the energy within an impulse is split into potential and kinetic energy, where only the kinetic energy is radiated. When traveling along a

homogeneous region, the energies are more or less conserved. But, during reflections, these energies get larger and smaller and cause a fluctuation in the kinetic energy and so in the radiation of sound.

9:30

4aMU3. Acoustical effects of reed cells in the American reed organ. Benjamin A. Franta and James P. Cottingham (Phys. Dept., Coe College, 1220 First Ave. NE, Cedar Rapids, IA 52402)

In a reed organ the free reed is mounted above the windchest and is typically surrounded by a reed cell cavity. The length of this tubelike cavity varies from just slightly longer than the reed in some instruments to several centimeters longer in others. Earlier research on effects of the reed cell focused mainly on the lowering of sounding frequency due to the presence of the cell [J. Cottingham, ''Effects of reed cell geometry on the vibration frequency and spectrum of a free reed,'' J. Acoust. Soc. Am. **101**, 3143 (1997)]. Reed cells are shown to have a significant effect on the radiated waveform and spectrum. The length and interior geometry of the cell are significant factors in determining this effect. Changes in reed cell dimensions have a greater effect on the sound spectrum than changes in reed dimensions. It is shown that the presence of a reed cell considerably enhances the upper harmonics of the spectrum. Higher vibrational modes in air-driven free reeds have recently been observed [A. Paquette, J. Vines, and J. Cottingham, ''Modes of vibration of air-driven free reeds in transient and steady state oscillation,'' J. Acoust. Soc. Am. **114**, 2348 (2003)]. The current measurements indicate that the reed cell substantially affects the higher vibrational modes of the reed, particularly enhancing the second and third transverse modes. [Research supported by the Carver Summer Scholar Program.]

9:45

4aMU4. Experimental and calculated frequencies of the Southeast Asian naw. Eric A. Dieckman (Truman State Univ., Kirksville, MO 63501, ead603@truman.edu) and James P. Cottingham (Coe College, Cedar Rapids, IA 52402)

The naw, an Asian free-reed gourd pipe, is constructed of pipes open at both ends with a free reed in the side of each pipe near one end. The pipes of the naw are cylindrical over a significant portion of their length but have a slightly tapered bore near the reed end. The bottom end of the pipe can be covered with the player's thumb to bend the pitch. Previous research on the khaen, a similar instrument with a cylindrical bore, found that the sounding frequencies of the khaen are well approximated using the simple equation for an open cylindrical pipe. This is not possible for the naw. A method of transmission matrices has been used to calculate the theoretical resonant frequencies of the naw pipes, both open and closed. The calculated values agree well with the resonant frequencies determined experimentally by measuring the input impedance of the pipe. When blown normally, the open pipe sounds near its calculated frequency, while the closed pipe sounds slightly lower than the open pipe frequency. By underblowing the reed, however, the closed pipe can be made to sound near its predicted resonant frequency. [Research supported by NSF REU Grant PHY-0354058.]

10:00–10:15 Break

10:15

4aMU5. Benchmarking the lattice Boltzmann method for the determination of acoustic impedances of axisymmetric waveguides. Andrey Ricardo da Silva, Philippe Depalle, and Gary Scavone (Music Technol. Area, Faculty of Music, McGill Univ., 555 Sherbrooke St. West, Montreal, Canada H3A 1E3, andrey.dasilva@mail.mcgill.ca)

A numerical technique for determining the acoustic impedance of axisymmetric waveguides based on the lattice Boltzmann method (LBM) is proposed here. This approach presents some desirable characteristics, namely, the possibility of explicitly considering the propagation of waves at low Mach number conditions with a relative low computational cost when compared to traditional continuum methods. The validation of the method is achieved by simulating the radiation of an open unflanged pipe using LBM and comparing the results with those obtained through a finite-element model and through the analytic solution derived by Levine and Schwinger. Finally, the results and the discussion of the method in terms of its computational efficiency and applicability to the simulation of wind instruments at more realistic conditions are presented. [The first author was supported by CAPES (Brazil).]

10:30

4aMU6. Keyboard characterization using spiral and Cartesian mappings of the harmonic series. Palmyra Catravas (ECE Dept., Union College, Schenectady, NY 12308, catravap@union.edu)

Spatial transformations of functions of frequency can facilitate the analysis of frequency spectra and the identification of characteristic spectral patterns. Here, spiral and Cartesian mappings of the harmonic series are applied to frequency spectra from several keyboard instruments (modern grand piano, modern upright piano, fortepiano). Their utility in comparing properties of the instruments, including inharmonicity, coupling between strings, tuning errors, and beat frequencies, and simply identifying a fundamental frequency will be described. An update on a previous application of the transformations (identification of the B-A-C-H motif in music) will also be presented.

Session 4aNS

Noise: Topics in Noise Control

Nancy S. Timmerman, Chair
Nancy S. Timmerman, P.E., 25 Upton St., Boston, MA 02118-1609

Contributed Papers

8:45

4aNS1. Development of artificial neural network classifier to identify military impulse noise. Brian A. Bucci and Jeffrey S. Vipperman (Dept. of Mech. Eng., Univ. of Pittsburgh, 648 Benedum Hall, Pittsburgh, PA 15261)

Civilian noise complaints from urban encroachment have curtailed military exercises that produce high levels of impulse noise. Currently, monitoring stations are in place around many military installations to monitor noise levels. However, these monitoring stations do not provide real-time data and suffer from false positive detections. This project sought to develop algorithms with improved noise classification accuracy. Various types of military impulse noise and nonimpulse noise were measured and processed. Approximately 1000 waveforms were field collected (670 nonimpulse and 330 military impulse). Although several metrics were examined, the conventional time domain metrics of kurtosis and crest factor as well as two novel frequency domain metrics, weighted square error and spectral slope, were found to work best. The computed metric values were used to train and verify the performance of the artificial neural network (ANN), which was able to achieve 100% accuracy on training data and 100% accuracy on verification data. [This research was supported wholly by the U.S. Department of Defense, through the Strategic Environmental Research and Development Program (SERDP).]

9:00

4aNS2. Analysis of high-amplitude jet noise using nonlinearity indicators. Kent L. Gee, Micah R. Shepherd (Dept. of Phys. and Astron., N283 ESC, Brigham Young Univ., Provo, UT 84602, kentgee@byu.edu), Lauren E. Falco, Anthony A. Atchley (The Penn State Univ., University Park, PA, 16802), Lawrence S. Ukeiley, and Bernard J. Jansen (The Univ. of Mississippi, University, MS 38677)

Nonlinear effects in high-amplitude jet noise are often characterized by examining spectral changes over some propagation range. However, in the case of typical laboratory-based measurements with model-scale nozzles, the study of nonlinearity via spectral changes is made difficult by both limited measurement bandwidth and short propagation distances. One analysis technique that potentially can overcome these difficulties is a single-point measurement of nonlinearity that has been studied most recently by Falco *et al.* [J. Acoust. Soc. Am. **117**, 2596 (2005)]. Another potential solution is bispectral analysis, which has been used in the past as an indicator of nonlinearity for a wide variety of applications. Both of these analysis methods have been applied to data from recent laboratory measurements at the National Center for Physical Acoustics. Observed trends for a variety of source conditions, measurement angles, and propagation distances are discussed. [Work supported in part by the Office of Naval Research.]

9:15

4aNS3. A comparison of differences in highway noise barrier design using TNM 2.5 and STAMINA. Ning Shu, Louis Cohn, Rowell Harris, and Teak Kim (Univ. of Louisville, Louisville, KY 40292)

In April 2004, the FHWA TNM 2.5 was released; all state transportation agencies are required to use it for all Federal funded projects by FHWA. TNM 2.5 has been modified from the previous versions to update the noise energy back to the source, and therefore influence the predicted insertion loss and absolute noise levels. The study was designed to compare the noise barrier design by both TNM 2.5 and TNM 2.0, as well as STAMINA 2.0. All these models were used to design noise barriers on the I-64 St. Louis, MO project with the length for the mainline of 11 mile.

9:30

4aNS4. Analysis of brake squeal noise using the finite-element method: A parametric study. Samir N. Y. Gerges, Mrio Trichs, Jr., and Alessandro M. Balvedi (Dept. of Mech. Eng., Federal Univ. of Santa Catarina, Trindade, Florianopolis, SC, Brazil)

Brake squeal noise has been under investigation by the automotive manufacturers for decades due to consistent customer complaints and high warranty costs. Brake squeal is a problem that has been faced by several numerical and analytical methods in the last decade. Among then, finite-element analysis is proving to be a useful tool to predict the noise occurrence of a particular brake system during the design stage. The finite-element method (FEM) has been employed to analyze the squeal propensity of the disc brake systems. The complex eigenvalue analysis conjugated with FEM has been widely used to detect unstable frequencies in numerical brake systems models. These unstable frequencies are related, in most cases, with high sound pressure levels generated during the brake system operation. This paper summarizes the application of the complex eigenvalue analysis in a finite element model. The effects of the operating parameters (friction coefficient, braking pressure, brake temperature) and wear on the dynamic stability of the brake system are examined. After identifying unstable frequencies and the behavior under different conditions, the performance of some control methods are tested. Changes in material properties and the application of brake noise insulators are presented and its effect discussed.

9:45

4aNS5. NOIDESc: A novel traffic noise description scheme. Brian Gygi, Werner Deutsch, Holger Waubke, and Anton Noll (Acoust. Res. Inst. of the Austrian Acad. of Sci., Reichratsstrasse 17, Vienna 1010 Austria)

A novel framework for the acoustic and psychoacoustic description of noise signals (railway and road) has been proposed in order to extend and improve traditional noise classification schemes. In addition to measures of averaged sound-pressure level and noise exposure duration, the description scheme includes psychoacoustic parameters such as spectral centroid, spectral spread, modulation spectrum, and further dynamic aspects of the sound events corresponding in part to the MPEG-7/4 standard. The dynamic changes between foreground and background are addressed in the model. The approach assumes the evaluation of several low-level acoustic parameters of the noise, to be integrated into few high-level features by means of statistical methods, such as Gaussian mixture models and cluster analysis. As the usage of noise monitoring systems is expected to increase in the near future, the cumulative collection of calibrated sound recordings will provide comprehensive regional and supraregional sound databases to serve as input for automated noise classification. [The work in progress is performed in cooperation with the division of Acoustics of the TGM Vienna and is supported by the Austrian FFG.]

10:00

4aNS6. New approach to studying traffic noise nuisance. Cobis Rabah Derbal and Hamza Zeghlache (BP 12 SMK, Constantine 25003, Algeria, cobis.derbal@free.fr)

More often the problem of noise and sonic pollution particularly in urban sites, is studied by different disciplines such as physics, acoustics, psychoacoustics, medicine, and others. Some studies are carried out in laboratories taking noise and sonic phenomena as samples cut off from their realistic context. Urban noise is studied as well by making abstraction of the different contextual parameters by idealizing a rather complex sonic environment. This derives often on incompatible results that can hardly be used in real situations of urban noise problems. The noise, according to this present approach, is supposed to react with the surrounding space, and it takes the form and the quality of the place by defining and requalifying it. It is found that the contextual aspects such as social, cultural, or even symbolic dimensions modulate the listening and the perception quality of the noise. The multiparameter dimension study of the noise in urban context is necessary. The small number of studies already done based on such multidisciplinary approach and the results that came out of such studies comfort well our effort and are very encouraging to go ahead with this methodological approach.

10:15–10:30 Break

10:30

4aNS7. Predicting acoustic and seismic pulses from outdoor explosions. Keith Attenborough, Patrice Boulanger (Dept. of Eng., The Univ. of Hull, Cottingham Rd., Hull HU6 7RX, UK), and Donald Albert (U.S. Army ERDC Cold Regions Res. and Eng. Lab., Hanover NH 03755-1290)

To support the formulation of a simple model that can be used for assessment and prediction of ground vibration levels due to above-ground explosions, a fast field program for layered air ground systems (FFLAGS), developed originally for continuous sound sources, has been extended to enable predictions of propagation from impulsive sound sources in a refracting atmosphere above layered porous and elastic ground. Using a mixture of measured and best-fit parameters, the pulse version of the code (PFFLAGS) has been found to give predictions of the air pressure spectrum above ground and the vertical component of solid particle velocity near the ground surface that compare tolerably well with published data for the spectra and waveforms of acoustic and seismic pulses at short range (60 m) in grass- and snow-covered ground and at longer range (3 km) in forested terrain. The predicted seismic pulses in the presence of the snow cover explain a phenomenon frequently observed for shallow snow covers whereby a high-frequency ground vibration is modulated by a lower-frequency layer resonance [Work supported in part by SERDP SEED Project SI-1410 and Contract Number W90C2K9887EN010001 through the European Research Office of the U.S. Army.]

10:45

4aNS8. Noise in the adult emergency department of Johns Hopkins Hospital. Douglas Orellana, Ilene J. Busch-Vishniac, and James E. West (Mech. Engr. Dept., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

While hospitals are generally noisy environments, nowhere is the pandemonium greater than in an emergency department, where there is constant flow of patients, doctors, nurses, and moving equipment. In this noise study we collected 24-h measurements throughout the Adult Emergency Department of Johns Hopkins Hospital, the top-ranked hospital in the U.S. for 15 years running. The emergency department is structured into pods, each of which specializes in particular types of treatment. We obtained measurements in each pod. Compared to data collected at other locations of the hospital, the average L_{eq} throughout the emergency department is about 5 dB (A) higher. Within the emergency department the triage area has the highest L_{eq}, ranging from 65 to 73 dB (A), primarily due to being at the entrance to the department. Sound levels in the emergency department are sufficiently high [on average between 62 to 67 dB(A)] to raise concerns regarding the communication of speech without errors, an important issue everywhere in a hospital and a critical issue in emergency departments because doctors and nurses frequently need to work at an urgent pace and to rely on oral communication.

11:00

4aNS9. Reducing the occlusion effect with active noise control. James G. Ryan (Gennum Corp., 232 Herzberg Rd., Kanata, ON, Canada K2K 2A1), Betty Rule, and Stephen W Armstrong (Gennum Corp., Burlington, ON, Canada L7L 5P5)

The occlusion effect refers to the build-up of low-frequency sound pressure in the ear canal when it is sealed by a hearing aid. Hearing aid wearers experiencing the occlusion effect often complain about the sound of their own voice. The traditional solution to this problem is to introduce a vent or opening through the hearing aid from the canal end to the faceplate. Unfortunately, the use of large vents, which is required for reducing occlusion, can result in significant feedback problems. This paper describes a solution for the occlusion problem that is based on active noise control techniques. This approach allows occlusion reduction without the feedback problems caused by large vents.

11:15

4aNS10. Operating room noise at Johns Hopkins Hospital. Jonathan Kracht, Ilene J. Busch-Vishniac, and James E. West (Mech. Engr. Dept., Johns Hopkins Univ., 3400 N. Charles St., Baltimore, MD 21218)

Very little reliable information exists on the sound levels present in an operating room environment. In an attempt to remedy this situation, rigorous acoustic testing of the operating rooms in Johns Hopkins Hospital was conducted. Sound levels in each of the operating rooms was recorded for a period of 24 h. The data were analyzed to determine background sound levels, average equivalent sound levels (L_{eq}'s), and frequency distribution. With the assistance of Johns Hopkins Hospital, we were able to match each surgery to the period of noise it produced and to associate sound levels with particular types of surgeries. This data was used to determine the average sound-pressure level for classes of surgery (e.g., orthopedic, neurological, etc.). Averaging over many surgeries, orthopedic surgery was found to have the highest L_{eq} at approximately 66 dB(A). Urology, cardiology, gastrointestinal, and neurosurgery followed closely, ranging from 62 to 65 dB(A).

11:30

4aNS11. Dynamic behavior analysis of a switched reluctance motor. Juan Llado, Beatriz Sanchez, Jesus Ortiz (Mech. Eng. Dept., Univ. of Zaragoza, Spain), Aberlardo Martinez, Francisco Perez, Bonifacio Martin del Rio, Tomas Pollan, Eduardo Laloya (Univ. of Zaragoza, Spain), and Javier Vicuna (Univ. of la Rioja, Spain)

Within the home appliances sector, the series motor practically shares most applications. However, no further progress in cost reduction or control performances is envisaged. But, the switched reluctance motor, SRM, shows fairly good characteristics such as robustness, easy assembly, and low manufacturing costs. Besides, it allows accomplishment of integrated designs with the product, and an improvement of efficiency up to 90% when compared to about 45% reached by series motors used for home appliances. However, two specific drawbacks of this machine are control complexity and acoustic noise. As a first step, with respect to the acoustic noise, the target of this work is to develop a theoretical modal model of an SRM of four phases with 24 poles in an external stator and 18 poles rotor placed inside. This requires a theoretical simulation by means of an FEM software and an experimental work in a test bench to analyze the motor structure. The comparison of the results obtained by both methods will validate the theoretical model analyzed with the FEM software, and introduce structural modifications that can be used to design quiet SRMs from the knowledge of mode shapes and mode frequencies.

11:45

4aNS12. Sound power emission of a washing-machine cabinet. Beatriz Sanchez and Juan Llado (Mech. Eng. Dept., Univ. of Zaragoza, Spain)

The noise emitted by a washing machine can be divided into airborne noise produced mainly by the electric motor, and structural-borne noise caused by the cabinet vibration. The aim of this work is to determine the contribution of the structural radiation of the cabinet to the total sound power emitted by a washing machine, and to know which element is the noisiest. The total sound power was measured according to ISO 3744, and in order to know which panel of the cabinet is the noisiest, the standard ISO TR 7849 was applied. This experimental procedure relates the noise radiated by a structure with its vibration velocity, and it was found that the front and lateral left panels were the most relevant. Once the noise transmission paths were characterized, the proposal to reduce noise was focused on reducing the excitation force transmitted by the shock absorbers. As the modification of the force applied by the shock absorber is initially not feasible because it would imply changing the dynamic of the oscillating group, as a first step the panels which support these elements were modified to make them more flexible. A sound power reduction of about 5 dBA was obtained.

THURSDAY MORNING, 8 JUNE 2006 ROOM 550AB, 8:00 TO 11:45 A.M.

Session 4aPA

Physical Acoustics: Landmine Detection; Scattering Phenomena

James M. Sabatier, Cochair
Univ. of Mississippi, National Ctr. for Physical Acoustics, Coliseum Dr., University, MS 38677

Alexander Ekimov, Cochair
Univ. of Mississippi, National Ctr. for Physical Acoustics, Coliseum Dr., University, MS 38677

Contributed Papers

8:00

4aPA1. Excitation techniques for nonlinear acoustic landmine detection. Douglas Fenneman and Brad Libbey (U.S. Army RDECOM CERDEC NVESD, 10221 Burbeck Rd., Fort Belvoir, VA 22060)

Nonlinear acoustic landmine detection techniques are currently being investigated. These detection methods employ two tones broadcast in the vicinity of a buried mine. Measurement of the resultant surface velocity yields a spectrum rich in nonlinearly generated combination frequencies. The nonlinear mechanism governing the mine/soil interface is not completely understood. Donskoy *et al.* have proposed that a discontinuity occurs when the mine/soil system is subjected to acoustic excitation. [Donskay *etal*, J. Acoust. Soc. Am. **111**, 2705–2714 (2002)] In this case, the mine and soil remain in contact during compression, but separate in tension. Concurrently, Korman and Sabatier have documented nonclassical nonlinear effects in conjunction with acoustic mine detection [Korman *etal*, J. Acoust. Soc. Am. **116**, 3354–3369 (2004)]. Specifically, they have observed conditioning, soil dynamics, and hysteresis. These results are consistent with those of geomaterials which exhibit mesoscopic elasticity. It is likely that the nonlinear response above a buried mine is a function of both classical and nonclassical effects. These experiments investigate the amplitude dependence of nonlinear effects above a buried mine as a function of linear velocity. A comparison of the nonlinear amplitude dependence in the presence of a mine versus soil alone is intended to provide insights into the nature of the mine/soil interface.

8:15

4aPA2. Laboratory experiments on nonlinear acoustic detection of buried objects. Keith Attenborough, Qin Qin (Dept. of Eng., The Univ. of Hull, Cottingham Rd., Hull HU6 7RX, UK), Jonathan Jefferis, and Gary Heald (DSTL, Kent, TN14 7BP, UK)

The surface velocity response to acoustic excitation of sand contained within a thick-walled PVC tank in the shape of a truncated cone has been investigated with and without buried objects. The excitation was provided by either tonal or white-noise spectra from a large loudspeaker, and the sound field was localized above the buried objects by means of an inverted cone and pipe arrangement. The surface velocity of the sand has been measured with an accelerometer. The loading effects of the accelerometer have been found to be minor. The surface admittance spectra above a buried plastic container are found to have uniquely resonant structures, the nature of which has been investigated. It has been found that if the surface of a granular medium, containing buried objects, is subject to high levels of sound using either a single frequency or two separate frequencies, then, as well as an enhanced linear surface velocity response at the excitation frequencies, increased quadratic and cubic nonlinear responses may be observed. The nonlinear response at a sand surface is found to be highly sensitive to the moisture content and the degree of compaction. [Work supported by DSTL (UK) Contract Number RD029 06495.]

8:30

4aPA3. Saturation of burial-induced resonance frequency shift in acoustic landmine detection experiments. W. C. Kirkpatrick Alberts, II, Roger Waxler, and James M. Sabatier (Univ. of Mississippi, NCPA, 1 Coliseum Dr., University, MS 38677)

Modes in landmines have become a topic of recent interest in acoustic landmine detection. Of equal importance is the effect of burial on the experimentally observed resonances of landmines. Complicated dependence of resonance frequency on depth has been observed in field measurements and has been, thus, the topic of experimental and theoretical study. As such, further burial simulation experiments have been performed on a landmine and an acoustic landmine simulant. Burial simulation in water shows a saturation of downward modal frequency shift that rapidly deviates from the shift predicted by simple mass loading theory. This saturation agrees with that predicted by Velea *et al.* [J. Acoust. Soc. Am. **115**, 1993–2002 (2004)]. Burial simulations performed in sand reveal resonance frequency depth dependence that saturates following a rapid downward frequency shift and a small resonance frequency increase. [This work was sponsored by US Army Research, Development, and Engineering Command, Night Vision and Electronic Sensors Directorate under Contract DAAB15-02-C-0024.]

8:45

4aPA4. Wide-bandwidth time-reversal techniques for landmine detection. Brad Libbey, Douglas Fenneman (U.S. Army RDECOM CERDEC NVESD, 10221 Burbeck Rd., Fort Belvoir, VA 22060, brad.libbey@nvl.army.mil), Alexander Sutin, Victor Kurtenoks, and Armen Sarvazyan (Artann Labs., Inc., Trenton, NJ 08618)

Detection of landmines using acoustic to seismic wave propagation has shown that accurate landmine localization is possible by sensing vibrations of a soil surface above the mine. Time-reversal acoustic (TRA) focusing allows concentration of elastic energy at a location in the soil that increases landmine vibration and improves signal/noise levels. Previous experiments with TRA landmine detection utilized narrow-band signals, but in the most recent experiments wide-bandwidth signals provide clear excitation of the mine resonant frequencies. A loudspeaker array broadcasts these signals and the resulting soil vibration is measured by a laser Doppler vibrometer. The experiment utilizes linear frequency sweeps and orthogonal signals with frequency content between 50 and 500 Hz. Using the orthogonal signals, the suggested TRA procedure can be implemented simultaneously with multiple transmitters to increase the scanning speed. The spatial distribution of energy on and off mine demonstrates that the TRA system concentrates elastic wave energy in the desired area with the focal spot dimension comparable to a wavelength. The spectral density of the TRA focused signal increases in the area near the resonant frequency of the landmine.

9:00

4aPA5. Performance of the scanning ultrasound vibrometer for sensing displacements of a randomly rough vibrating surface. Ameya Galinde, Mark Andrews, Ninos Donabed, Carey Rappaport (Northeastern Univ., 409 Dana Res. Ctr., Boston, MA 02115-5000, ameya.g@gmail.com), Douglas Fenneman (U.S. Army RDECOM CERDEC Night Vision and Electron. Sensors Directorate; Fort Belvoir, VA 22060), and Purnima Ratilal (Northeastern Univ., Boston, MA-02115-5000)

An analytic model is developed for the ultrasound field reflected off a randomly rough vibrating surface for a scanning ultrasound vibrometer system in bistatic configuration. Kirchhoff's approximation to Green's theorem is applied to model reflection and scattering from the vibrating rough surface. The model incorporates the beampattern of both the transmitting and receiving ultrasound transducers and the statistical properties of the rough surface. Conventional approaches for estimating surface displacement and velocity based on the laser vibrometer, such as incoherent Doppler shift spectra and coherent interferometry, are applied to the ultrasound system. Simulations with the model indicate that surface displacement and velocity estimation is highly dependent upon scan velocity and operating frequency of the ultrasound vibrometer relative to the roughness height and correlation length scales of the rough surface. The model is applied to determine optimal sonar design parameters and measurement geometry for the scanning ultrasound vibrometer for measuring ground displacements from acoustic/seismic excitation in land-mine detection.

9:15

4aPA6. Effect of particle friction on the vibrations of a plate loaded with granular media. Wonmo Kang, Liyong Yang, Joseph A. Turner, Kitti Rattanadit, and Florin Bobaru (Dept. of Eng. Mech., Univ. of Nebraska-Lincoln, Lincoln, NE 68588-0526)

Acoustic methods have recently been shown to hold promise for detection of nonmetallic mines. These methods rely on the vibration response of the soil-plate system. The influence of particle size on the frequency response was previously illustrated using sieved sand as the loading material. Here, the influence of particle friction is studied using spherical glass beads. The first resonance of a plate was measured as a function of layer mass for a layer of monodispersed glass beads of various sizes (from 38 μm to 2.5 mm). The plate exhibited the same qualitative behavior as with sand loading but the response was quantitatively different. In particular, the minimum frequency attained for the glass bead layer was consistently higher than that for sand. However, the dependence on particle size followed the similar trends. Mechanisms related to both interparticle friction and layer-plate interface friction were explored. Experimental results with a roughened plate indicate that the resonance behavior is not dependent on the interaction between the layer-plate interface. Finally, numerical results based on dynamic simulations of discrete, interacting particles are discussed as applied to the experiments. The results are expected to impact quantitative aspects of mine detection by acoustic methods. [Work supported by ARL.]

9:30

4aPA7. Comparisons between predicted and measured acoustic scattering from a wedge. Ronald Hughes, Jan Niemiec (NSWCCD, 9500 MacArthur Blvd., West Bethesda, MD 20817), and Herbert Uberall (Catholic Univ. of America, Washington, DC)

Our earlier work has mainly dealt with developing a mathematical approach for predicting acoustc scattering from a wedge for various boundary conditions and wedge material properties. We initially reported scattering for wedge surfaces that satisfy a Dirichlet or a Neumann boundary condition. Recently, our approach was extended to include wedge boundary conditions that can be characterized by an impedance, including absorption. Numerical predictions are found using the Kontorovich-Lebedev transform [A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953)] based upon MacDonald functions, requiring very high precision numerical techniques. We now apply this numerical approach to calculate the monostatic scattering amplitude for a fabricated wedge structure that has been measured in a nonlaboratory environment. Comparisons will be presented and discussed of the results obtained from the predicted and measured monostatic scattering pressures for a range of insonification frequencies at selected offset distances and angles relative to the edge of the wedge.

9:45

4aPA8. Scattering of sound by a cylinder above an absorbing boundary. Kai Ming Li and Wai Keung Lui (The Ray W. Herrick Labs., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47907-2031)

The problem of acoustic scattering by an impedance infinite cylinder irradiated by a point source is investigated. A theoretical model has been developed to consider the case of the sound scattered by an impedance cylinder above an outdoor ground surface of finite impedance. The formulation is based on the use of the appropriate wave-field expansions, the classical image source method, and the translational addition theorem for cylindrical wave function. The solutions can be expressed in the form of a multiscattering interaction between the cylinder and the image cylinder near a flat boundary. The effect of boundary impedance on the reflected sound fields is incorporated in the numerical model by using the well-known Weyl–van der Pol formula. The numerical results are compared with the published data. The effects of various source and receiver locations around the cylinder above the impedance boundary on the sound field are also studied.

10:00–10:15 Break

10:15

4aPA9. A ray/diffraction model of high-frequency backscattering for the study of caustic merging in simple targets. Benjamin R. Dzikowicz (NSWC-Panama City, Code HS-11, 110 Vernon Ave., Panama City, FL 32407)

The backscattering from certain finite geometric targets can be enhanced due to caustic merging when a geometric parameter, such as aspect, is varied. For example, Blonigen and Marston [Blonigen & Marston, ''Backscattering enhancements for tilted solid plastic cylinders in water due to the caustic merging transition: Observations and theory,'' J. Acoust. Soc. Am. **107**, 689–698 (2000)] have shown that there occurs an enhancement for doubly reflected shear waves within a finite plastic cylinder at certain tilt angles. Here, the contributions of single ray classes, such as the doubly reflected shear waves in the example above, are numerically

summed by following ray patches on an incoming reference plane through the target and back, forming a patch on an outgoing diffraction plane. The phase and magnitude of the returning patch depend on reflection and refraction coefficients and attenuation in the target material. When integrated, these patches can form singularities when rays merge at the screen and the patch sizes become infinitesimal. To solve this problem, a change in variables is made and the integration performed over the incoming, rather than the outgoing, patches. By studying the contribution of a single class of rays, individual caustic merging transitions can be studied. The total scattering at steady state can be written as an infinite sum of all classes of rays. [Work supported by ONR.]

10:30

4aPA10. Coherent backscattering without a source. Eric Larose, Oleg I. Lobkis, and Richard L. Weaver (Dept. of Theoretical and Appl. Mech., Univ. of Illinois, 104 S. Wright St., Urbana, IL 61801)

Coherent backscattering is due to constructive interferences of reciprocal paths and leads to an enhancement of the intensity (by factors of up to 2 or 3, typically) of a multiply scattered field near its source. To observe this enhancement an array of receivers must be placed close to the source, within one or two wavelengths. It has been proposed that the effect may be used for the assessment of mean free paths against scattering. Here, we propose and demonstrate a passive technique that allows observation of the enhancement without need of sources. It is performed using time correlation of diffuse fields [Weaver and Lobkis, Phys. Rev. Lett. **87**, 134301 (2001)] obtained by receive-only transducers. It is shown that this passive enhancement cone has an excellent spatial resolution, comparable to active cones previously obtained by others [Tourin *et al.*, Phys. Rev. Lett. **79**, 3637–3639 (1997), Lobkis and Weaver, Phys. Rev. Lett. **90**, 254302 (2003)].

10:45

4aPA11. Determination of effective properties and effective thickness of resonant acoustic metamaterial. Vladimir Fokin, Muralidhar Ambati, Cheng Sun, and Xiang Zhang (NSF Nanoscale Sci. and Eng. Ctr. (NSEC), Univ. of California, 5130 EtcheVerry Hall, Berkeley, CA 94720-1740, xiang@berkeley.edu)

Multiple scattering in periodic structures with strong modulation of elastic constants leads to phononic band structures. According to the Bragg's theory, the spatial modulation of the periodic structure must be of the same order as the wavelength of considered elastic waves. A resonant acoustic metamaterial's respective wavelength is about two orders of magnitude larger than the lattice constant. Each unit cell consists of a locally resonant structural unit in contrast to simple acoustic scatterers in phononic crystals. In this report, we will address the design of locally resonant sonic materials. These materials may offer dynamic effective negative material properties around resonance frequency. Locally resonant sonic crystals, such as an array of Helmholtz resonators and cylinders or spheres coated with acoustically soft material, were analyzed in the literature. However, existing literature lacks a systematic study of the dependence of effective properties for a finite slab of these metamaterials both on the geometry and the properties of its constituent materials. In this report, we present a method for obtaining effective properties and effective thickness of a slab of acoustic metamaterial. The dependence of effective acoustic properties of a metamaterial on the geometrical and acoustic properties of constituent materials will be discussed.

11:00

4aPA12. Homogenization of two-dimensional clusters of rigid rods in air. Daniel Torrent, Andreas Håkansson, Francisco Cervera, and José Sánchez-Dehesa (Wave Phenomena Group, Nanophotonics Technol. Ctr., Universidad Politecnica de Valencia, C/Camino de Vera s/n, E-46022 Valencia, Spain)

The scattering of sound by finite-size clusters consisting of two-dimensional distributions of rigid cylinders in air is studied in the low-frequency limit (homogenization), both theoretically and experimentally. First will be reported analytical expressions for the effective density and sound speed obtained, in the framework of multiple scattering theory, for a circular-shaped cluster. Homogenization is theoretically analyzed as a function of the filling fraction, the type of arrangement of the cylinders in the cluster (hexagonal and square lattice), and the dimension of the cluster. Finally, experiments will be presented for the case of a circular-shaped cluster of 151 wooden cylinders in air. It is demonstrated that, in the long-wavelength limit, this cluster acoustically behaves as a cylinder of argon gas. Moreover, it is shown that homogenization works for frequencies lower than 3 kHz, which corresponds to a wavelength that is four times the parameter of the embedded lattice and is only a quarter of cluster's diameter.

11:15

4aPA13. Wigner distribution of ultrasonic transducers applied to scattering in polycrystalline media. Goutam Ghoshal, Joseph A. Turner (Dept. of Eng. Mech., Univ. of Nebraska—Lincoln, W317.4 Nebraska Hall, Lincoln, NE 68588-0526), and Richard L. Weaver (Univ. of Illinois at Urbana-Champaign, Urbana, IL 61801)

Diffuse ultrasonic backscatter measurements have been especially useful for extracting microstructural information and for detecting flaws in materials. Accurate interpretation of experimental data requires robust scattering models. Quantitative ultrasonic scattering models include components of both transducer beams as well as microstructural scattering information. Here, the Wigner distribution is used in conjunction with a radiative transfer equation (RTE) to model this scattering problem. The Wigner distribution represents a distribution in space and time of spectral energy density as a function of wave vector and frequency. An RTE is derived within the context of the Wigner distribution of various transducer beam models of common interest such that the source and receiver distributions are included in the analysis in a rigorous fashion. The resulting RTE is then simplified in the single-scattering limit typical of many diffuse backscatter experiments. Such experiments, usually done using a modified pulse-echo technique, utilize the variance of the signals in space as the primary measure of microstructure. The resulting backscatter model derived here is compared with previous derivations and experimental backscatter results for materials of common interest. These results are anticipated to impact ultrasonic nondestructive evaluation of polycrystalline media. [Work supported by U.S. DOE.]

11:30

4aPA14. Theoretical prediction of super-radiant modes in ideal and distorted hexagonal clusters of two-dimensional sonic crystals. José Sánchez-Dehesa, Yuriy A. Kosevich, and Daniel Torrent (Wave Phenomena Group, Nanophotonics Technol. Ctr., Polytechnic Univ. of Valencia, E-46022 Valencia, Spain)

Resonance modes confined in hexagonal and threefold symmetry clusters consisting of two-dimensional arrays of rigid cylinders in air have been studied analytically and numerically using a multiple scattering method. It is shown that in the low-frequency (homogenization) limit the triangular modes play a fundamental role in the radiation behavior of the cluster. Analytical formulas are found to predict the frequencies of weakly confined resonance modes, whose pressure patterns have triangular symmetry. In the case of ideal hexagonal cluster, there are two degenerate triangular modes, which result in sixfold symmetry of the far-field radiation. In the case of distorted hexagonal clusters with threefold symmetry, two triangular modes with different frequencies are predicted. These modes have the property of being strongly radiating along three normals to the sides of the distorted hexagonal cluster, having their radiation properties not equivalent. Multiple scattering simulations confirm our analytical formulas and show the predicted threefold symmetry of the far-field radiation from the distorted hexagonal cluster. [Work supported by MEC of Spain.]

Session 4aSAa

Structural Acoustics and Vibration: Vibration—Plates, Shells, and Beams

Jeffrey E. Boisvert, Chair
NAVSEA Newport, Bldg. 1258, Newport, RI 02841

Contributed Papers

8:30

4aSAa1. Perturbation theory of resonance amplification for skewed fan blades. R. Martinez and A. Stokes (Anteon Corp., 84 Sherman St., Cambridge, MA 02140, and Applied Physical Sciences Inc., 4 Muzzey St., Lexington, MA 02421)

The vibrational response of generic turbomachinery blades has been analyzed via a Timoshenko-Mindlin theory of flexure for thick beams formally extended to include torsion due to 3D skew. An alternate interpretation is the generalization of the governing equations of the special case of a radially stiff rib of prismatic cross section to include thick-beam flexure for displacements in the axial direction. The development identifies an artificially small parameter for the coupling of bending and torsion. A subsequent perturbation analysis yields an expression for the system's resonance-amplification envelope in terms of its material loss factor and flexural modal density as affected by torsion. The asymptotic theory reveals how a blade's skew could be exploited to transfer vibrational energy from flexure to torsion and thereby maximize a turbomachine's overall damping for bending. The greatest benefits are found for a band of high frequencies straddling the point at which the dispersive phase speed for thick flexure matches the prismatic member's effective torsional wave speed. Comparisons are made between the closed-form predictions of the perturbation analysis and exact calculations that do not rely on a presumed smallness of the coupling between flexure and torsion.

8:45

4aSAa2. Solution of the 3-D equations of elasticity in prolate spheroidal coordinates. Jeffrey E. Boisvert (NAVSEA Newport, Code 1514 Bldg 1258, Newport, RI 02841) and A. L. Van Buren (Cary, NC 27519)

The 3-D equations of elasticity in prolate spheroidal coordinates are solved using a Helmholtz decomposition for the displacement vector. The displacement vector is written in terms of a scalar potential (dependent on the dilatational wave number) and a vector potential (dependent on the shear wave number). These wave potentials are solutions, respectively, of the scalar and vector wave equations cast in prolate spheroidal coordinates. An eigenfunction expansion of spheroidal wave functions is employed to solve the scalar wave equation. The solution of the vector wave equation is facilitated using expansions of prolate spheroidal vector wave functions. These vector wave functions are constructed by applying certain vector differential operators to the scalar wave functions. The canonical problem of free vibration of a traction-free confocal elastic layer (shell) is considered. The shape of the prolate spheroid is defined by the major-axis length L and minor-axis length D. Since the sphere is a limiting form of the prolate spheroid, the analytical approach is validated by comparing the frequency spectra of a prolate spheroidal shell with $L/D \sim 1$ to those of a spherical shell using a 3-D solution in spherical coordinates. [Work supported by the NAVSEA Newport ILIR Program.]

9:00

4aSAa3. Energy finite-element analysis of the NASA aluminum test-bed cylinder. Aimin Wang, Nickolas Vlahopoulos (NA&ME Dept., Univ. of Michigan, 2600 Draper Rd., Ann Arbor, MI 48109), Ralph Buehrle, and Jacob Klos (NASA Langley Res. Ctr., Hampton, VA 23681)

A formulation was developed in the energy finite-element analysis (EFEA) for modeling the vibration of cylindrical structures with periodic axial and circumferential stiffeners. Appropriate power transfer coefficients are computed from the values of the propagation constants. The joint matrices of the EFEA formulation are computed based on the power transfer coefficients derived from periodic structure theory. EFEA analyses are performed for the NASA aluminum test-bed cylinder and simulations are compared to experiments. Excitation is applied on the test-bed cylinder by four shakers and the vibration is measured at 40 bays formed by axial and circumferential stiffeners. The EFEA results are compared successfully to test data between the $\frac{1}{3}$ octave bands of 315 and 6300 Hz. [Work sponsored by NASA Langley Structural Acoustics Branch.]

9:15

4aSAa4. Detecting and classifying adhesive flaws between bonded elastic plate: The inverse problem. Ricardo Leiderman and Paul E. Barbone (Dept. of Aerosp. and Mech. Eng., Boston Univ., 110 Cummington St., Boston, MA 02215)

A time-harmonic inverse scattering problem to reconstruct the effective stiffness distribution of an adhesive bond in a layered elastic plate was formulated and solved. The motivation is based on the assumption that localized adhesion flaws that diminish bond stiffness also tend to diminish bond strength. The formulation and solution are based on the invariant embedding method, apply to isotropic and anisotropic elastic layers, and are essentially that of identifying embedded acoustic sources in elastic layered structures. Two solutions of the inverse problem are presented: the Born approximation and the exact solution. In the example calculations we compare the two solutions and show that when imperfections are too large the accuracy of the Born approximation breaks down. The impact of noise and uncertainties in the background properties in the inversion is also investigated.

9:30

4aSAa5. Using a scanning laser Doppler vibrometer to measure acoustic intensity. Andrew J. Boone, Jonathan Blotter, Scott D. Sommerfeldt, and Timothy W. Leishman (Dept. of Mech Eng., BYU, 435 CTB, Provo, UT 84602, boone.andrew@gmail.com)

Two different scanning methods were compared to see if acoustic intensity can be accurately measured using a scanning laser Doppler vibrometer (SLDV). Both intensity scans were performed on the same vibrating object, the side panel of a diesel generator enclosure. The first scan was done the traditional way using an intensity probe and scanning grid. The second scan was done using an SLDV. The results from each intensity scan were compared and analyzed. From these results it appears that the SLDV scan is measuring intensity contributed by the vibrating panel alone, while the traditional scan is measuring intensity contributions from both the vibrating panel and the enclosed generator.

9:45

4aSAa6. Simulation of the propagation of the first antisymmetric Lamb mode in the audible range. Guillemette Ribay, Stefan Catheline, Didier Cassereau, and Mathias Fink (Laboratoire Ondes et Acoustique, 10 rue Vauquelin, 75005 Paris, France)

When somebody knocks on a plate-shaped object, the first antisymmetric Lamb wave is created, propagates, and is reflected a lot before vanishing. It has been recently shown [Appl. Phys. Lett. **87**, 204104 (2005)] that if the wave is recorded by an accelerometer, it is possible to deduce the knock position thanks to a cross-correlation technique. Now it is of interest to be able to simulate the wave field created in such experiments, for example in order to study the influence of various parameters, or possibly to try to replace the learning phase of this localization technique by a numerical prediction of the impulse responses. Therefore, in this talk, a finite difference code is presented, simulating the A0 wave in the frequency range excited by a knock on a plate, i.e., [0, 10 kHz]. It is based on a new wave equation of the flexural wave, which takes into account shear correction. Experiments on plates have been carried out to validate the simulation. It is shown that the spectrum of the experimental waves of this reverberant cavity is the same as the simulated one. [This work was financially supported by the TAICHI European project.]

THURSDAY MORNING, 8 JUNE 2006 ROOM 551B, 10:15 A.M. TO 12:00 NOON

Session 4aSAb

Structural Acoustics and Vibration: Vibration in Elastic Structures

Andrew A. Piacsek, Chair
Central Washington Univ., Dept. of Physics, 400 E. University Way, Ellensburg, WA 98926-7422

Contributed Papers

10:15

4aSAb1. Adaptive mechanical model of human footsteps. Alexander Ekimov and James M. Sabatier (The Univ. of Mississippi, NCPA, 1 Coliseum Dr., University, MS 38677, aekimov@olemiss.edu)

The vibration signature from footsteps can be exploited for human characterization and recognition. Footsteps are repeatable cycles of supported surface loading by a dynamic force. This force produces the vibration response of the ground/floor in a broadband frequency range. The vibration response at the distance R from the applied force is proportional to the dynamic force and results from the product of the footstep force and the ground/floor specific transfer function. The transfer function describes the propagation of vibration between the force location and the location of the detector. It was experimentally shown that walkers could vary the dynamic force of footsteps by controlling leg stiffness. An adaptive model of a walking person as a mass-spring system with controlled leg-spring stiffness was developed. This model was tested indoors and outdoors. Three distinct types of walking styles (regular, soft, and stealthy) with the same speed of motion were described and investigated. These styles result in different vibration magnitudes in the low-frequency range (below 100 Hz) due to differences in the walker's leg stiffness. The transfer functions in the building and outdoors were measured and results are presented. [Work supported by Department of the Army, Army Research Office Contract W911NF-04-1-0190.]

10:30

4aSAb2. Elastic and vibrational properties of a regular tensegrity structure. Anthony R. Smith and Andrew A. Piacsek (Central Washington Univ., 400 E. University Way, Ellensburg, WA 98926-7422)

Invented by R. Buckminster Fuller and Kenneth Snelson between 1948 and 1961, tensegrity structures consist of stiff rods connected to flexible strings such that no rod is joined to any other rod. A prestressed tensegrity structure, in which all the strings are under tension, maintains a distinct shape, deforming slightly under a load. For the present study, tensegrity masts consisting of repeated layers of three-, four-, or five-rod patterns were constructed. With 15-cm rods, an eight-layer mast is 50 cm long and has a diameter of 20 cm. All masts exhibited a hysteretic nonlinear elastic response to applied stress along the major axis. Young's modulus for the eight-layer mast ranged from 2500 to 5000 Pa at 8% strain. Increasing the tension in the strings by a factor of 2 resulted in a 14% increase in the Young's modulus at small strains. Implications of this nonlinear behavior for wave propagation along the axis of a tensegrity mast is currently being investigated. Results will be shown for resonance frequencies as a function of drive amplitude. [Work supported by the Murdock Charitable Trust and Central Washington University.]

10:45

4aSAb3. Models of rattling objects for a moving base excitation. Daniel H. Robinson, Robert J. Bernhard, and Luc G. Mongeau (Ray W. Herrick Labs., Purdue Univ., 140 S. Intramural Dr., West Lafayette, IN 47906)

Rattle noise is of particular concern in housing structures near airports. This secondary noise source is often related to large-amplitude low-frequency noise excitation of floors and walls and their interaction with loosely supported objects such as windows, wall hangings, and bric-a-brac. Simple theoretical models were developed to describe interactions between rigid bodies and vibrating surfaces. Rattle models were divided into two classes to describe resonant and nonresonant systems. The models allow a better understanding of the causes of rattle and help identify suitable criteria to mitigate rattle noise. In this presentation, thresholds of rattle onset are discussed and mitigation strategies to reduce and possibly eliminate rattle noise are identified. [Work supported by the Federal Aviation Administration.]

11:00

4aSAb4. Spatial statistics of diffuse elastodynamic modes. Alexei Akolzin and Richard L. Weaver (Dept. of Theoretical and Appl. Mech., Univ. of Illinois, 104 S. Wright St., Urbana, IL 61801)

Spatial statistics of the modes in an elastic body plays a useful role in predicting certain average elastodynamic responses of complex structures, such as transmitted power or nonlinear spectral energy redistribution. For single wave-speed fields, extensive evidence exists that high-frequency modes in the interior of a generic ray-chaotic body resemble a superposition of plane waves with random amplitudes and phases. This implies centered Gaussian statistics for the field amplitude and a specific form for the two-point correlation function. The former can also be linked to the imaginary part of the Green's function averaged over a frequency band or over a characteristic ensemble of body realizations, which allows extension to the boundaries and inclusion of effects of finite body size. In the present study we analyze and compare theoretical expressions for a num-

ber of spatial statistics, such as intensity distribution, wave-type energy equipartition, and two-point correlation function, with numerical results for an ensemble of 2D plane-strain elastic bodies with rough boundaries.

11:15

4aSAb5. Transmission and localization in connected ergodic reverberation rooms. Igor Rozhkov and Oleksandr Zelyak (Dept. of Phys. and Astron., Univ. of Kentucky, 177 Chem.-Phys. Bldg., 600 Rose St., Lexington, KY 40506-0055, irozh2@uky.edu)

The wave power redistribution in two connected reverberant structures represents a paradigm problem in studies of transmission through a complex structure consisting of several reverberant substructures. We analyze randomly coupled ergodic reverberation rooms with the goal to extend such analysis to a more general situation. The two-room problem is, however, quite challenging on its own [R. L. Weaver and O. I. Lobkis, J. Sound Vibr. **231**, 1111–1134 (2000)]. Although the ergodicity assumption allows one to use random matrix theory (RMT), by describing both subsystems with random Gaussian Hamiltonians, the nonperturbative calculations of power transmission and localization have not yet been done. The only attempt made in this direction was the corresponding study of two coupled quantum dots [A. Tschersich and K. B. Efetov, Physi. Rev. E **62**, 2042–2045 (2000)], which used the supersymmetry technique to calculate the joint distribution of local densities of electron wavefunctions and found spatial correlations absent in a single chaotic system. Here, we present both perturbative and nonperturbative (in ensemble averaging) RMT calculations of simplest nontrivial characteristics of power transmission and localization, and compare it with numerics.

11:30

4aSAb6. Free vibrations of a polar elastic body. Glay Altay (Dept. of Civil Eng., Faculty of Eng., Bogazii Univ., Bebek, 34342 Istanbul, Turkey, askarg@boun.edu.tr) and M. Cengiz Dokmeci (Istanbul Tech. Univ., 34430 Istanbul, Turkey)

This paper concerns certain features of the equations governing the time-harmonic free vibrations of a polar elastic body. The governing equations of micropolar elasticity are expressed in differential form; then, the uniqueness in their solutions is investigated. The conditions sufficient for the uniqueness are enumerated using the logarithmic convexity argument without any positive-definiteness assumptions of material elasticity. Applying a general principle of physics and modifying it through an involutory transformation, a unified variational principle is obtained for the free vibrations. The principle leads to all the governing equations of the free vibrations, as its Euler-Lagrange equations. The governing equations are alternatively expressed in terms of the operators related to the kinetic and potential energies of the body. The basic properties of vibrations are studied and a variational principle in Rayleigh's quotient is given, and the high-frequency vibrations of a micropolar plate [cf., Altay and Dkmeci, Int. J. Solids Struct. (2006)] are likewise treated. [Work supported by TUBA.]

11:45

4aSAb7. Progress towards a diffusion theory of wave energy transport in large irregular structures. Richard Weaver and Nicholas Wolff (Dept. of Theoretical and Appl. Mech., Univ. of Illinois, 216 Talbot Lab., 104 So. Wright, Urbana, IL 61801)

We address the problem of constructing predictions for mean-square bandlimited displacement (''energy'') at points in a complex structure, and doing so with information gleaned from short-time cost-effective direct numerical simulations (DNS). Past work has shown that a DNS over short times subsequent to the action of a transient force allows good estimates of band-averaged admittances and therefore also of sensitivity to local energy density. (A further assumption of equipartition and no damping then permits estimates of response levels at infinite time.) We note that short-time DNSs also reveal a degree of transport and find that the transport operator observed to govern flow over short times may be iterated to predict energy distributions at later times. When carefully constructed, this iteration does a good job of predicting actual energies—at least for the special cases studied so far.

THURSDAY MORNING, 8 JUNE 2006 BALLROOM D, 8:30 TO 11:45 A.M.

Session 4aSC

Speech Communication: Production and Perception in First and Second Languages

Harriet S. Magen, Chair

Rhode Island College, Dept. of Communications, 600 Mt. Pleasant Ave., Providence, RI 02908

Contributed Papers

8:30

4aSC1. Context matters!—The influence of listening materials on L2 speakers' performance of devoiced liquids in English. Sally Chen and Janice Fon (Grad. Inst. of Linguist., Natl. Taiwan Univ., 1 Roosevelt Rd., Sec. 4, Taipei 106, Taiwan)

For previous L2 studies, the influence of context could be categorized either as the duration speakers live in an environment with their L2 as the official language or as the retaining effect after a certain amount of training. This study aims to investigate the variations on L2 learners' production of English devoiced liquids with contexts of different degrees of directness provided. Sixteen subjects are divided into three groups. They are given an equal length of story (indirect context), wordlist (direct context), or music (control) in the listening task between two recordings of their production of a wordlist consisting of 14 monosyllabic English words containing a voiceless stop in the first position and a liquid in the second (e.g., play). Duration and ratio of the consonant clusters are measured and so is the voiceless portion of the clusters. Preliminary results showed that after the listening task, participants had a significantly longer duration for both cluster length and voiceless parts. Cluster duration of the two experimental groups was also lengthened. Furthermore, the wordlist group had their cluster/word and voiceless/cluster ratio closer to the native speakers in the second recording. Thus, practice effect, context effect, and directness effect were all supported.

8:45

4aSC2. The hybrid system of voice onset time in French/English bilinguals. Claire Gurski (Univ. of Western Ontario, 509-880 Kipps Ln., London, ON N5Y 5K4, Canada, mchough@uwo.ca)

This study will look at the production of occlusive consonants by bilinguals in both their first and second languages, namely French and English. Due to the inherent differences in the voice onset times (VOTs) of these two languages, do second language speakers acquire a native-like voice onset time when speaking in their second language or do traces of their first language remain? It will be shown that although French and English have categorical distinctions in their VOT systems [Keating ''Phonetic and phonological representation of stop consonant voicing,'' Language, **60**, 286–319 (1984) Lisker & Abramson, ''A cross-language study of voicing in initial stops: Acoustic measurements,'' Word. **20**, 384–422 (1964)] a hybrid system of the two languages is produced regardless of first language. This research is part of a larger continuing study looking at speaker identification of bilinguals. Due to the right to a trial in either of the official languages of Canada, French or English, it is impossible to force a bilingual suspect to supply a voice sample that can be compared with an incriminating recording in one of the two languages. In such a situation, is it possible to base an identification (or a nonidentification) on the comparison of samples in the two languages of the suspect?

9:00

4aSC3. The language interference of pitch accent language on tone language—A case study of Mandarin and Swedish. Yi-Chen Tung (NCTU Dept. of Foreign Lang. and Lit.s, 1001, Ta-Hsueh Rd., Hsinchu, Taiwan, 300)

The study aims to find how pitch accent language influences learning of tone language. The target language is Mandarin which has three contour lexical tones and one level lexical tone. The source language is Swedish which has two contour accents. It is assumed that Swedish cannot produce a Mandarin words with contour tone applied on single syllable since Swedish always exerts contour tone over two or more syllables. In experiment 1, it was found that the Swedish could produce a Mandarin contour tone completely. On the other hand, it was found that the Swedish subject used accent 1 to produce FR (falling+rising) and RL (rising +falling-rising) tonal combinations and accent 2 to produce RF (rising + falling) and LR (falling-rising+rising) tonal combinations. To further investigate the learning pattern of tone language, all the possible tonal combinations in Mandarin were tested in experiment 2. It was found that the Swedish subject applied accent 1 to produce all the tonal combinations. The result revealed that the language interference of pitch accent on tone language could be illustrated on the manipulation of direction of F0 movement.

9:15

4aSC4. Comprehension of envelope-only sentences by English-speaking adults and children and Mandarin-speaking adults. Susan Nittrouer (Speech and Hearing Sci., Ohio State Univ., 1070 Carmack Rd., Columbus, OH 43210) and Robert Packer (Washington State Univ., Pullman, WA 99164)

Listeners comprehend sentences when only amplitude envelope information for a few bandlimited channels is available, but there is variability in their abilities to do so. This experiment asked whether the variability in listeners' abilities to comprehend these signals is best explained by variability in magnitude of linguistic context effects, in auditory processing, or in sensitivity to structural organization at a global level. Stimuli were 30 four-word syntactically appropriate, semantically anomalous sentences, processed to preserve envelope information in 4 or 8 channels. Listeners were two groups of native English speakers (adults and 7-year-old children) and one group of native Mandarin speakers (adults) who were proficient enough in English to be graduate students at an American university. The sentences provided linguistic context that should affect comprehension equally for all these listeners, so that possible outcomes were (1) All three groups would perform similarly if abilities to use linguistic context best explained previously reported variability; (2) both adult groups would perform similarly and better than children if age-related differences in auditory processing best explained variability; and (3) English-speaking children and Mandarin-speaking adults would perform similarly and poorer than English-speaking adults if sensitivity to global structure best explained variability. Results matched possible outcome #3. [Work supported by NIDCD Grant No. DC-00633.]

9:30

4aSC5. The error patterns of different nonword repetition tasks in 3-year-old Mandarin-speaking children. Naihsin Li, Joanna Lee, Janice Fon, and Hintat Cheung (Grad. Inst. of Linguist., Natt. Taiwan Univ., No. 1, Sec. 4, Roosevelt Rd., Taipei, 10617 Taiwan, r93142004@ntu.edu.tw)

This study investigates qualitatively and quantitatively 3-year-old Mandarin-speaking children's error patterns in two nonword repetition tasks which differ in their degrees of wordlikeness, i.e., the nonce word and the gap word sets. Their performances in the two tasks were transcribed and analyzed from two respects: (1) the rate of a specific unit retaining the target structure and (2) the cross comparison between linguistic units and error processes. Results showed that tones and syllable structures of the target forms were easier to retain than the contents. And even though children's error patterns approximated the adults patterns in gap word repetition task, they had strategic difference in dealing the weight decay in memory: while the children used syllable and rhyme substitutions, the adults used syllable omissions. Our study has implied that children have the same processing mechanism with adults in learning novel words. However, their strategic difference may have an implication for language learning. In addition, the present study also contributes the processing of tone to the linguistically constrained model (Hartley & Houghton, ''A linguistically constrained model of short-term memory for nonwords,'' J. Mem. Lang. **35**, 1–31 1996), so that it can become a more complete model in demonstrating how novel sound sequences are learned and recalled in Mandarin-speaking young children.

9:45

4aSC6. Response to subthalamic deep brain stimulation on speech intelligibility in Parkinson's disease. Emily Wang (Dept. of Commun. Disord. and Sci., Rush Univ. Medical Ctr., 1653 West Congress Pkwy., Ste. 203 SENN, Chicago, IL 60612), Leo Verhagen Metman (Rush Univ. Medical Ctr., Chicago, IL 60612), Roy Bakay (The Chicago Inst. of Neurosurgery and Neuroresearch), and Bryan Bernard (Rush Univ. Medical Ctr., Chicago, IL 60612)

The study reports on intelligibility of reading and monologues produced by Parkinsonian patients who underwent unilateral subthalamic deep brain stimulation (STN DBS). Twenty right-handed individuals with advanced Parkinson's disease (PD) and dysarthria underwent unilateral STN DBS. Ten were operated on the right hemisphere and ten on the left hemisphere. The side receiving STN DBS had more affected motor function. Speech was evaluated before surgery and three-to-six months after surgery with stimulator-off and with stimulator-on; all were off anti-Parkinsonian medication for 12 h before evaluation. Evaluators and subjects were blinded to the subjects stimulator status at the postsurgery evaluations. Nonspeech motor performance was assessed with UPDRS-III. Each subject read a standard passage and produced a monologue on a preselected topic. The readings and monologues were rated for intelligibility using UPDRS-III item 18. Sixteen student clinicians in speech pathology served as judges. Each speech sample was repeated three times. The reading and monologue samples were presented in two blocks. Within each block, the samples were randomly presented to each judge over headphones. Presentation and response collection were completed using a computer procedure written in Praat. The results are discussed with regard to speech and nonspeech motor response and to hemisphere-specific response on speech intelligibility.

10:00–10:15 Break

10:15

4aSC7. Vowel imitation of 4-year-old children. Richard S. McGowan (Haskins Labs. and CReSS LLC, 1 Seaborn Pl., Lexington MA 02420, rsmcgowan@earthlink.net)

Published formant scalings for vowels of American 4-year-olds who were imitating adult production are used along with published data of American adult male vowel production to synthesize /ah, ae, uu, ii/. Perturbations to these vowels were also synthesized. Adult listeners were asked to categorize these synthetic vowels in a forced-choice task. Except for a preference for slightly more extreme first formant frequencies and higher second formant for /ae/, the formant frequencies preferred for the vowels /ah, ae, uu, ii/ were close to the values used for synthesis. A five-tube model was used in an algorithm to infer vocal-tract shape from the first three formant frequencies of the adult productions, the children's synthesized /ah, ae, uu, ii/, and the formant frequencies derived for these vowels based on the listening experiments. It was found that the rear tube length for the children, in proportionate terms, was nearly always greater than that of the adult. In proportionate terms the rear tube length was twice as long in children compared to adults for the vowel /uu/. Tongue root flexibility may be more important than pharynx length in determining formant scalings for young children. [Work supported by NIHD-003782 to Haskins Laboratories.]

10:30

4aSC8. Pitch and voice quality in Burmese vowels. James F. Gruber and Zhaleh Feizollahi (Georgetown Univ., Washington, DC 20057, jfg27@georgetown.edu)

Burmese vowels are pronounced with a four-way contrast in laryngeal configuration, described as high, low, creaky, and killed registers [Bradley, Tonation, Pacific Linguistics Series A-62 117–132 (1982)]. However, acoustic descriptions differ as to whether the registers should be defined in terms of pitch or voice quality [Watkins SOAS Working Papers in Linguistics: **10**, 139–149 (2000)]. This study investigates whether voice quality or pitch is the more consistent carrier of contrast across the four registers. Four speakers pronounced 20 words (5 vowels $\times 4$ registers) inside a frame sentence surrounded by syllables of low register. Each sentence was repeated five times consecutively. For each syllable, F0 measures were taken at each quartile, as well as peaks and valleys. Voice quality was quantified in terms of spectral slope [Gordon and Ladefoged, J. Phonetics **29**, 383–406 (2001)], specifically the difference in dB between the first and second harmonics in an FFT spectrum over a 50-ms window. Results indicate that pitch more reliably indicates register than voice quality. While there was no significant difference in spectral slope between registers, the four registers differed consistently in height and timing of pitch peaks.

10:45

4aSC9. Initial pitch informs sentence duration. Johanna Tan, Joel Dunham, Caleb Lee, and Eric Vatikiotis-Bateson (Dept. of Linguist., Univ. of British Columbia, E270-1866 Main Mall, Vancouver, Britsh Columbia V6T 1Z1, Canada)

This study re-examines the potential relation between sentence duration and sentence-initial fundamental frequency ($F0$), in both speech production and perception, and provides perceptual evidence for speech planning. Liberman and Pierrehumbert [in *Language Sound Structure*, edited by Aronoff & Oerhle (MIT Press, Cambridge, 1984)] argued that a correspondence between sentence length and initial $F0$ would accommodate declination and could be evidence for speech planning. Unfortunately, their study did not provide strong evidence either for or against this intriguing possibility. The present study has two parts: First, read sentences were recorded for native English speakers and the relation between sentence length and $F0$ was analyzed. With a few exceptions, longer sentences had higher initial $F0$. Second, 24 listeners were presented with initial portions (300 and 700 ms) of the recorded sentences and asked to make both relative and absolute predictions of total sentence duration. Subjects consistently associated sentence fragments containing higher initial $F0$ with longer sentences. [Research supported by NSERC and CFI.]

11:00

4aSC10. Position variability in stop and fricative production. Anders Lofqvist (Haskins Labs., 300 George St., New Haven, CT 06511, lofquist@haskins.yale.edu)

This study analyzed position variability in stop and fricative consonants, examining potential differences related to precision requirements. It has often been claimed that fricatives require more precision in their production than stops due to the management of airflow. However, there is very little, if any, empirical evidence for this claim which is mostly based on different developmental trends. Tongue movements were recorded using a magnetometer system. Measurements were made of tongue position at the onset of the consonant, defined in the acoustic signal. The results do not show any consistent pattern within or across subjects for stops and fricatives. Thus, there is not less position variability for fricatives than for stops. Taken together with results showing that the tongue movement during the closure is no less constrained for a fricative than for a stop, the present results thus indicate that the precision requirements for stops and fricatives do not differ. These results most likely reflect the fact that speech is an overlearned skill, so that any differences related to precision are very small, or nonexistent, in adults, and might only be seen in the early stages of speech development or second language learning. [Work supported by NIH.]

11:15

4aSC11. $F0$ and the voicing states of initial consonants in Mandarin. Chunling Zhang and Terrance M. Nearey (Dept. of Linguist., Univ. of Alberta, 4-32, Assiniboia Hall, Edmonton, AB, T6G 2E7 Canada, chunling@ualberta.ca)

Studies have shown inconsistent $F0$ changes following the plain voiceless stops ($F0_{\mathrm{PLAIN}}$) compared to (1) voiced stops ($F0_{\mathrm{VOICED}}$) and (2) aspirated voiceless stops ($F0_{\mathrm{ASPIRATE}}$). Aerodynamic, laryngeal setting and auditory enhancement models of [$\pm$voice] have all been implicated in $F0$ variations [J. Kingston and R. L. Diehl, Language **70**, 419–454 (1994)], but each explains only certain aspects of the prevocalic voicing effects on $F0$. Studies in English show that $F0$ of a vowel is simultaneously affected by multiple phonetic factors, such as voicing state, vowel identity, intonation, and place of articulation. Previous small-scale experiments in Mandarin also exhibit interactions of the phonetic distinctions on $F0$, but only certain aspects of the contextual variations of $F0$ have been examined. This study carries out a large-scale recording of Mandarin CV syllables in isolation, examining the interactive patterns of acoustic correlates of the voicing states of initial consonants. Preliminary results show dramatically different $F0$ changes from previous studies in Mandarin [C. X. Xu and Y. Xu, JIPA, **33**, 165–181 (2003)] but show similar patterns to English, such that $F0_{\mathrm{PLAIN}}$ is similar to $F0_{\mathrm{VOICED}}$ (nasals), with $F0_{\mathrm{ASPIRATE}}$ significantly higher. This study provides a better understanding of the cross-linguistic phonetic realization of voicing distinctions.

11:30

4aSC12. The nature of aspiration in stop consonants in English. Helen M. Hanson and Kenneth N. Stevens (Speech Commun. Group, Res. Lab. of Electron., MIT, 77 Massachusetts Ave., Cambridge, MA 02139, hanson@speech.mit.edu)

The releases of aspirated stops in English are typically modeled as having three consecutive phases, which overlap somewhat in time: (1) transient, (2) frication, and (3) aspiration. Close examination reveals that the noise spectrum in the aspiration phase is sometimes dominated by one spectral prominence, rather than several prominences as expected with a glottal source. In this work, the possibility that frication noise generated during the third phase may sometimes dominate the aspiration noise is explored. The nature of the radiated sound during the production of both voiced and unvoiced stop consonants is examined for the three places of

articulation in English and with several different following vowels. Data from eight subjects are observed. Results suggest that some subjects follow the classical model, but other subjects produce a mix of frication and aspiration during the third phase. There is considerable variability in these acoustic attributes, both within a speaker and across speakers. It is suggested that speakers can choose between using (1) an extended burst or (2) formant transitions to provide enhancing cues to place of articulation. The relation of these data to acoustic theories of sound production and to perceptual saliency is discussed. [Work supported by NIH Grant No. DC00045.]

THURSDAY MORNING, 8 JUNE 2006 ROOM 556AB, 9:00 TO 11:45 A.M.

Session 4aSP

Signal Processing in Acoustics, Underwater Acoustics, and Psychological and Physiological Acoustics: Aural Processing of Sonar Signals

Paul C. Hines, Chair
Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, NS, B2Y 3Z7, Canada

Chair's Introduction—9:00

Invited Papers

9:05

4aSP1. Perception and cognition of acoustic signals. Laurie M. Heller (Dept. of Cognit. and Linguist. Sci., Brown Univ., Providence, RI 02912)

Sonar operators may distinguish the sound of a submarine from the sound of a fishing trawler amidst a background of snapping shrimp. This accomplishment requires integrating auditory sensation and perception with auditory cognition, experience, and the ability to adapt to context. The fundamentals of auditory perception and cognition required for this task will be reviewed in this talk. I will cover a range of levels in the auditory system: from the ability to detect a signal in a background of noise, to the ability to distinguish a complex pattern from a complex background, to the ability to analyze the physical properties of an event. This review of both empirical data and research will cover a variety of psychoacoustic testing methods that is appropriate for different situations that are likely to be encountered in underwater signal processing, such as detection, discrimination, and identification.

9:25

4aSP2. Active listening in passive-sonar soundscapes. Gregory Wakefield (Elec. Eng. and Comput. Sci. Dept., Univ. of Michigan, 1301 Beal Ave., Ann Arbor, MI 48109, ghw@umich.edu) and Thomas Santoro (Naval Submarine Medical Res. Lab., Groton, CT)

Listening tasks differ across a variety of dimensions. Tasks that are characteristic of competing channels are those in which command and control signals are delivered simultaneously. Such signals are typically highly uncertain, statistically independent, and nonredundant. Tasks that are characteristic of cooperative channels are those in which sources vary in some coordinated or collaborative fashion. Such signals are typically of much greater certainty, statistically correlated in several of their parameters, and highly redundant. We report on results from several types of cooperative-channel tasks in which listeners attend to targets in multiple-source soundscapes that resemble the passive sonar listening environment. Several levels of rendering are considered ranging from monaural to individualized binaural auditory displays with head motion. As the number of sources increases, performance degrades but at a slower rate than observed for competing channels. The fidelity of the binaural auditory display also affects performance, with greater performance gains being realized by more accurate realization of the listener's head-related transfer functions. Finally, we introduce results concerning the perceptual asymmetry between source and head rotation and discuss the importance of head rotation in cooperative-channel tasks. [Research supported by a grant from the Office of Naval Research.]

9:45

4aSP3. Aural classification of impulsive-source active sonar echoes. James Pitton (Appl. Phys. Lab, Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698), Scott Philips, Les Atlas (Univ. of Washington, Seattle, WA 98195-2500), James Ballas, Derek Brock, Brian McClimens (Naval Res. Lab., Washington, DC 20375-5337), and Maxwell Miller (ARCON Corp., Washington, DC 20375-5337)

The goal of this effort is to develop automatic target classification technology for active sonar systems by exploiting knowledge of signal processing methods and human auditory processing. Using impulsive-source active sonar data, formal listening experiments were conducted to determine if and how human subjects can discriminate between sonar target and clutter echoes using aural cues alone. Both trained sonar operators and naive listeners at APL-UW were examined to determine a baseline performance level. This level was found to be well above chance for multiple subjects in both groups, validating the accepted wisdom that there are inherent aural cues separating targets from clutter. In a subsequent experiment, feedback was provided to the naive listeners and classification

performance dramatically improved, demonstrating that naive listeners can be trained to a level on par with experts. Using these trained listeners at APL-UW, a multidimensional scaling (MDS) listening experiment was designed and conducted. The results of these experiments and an analysis of the data, particularly its correlation with the physical attributes of target and clutter echoes (i.e., signal features), will be presented.

10:05–10:20 Break

Contributed Paper

10:20

4aSP4. Study on the human ability to aurally discriminate between target echoes and environmental clutter in recordings of incoherent broadband sonar. Nancy Allen, Paul C. Hines, Victor W. Young, and Douglas A. Caldwell (Defence R&D Canada Atlantic, P.O. Box 1012, Dartmouth, NS, Canada B2Y 3Z7, nancy.allen@drdc-rddc.gc.ca)

Unacceptably high false-alarm rates due to the inability to discriminate between target echoes and environmental clutter are an issue for existing low-frequency active sonar systems operating in coastal environments. A research project at Defence R&D Canada—Atlantic is investigating the potential use of aural cues to tackle this challenge. One aspect of the project is to evaluate the human ability to aurally discriminate between target echoes and environmental clutter. The design and preliminary results from the study are presented here. Human subjects are presented with a series of sounds containing target echoes and clutter obtained from recordings of an incoherent broadband sonar experiment. The quantitative data collected in the study are the subjects' decisions as to whether the echo heard was a target echo or clutter and their level of confidence associated with the decisions. Receiver-operating characteristic (ROC) analysis is used to produce a statistical model of the subjects' performance. The study also includes a questionnaire: answers may prove useful in supporting the quantitative results and in providing a better understanding of the cues and decision techniques used by the subjects.

Invited Papers

10:35

4aSP5. Generalizing timbre space data across stimulus contexts: The meta-analytic approach. Stephen McAdams (CIRMMT, Schulich School of Music, McGill Univ., 555 Sherbrooke St. W., Montréal, QC, H3A 1E3, Canada, smc@music.mcgill.ca) and Bruno L. Giordano (McGill Univ., Montréal, QC, H3A 1E3, Canada)

Studies of the multidimensionality of the complex auditory attribute called timbre often represent the perceptual structure in terms of a spatial model in which distance has a monotonic relation to perceived dissimilarity. Acoustical correlates are sought for the dimensions of these timbre spaces in order to quantify them psychophysically. An important question concerns the extent to which the results obtained with one stimulus set generalize to other stimulus sets. A meta-analytic study of data from several timbre spaces using musical tones was performed with the same multidimensional scaling algorithm. The effect of changing a subset of the sounds on the perceptual relations among common sounds was evaluated in conditions where the degree of variation along perceptual dimensions across sets was similar or different (new salient dimensions added). A change in the sounds present in the stimulus set, keeping salient dimensions constant, had little effect on perceptual structure of the common sounds in the two sets. Changing salient acoustical correlates present, while resulting in new dimensions, did not strongly affect the perceptual structure of the common sounds. Results demonstrate that, while a change in stimulus set may introduce new perceptual dimensions, the perceptual structure among sounds common to two sets remains similar.

10:55

4aSP6. Active sonar classification using perceptual signal features from musical acoustics. Victor W. Young, Paul C. Hines, and Sean Pecknold (Defence R&D Canada-Atlantic, Box 1012, Dartmouth, NS B2Y 3Z7, Canada, victor.young@drdc-rddc.gc.ca)

Because of the relatively low frequencies at which they operate, navy active sonars are often plagued by false-alarm returns resulting from geological structures. In the lexicon of sonar operators these false returns are referred to as geoclutter or simply clutter. Despite mounting evidence that human operators can aurally discriminate target returns from clutter, attempts to develop robust automatic classification algorithms have, as yet, met with limited success. This paper investigates the possibility of improving the performance of automatic classifiers by exploiting the signal processing employed in the human auditory system. This amounts to replacing the statistical signal features used by conventional automatic classifiers with perceptual signal features that reflect the way a human listener would perceive a given return. Drawing an analogy between active sonar returns and percussive musical timbre, this paper considers perceptual signal features—like loudness centroid, sub-band attack/decay, and sub-band correlation—that have previously been identified as underlying the perception of timbre. Values for each of these timbre features are measured for a series of target and clutter returns recorded during an experimental sea trial. The effectiveness of these perceptual features as target-clutter discriminators is then evaluated using a Gaussian linear discriminant classifier.

Contributed Papers

11:15

4aSP7. Automatic feature identification using multidimensional scaling data. Scott Philips, Les Atlas (Elec. Eng. Dept, Univ. of Washington, Campus Box 352500, Seattle, WA 98195-2500), and James Pitton (Univ. of Washngton, Seattle, WA 98105-6698)

Multidimensional scaling (MDS) listening experiments have long been used to better understand the perceptual cues human use when distinguishing between sounds in a given dataset. By collecting aural similarity measures between sound pairings, one can find the position of each sound in a D-dimensional perceptual space. The space allows the researcher to explore the perceptual attributes used to distinguish sounds. Typically, candidate signal features are correlated to each dimension to find which features accurately predict the signal positions in the perceptual space. Signal features generally exhibit varying degrees of correlation, with the optimal (highest correlated) feature declared to be the one used in perception. This method is not only ambiguous as to which feature (or combination of feature) is important to perception, but also is limited to previously defined signal processing features. Instead of optimizing over a limited set of features, one can optimize over a function which maps our signals to the

perceptual MDS results. This approach allows one to find a transform of the data that better represents the complicated perceptual space. A number of solutions to the problem will be presented based on the results of an MDS experiment using active sonar echoes.

11:30

4aSP8. Information-theoretic limits on spatial auditory display of sonar beam data. Jason E. Summers and Charles F. Gaumond (Acoust. Div., Code 7142, Naval Res. Lab., Washington, DC 20375-5350, jason.summers@nrl.navy.mil)

A signal-processing-based method that determines the number of independent channels in a set of head-related impulse responses (HRIRs) [B. Gardner and K. Martin, http://sound.media.mit.edu/KEMAR.html] is presented. The HRIRs in the horizontal plane are analyzed using information-theoretic methods and the resulting number of independent channels is shown as a function of noise level. These results are compared to previous psychophysical measurements of absolute and relative localization thresholds as a function of signal spectrum and target angle. This comparison leads to a noise level related to human perception. Additionally, the information-theoretic treatment is shown to yield a representation of the HRIR in the horizontal plane as a one-dimensional manifold in multidimensional coefficient space. An interpolation method is presented that is based on a distance measure on this manifold and is related to previous work [R. Duraiswami and V. C. Raykar, Proc. ICASSP **3**, 285–288 (2005)]. Different distance measures are compared. Finally, these results are discussed in context of presenting beamformed sonar data via a spatial auditory display using a mapping of beams to sources in a virtual auditory space [G. R. Arrabito *et al.*, Appl. Acoust. **66**, 986–1005 (2005).] [Work supported by ONR.]

THURSDAY MORNING, 8 JUNE 2006 ROOM 552A, 8:25 TO 11:20 A.M.

Session 4aUW

Underwater Acoustics: High-Frequency Acoustic Propagation and Applications I

Mohsen Badiey, Cochair
Univ. of Delaware, College of Marine Studies, Robinson Hall, Newark, DE 19716

James C. Preisig, Cochair
Woods Hole Oceanographic Inst., Dept. of Applied Ocean Physics and Engineering, Woods Hole, MA 02543-1053

Chair's Introduction—8:25

Invited Papers

8:30

4aUW1. Internal and surface wave effects on inverted echo sounder measurements. David M. Farmer (Grad. School of Oceanogr., Univ. of Rhode Island, Narragansett, RI 02882) and Timothy F. Duda (Woods Hole Oceanographi Inst., Woods Hole, MA 02543)

Inverted echo-sounders operating at 10 kHz, long used to observe mesoscale synoptic variability from the sea-floor through changes in acoustic travel time and pressure, were recently deployed in the South China Sea to explore their potential for detecting nonlinear internal waves. The observations, based on a 6-s transmission rate, reveal both the internal tide and high-frequency nonlinear internal wave packets and are consistent with remotely sensed images and nearby temperature time series. The results motivate exploration of a refinement in instrument operation in which the raw signal is digitized and stored at the full transducer bandwidth, rather than the present scheme of narrow band filtering and threshold detection. If the raw signal is available, it will be related to the surface wave field, leading to the prospect of ocean surface environmental measurements from the sea-floor. An intriguing possibility is that internal wave modulation of the surface wave field, so evident in remotely sensed images, including the consequence of surface wave breaking in convergence zones, might be detectable in this way from the ocean floor, thus linking internal wave observations directly with surface wave modulation. Model calculations of surface scatter are used to gain insight on this approach.

8:50

4aUW2. HF propagation in the ocean waveguide: Progress and open issues. Michael Porter (Heat, Light, and Sound Research, 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130, mikeporter@hlsresearch.com)

The HF band, defined for this talk as 5–50 kHz, has become of interest to us for many reasons. This operating band includes the musical chords and less-musical hisses of acoustic modems, the whistles and clicks of the common dolphin, and the newly recorded sounds of beaked whales. This band is of interest for mine hunting and torpedoes. Further, new HF tomography systems are emerging that exploit this band to produce lightweight systems to image the ocean weather, including both thermal fronts and currents. In turn, the latter are particularly important to lightly powered autonomous vehicles and gliders. To the surprise of many used to lower-frequency systems, sound in this band can produce identifiable echoes involving many reflections from a sea surface broken by waves and quite rough ocean bottom. Thus, the role of both dynamic and static boundaries is important, as well as refractive effects of the ocean volume. We review recent progress and discuss open issues in this area.

9:10

4aUW3. The coupled oceanographic-tomographic analysis and prediction system: A practical means for utilizing high-frequency tomography in the world's oceans. Peter J. Stein (Sci. Solutions, Inc., 99 Perimeter Rd., Nashua, NH 03063), James K. Lewis (Sci. Solutions, Inc., Kalaheo, HI 96741), and Subramaniam D. Rajan (Sci. Solutions, Inc., Nashua, NH, 03063)

Over the last several years, technology has been developed to transition ocean tomography out of the laboratory and into practical use. This technology, the coupled oceanographic-tomographic analysis and prediction system (COTAPS), couples an ocean circulation model in a region with travel time information from a network of high-frequency acoustic nodes in the same region. A data assimilation technique is used to impose corrections to the ocean model temperatures and salinities based on observed arrival time anomalies and a Kalman gain matrix. Details of the features of COTAPS are presented and discussed. This includes the technique and associated software for assessing ray path information in the area in which the network of nodes is deployed, an automated method for calculating the Kalman gain, and a suggested design of the acoustic nodes. We also discuss a few issues that remain to be researched. These include significant variability of observed arrival times over short time intervals (less than 5 min) and how to mitigate the impact of such variability, the use of observed arrival times to account for errors in the locations of the nodes in a network, and the effect of assimilating arrival time information from multiple source-receiver pairs.

9:30

4aUW4. Modeling of the intensity channel impulse response for the East China Sea. Jee Woong Choi and Peter H. Dahl (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698, choijw@apl.washington.edu)

Results of modeling the intensity channel impulse response in shallow and comparison with measurements are presented. The measurements were made in the East China Sea (depth 100 m, range ~ 500 m) as part of the Asian Seas International Acoustics Experiment (ASIAEX). The intensity channel impulse response is obtained by incoherent summation of separate impulse response functions for each path (eigenray); the latter are modeled using ray and bistatic scattering methods. Measured sea surface spectral information and inverted seabed properties (from previously published works) are incorporated in the modeling. The modeled channel impulse response is convolved with representations of the CW (center frequency 8 and 16 kHz) or FM (frequency range 8–16 kHz) waveforms that were transmitted for comparison with observations. This talk will discuss (1) the influence of two-dimensional (out-of-plane) scattering including near-surface bubbles and the volume inhomogeneities in the sediment, on the intrapath time spread, and (2) the contribution of each arrival (up to the sixth path) to the channel impulse response. [Work supported by Office of Naval Research.]

Contributed Papers

9:50

4aUW5. Environmental effects on high-frequency propagation during the Makai experiment. Thomas Folegot (NATO Undersea Res. Ctr., MILOC, Viale S. Bartolomeo 400, 19138 La Spezia, Italy), Vincent K. McDonald (Space and Naval Warfare Res. Systems Ctr., San Diego, CA), Sergio M. Jesus (Univ. of Algarve, Faro, Portugal), Mohsen Badiey (Univ. of Delaware, Newark, DE 19711), Paul Hursky, and Michael B. Porter (Heat, Light, and Sound Res., Inc., San Diego, CA)

The Makai experiment was conducted from 15 September to 2 October 2005 off Kauai Island to provide a comprehensive study of acoustic propagation in the 8–14-kHz band for diverse applications. Thermistor string measurements show how the interaction between the barotropic tides and the bathymetry generates a highly dynamic environment. Strong variability in the sound-speed profile has been observed, with diurnal and semidiurnal thermocline depth changes up to 50 m. Acoustic waves have been propagated at several ranges between two sources and receiver arrays, using fixed and drifted setup. The structure of the received impulse functions shows strong variability, although resolved arrivals are persistent.

10:05–10:20 Break

10:20

4aUW6. Modeling Doppler effects for acoustic communications. Martin Siderius, Michael B. Porter, Paul Hursky (HLS Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130), and Vincent McDonald (SPAWAR Systems Ctr., San Diego, CA 92152)

Performance of high-data-rate, bandwidth-efficient communications signals can be significantly degraded by failing to compensate for Doppler effects. Significant Doppler spread can be introduced simply from the sound interacting with the moving sea surface; however, the effects can be much greater when the source and receiver are also in motion. Numerical modeling that accurately accounts for Doppler effects can be used to develop and test communications algorithms. In this presentation, a simple technique will be presented for modeling communication signals in the ocean that includes Doppler effects. The challenge for modeling is due, in part, to the broad band and high frequencies (e.g., 3–50 kHz) typically used in underwater communications. The frequency band suggests using Gaussian beam tracing to determine the field. In addition, the Gaussian beams are conveniently interpolated and extrapolated to allow for the treatment of Doppler effects due to environmental and platform motion. The technique will be tested using simulations and experimental data from Doppler-sensitive transmissions.

10:35

4aUW7. Scattering by a set of sediment rough layers from a normal incident beam. Dajun Tang (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105)

The seabed can often be modeled as layers, which are in general not flat. The ability to model scattering from such rough layers is important for interpreting data from acoustics systems such as chirp sonar and echo sounders. The goal of this work is to investigate this scattering problem for a normal incidence scenario. A set of one-dimensional fluid rough interface realizations was generated using Monte Carlo method. Integral equations for the problem are solved, which include all multiple scattering effects. The solution is simplified when scattering matrices for each individual rough interface are pre-calculated. Based on the exact solution, the applicability of Kirchhoff approximation under various conditions is evaluated. The approximated method can then be applied to two-dimensional rough interfaces. [Work supported by ONR.]

10:50

4aUW8. Accurate imaging of the current and temperature structures in coastal oceans by acoustic data assimilation. Tokuo Yamamoto (Appl. Marine Phys. Div., RSMAS, Univ. of Miami, Miami, FL 33149) and Arata Kaneko (Hiroshima Univ., Hiroshima, Japan)

Both ocean observations and models have limitations. To eliminate the limitations, the idea of data assimilation has been tried in recent years but the results still have been very limited. We have made reciprocal transmissions of 2.5-kHz sound signals along six paths in the Japanese Inland Sea. An extensive ADCP survey was made during this acoustic transmission experiment for the ground truth. The acoustic data were directly inverted for imaging the current and temperature structures. The acoustic data were also assimilated into a Princeton ocean model (POM) using a Kalman filter. The acoustic data assimilation produced very accurate structures of current and temperature, while the direct inversions produced inaccurate and distorted structures. The time interval of data assimilation controls the accuracy of data assimilation technique.

11:05

4aUW9. M-ary frequency-shift-keying bit error rate analysis for a high-frequency underwater acoustic fading channel. Wen-Bin Yang and T. C. Yang (Naval Res. Lab., Code 7120, 4555 Overlook Ave., SW, Washington, DC, 20375, wyang@wave.nrl.navy.mil)

One of the prime advantages of M-ary frequency-shift-keying (MFSK) modulation is its ability to use an incoherent receiver. The incoherent receiver does not need precise frequency and phase estimates so that it can perform well, particularly in a significantly faded underwater acoustic communications channel. MFSK bit error rate (BER) performance using an incoherent receiver is determined by fading statistics of received signal amplitude. From our recent study in high-frequency underwater acoustic channel characterization, we found the fading statistics has a non-Rayleigh behavior and may be modeled as a K-distribution. In this paper, the MFSK BER is derived in terms of the parameters of the K-distribution. The theoretical BER is compared with measured BER of frequency-hopping (FH) FSK signaling data collected off the coast of New Jersey during the TREX04 experiment. The FH-FSK signal is a modified version of MFSK signals and was centered at 17 kHz with 4-kHz bandwidth. [Work supported by the Office of Naval Research.]

THURSDAY AFTERNOON, 8 JUNE 2006 ROOM 557, 1:00 TO 6:00 P.M.

Session 4pAAa

Architectural Acoustics, Musical Acoustics and Engineering Acoustics: Composed Spaces Loudspeaker Concert II (Performance Session)

Alexander U. Case, Chair

Fermata Audio + Acoustics, P.O. Box 1161, Portsmouth, NH 03802

Works of music and sound art that were presented as a concert in Session 4aAAa will be repeated in this session. No lecture papers will be presented. Please see session 4aAAa for a description of the compositions. Attendees from all disciplines are invited to attend.

THURSDAY AFTERNOON, 8 JUNE 2006 BALLROOM B/C, 1:30 TO 5:30 P.M.

Session 4pAAb

Architectural Acoustics: Spaces of Worship—Another Quarter Century of Experience (1984–Present) II (Poster Session)

Adam R. Foxwell, Chair

PMK Consultants, 1420 W. Mockingbird Ln., Suite 400, Dallas, TX 75247

Contributed Papers

All posters will be on display from 1:30 p.m. to 5:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:30 p.m. and contributors of even-numbered papers will be at their posters from 3:30 p.m. to 5:30 p.m.

4pAAb1. Church of the Epiphany, Miami, FL. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb2. Church of the Transfiguration, Orleans, MA. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb3. Friendship Missionary Baptist Church, Charlotte, NC. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb4. Community of Christ (formerly Reorganized Church of Latter Day Saints) Independence, MO. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb5. Joan and Stanley Golder Chapel, Temple Jeremiah, Northfield, IL. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb6. Rose-Hulman Institute of Technology—White Chapel, Terre Haute, IN. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb7. United Theology and Seminary—Bigelow Chapel, New Brighton, MN. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb8. First Baptist Church, Greensboro, NC. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb9. St. Timothy's Episcopal, Winston, Salem, NC. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb10. Trinity Church, Boston, MA. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb11. Hinsdale United Methodist Church, Hinsdale, IL. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb12. St. James Cathedral, Seattle, WA. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb13. Palm Beach Atlantic University Chapel, W. Palm Beach, FL. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb14. First United Methodist Church Cumming, Cumming, GA. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb15. St. Bede's Atlanta, GA. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb16. Redeemer Lutheran, Atlanta, GA. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb17. First Congregational Catholic Church, Battle Creek, MI. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb18. Camelback Bible Church, Paradise Valley, AZ. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb19. St. Turibius Chapel, Columbus, OH. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb20. Worthington Christian Church, Columbus, OH. R. Lawrence Kirkegaard (Kirkegaard Associates, 801 West Adams St., 8th Fl., Chicago, IL 60607)

4pAAb21. Second Presbyterian, Bloomington, IL. R. Lawrence Kirkegaard, FASA, Hon. (Kirkegaard Associates, 801 West Adams St., Chicago, IL 60607)

4pAAb22. Mt. Paran Church of God, Atlanta, GA. Ronald Eligator (Acoustic Dimensions, 145 Huguenot St., Ste. 406, New Rochelle, NY 10801)

4pAAb23. Riverside Church, New York, NY. Ronald Eligator (Acoustic Dimensions, 145 Huguenot St., Ste. 406, New Rochelle, NY 10801)

4pAAb24. St. Andrews Church, Dallas, TX. Ronald Eligator (Acoustic Dimensions, 145 Huguenot St., Ste. 406, New Rochelle, NY 10801)

4pAAb25. Willow Creek Church, S. Barrington, IL. Ronald Eligator (Acoustic Dimensions, 145 Huguenot St., Ste. 406, New Rochelle, NY 10801)

4pAAb26. Roseville Lutheran Church, Roseville, MN. Ronald Eligator (Acoustic Dimensions, 145 Huguenot St., Ste. 406, New Rochelle, NY 10801)

4pAAb27. Crossroads Community Church, Cincinnati OH. Adam Foxwell, David Marsh, Jerrold Stevens, and Melvin Saunders, IV (PMK Consultants, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

4pAAb28. Broadway Baptist Church, Fort Worth, TX. Adam Foxwell, David Marsh, Jerrold Stevens, and Melvin Saunders, IV (PMK Consultants, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

4pAAb29. Asbury United Methodist Church, Tulsa, OK. Adam Foxwell, David Marsh, Jerrold Stevens, and Melvin Saunders, IV (PMK Consultants, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

4pAAb30. Crown Center at Pensacola Christian College, Pensacola FL. Adam Foxwell, David Marsh, Jerrold Stevens, and Melvin Saunders, IV (PMK Consultants, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

4pAAb31. St. Ann Catholic Church, Coppell, TX. Adam Foxwell, David Marsh, Jerrold Stevens, and Melvin Saunders, IV (PMK Consultants, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

4pAAb32. North Richland Hills Baptist Church, North Richland Hills, TX. Adam Foxwell, David Marsh, Jerrold Stevens, and Melvin Saunders, IV (PMK Consultants, 1420 W. Mockingbird Ln., Ste. 400, Dallas, TX 75247)

4pAAb33. Parls Chapel Freewill Baptist Church, Fayetteville, NC. Joseph F. Bridger, Aaron Farbo (Stewart Acoustical Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607), and Edward Strickand (AVCON, Inc., Cary, NC 27519)

4pAAb34. St. Francis of Assisi Catholic Church, Raleigh, NC. Noral D. Stewart, Joseph F. Bridger (Stewart Acoustics Consultants, 7406 L Chapel Hill Rd, Raleigh, NC 27607), and James S. Brawley (James S. Brawley & Associates, Inc., Clemson, SC 29631)

4pAAb35. Immaculate Conception Catholic Church, Durham, NC. Joseph F. Bridger (Stewart Acoustical Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607)

4pAAb36. North Raleigh United Methodist Church, Raleigh, NC. Joseph F. Bridger and Aaron Farbo (Stewart Acoustical Consultants, 7406 L Chapel Hill Rd, Raleigh, NC 27607)

4pAAb37. First Free Will Baptist Church, Wilson, NC. Joseph F. Bridger (Stewart Acoustical Consultants, 7406 L Chapel Hill Rd, Raleigh, NC 27607) and Edward Strickland (AVCON, Inc., Cary, NC 27519)

4pAAb38. Wake Chapel Baptist Church, Raleigh, NC. Joseph F. Bridger, Mathew M. George (Stewart Acoustical Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607), Dan Wood (One Source, Winston-Salem, NC 27101), and James S. Brawley (James S. Brawley & Associates, Inc., Clemson, SC 29631)

4pAAb39. Temple Emanuel, Winston-Salem, NC. Joseph F. Bridger and Mathew M. George (Stewart Acoustical Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607)

4pAAb40. Unitarian Universalist Fellowship of Raleigh, Raleigh, NC. Joseph F. Bridger (Stewart Acoustical Consultants, 7406 L Chapel Hill Rd., Raleigh, NC 27607) and Joe Davidson (Davidson Audio Visual, Inc., Pfafftown, NC 27040)

Session 4pAAc

Architectural Acoustics: Acoustics of Libraries

Benjamin E. Markham, Chair
Acentech Inc., 33 Moulton St., Cambridge, MA 02138-1118

Chair's Introduction—1:00

Invited Papers

1:05

4pAAc1. Fine-tuning the acoustics of libraries at colleges and universities. Joyce E. Rosenbaum, Carol L. Krumhansl, and Michael J. Hove (Dept. of Phys., Cornell Univ., Ithaca, NY 14853, jer42@cornell.edu)

New technology and changing student work habits challenge academic libraries to accommodate different types of study by designing different rooms for distinct purposes. This requires not only a variety of technologies, room setups, and seating, but also acoustic conditions that stimulate efficient and comfortable productivity—whether with computers, textbooks, or interactive study groups. Acoustic recordings were taken in 13 different ''favorite'' rooms, with diverse characteristics and personalities, within the two main libraries at Cornell University. The acoustic measurements were compared with observations of student activity within these spaces to determine relationships between acoustic measurements and student behavior. Factor analysis revealed three types of student-activity categories: quiet concentration, relaxing and socializing, and computer use. The analysis also revealed three categories of room type: quiet study rooms, rooms with computer use, and cafés. These categories of activities and room types correlate with acoustic measurements, pointing to the existence of distinct niches within the library system. Examining the acoustic conditions conducive to certain activities may help future designers of academic libraries to optimize the acoustics of study spaces for different purposes.

1:25

4pAAc2. Rethinking the library for the 21st century? Peter Nobile (Shepley Bulfinch Richardson and Abbott, 2 Seaport Ln., Boston, MA 02210)

In the past few decades, the evolution of libraries has been increasingly dramatic and complex. Technology has fundamentally revolutionized information access, manipulation, and utilization. This transformation has challenged librarians and architects to evaluate the implications of these changes and their impact on the core mission of the library, its architecture, and the institution as a whole. Shepley Bulfinch is often engaged by our clients to help usher in the library for the 21st century. Such a facility is expected to fully integrate the digital and print worlds and introduce new spaces for media access, cyber cafs, 24/7 use, seminar and general instruction, faculty development labs, and GIS services. Additionally, such projects create flexible spaces that can be easily transformed to meet changing needs and uses. These projects bring with them significant acoustical isolation and technological access problems. However, such challenges are to be expected as these projects redefine the research and learning environment by integrating library resources, services, and users in spaces that reflect the overall evolution of the library.

1:45

4pAAc3. The digital preservation facility. Elizabeth Cohen (Consulting Professor of Elec. Eng., Stanford Univ., 161 Packard Bldg., 350 Serra Mall, Stanford, CA 94305)

Critical listening should be an essential part of all archiving and restoration facilities quality control. We review the priorities and requirements for listening spaces ranging from the individual collector and small community archives to large-scale facilities. Examples discussed include the Library of Congress Culpepper facility, university libraries, and commercial facilities. Adapting listening rooms to the requirements of n-channel audio are discussed. Public recommendations of the Sound Preservation Board of the Library of Congress will be reviewed.

2:05

4pAAc4. Acoustical design issues for library facility planners. Cristina L. Miyar, Elaine Y. Hsieh, and Charles M. Salter (Charles M. Salter Associates, Inc., 130 Sutter St., 5th Fl., San Francisco, CA 94104, cristina.miyar@cmsalter.com)

Acting under the provisions of the Library Services and Technology Act, the U.S. Institute of Museum and Library Services retained Charles M. Salter Associates, Inc. to provide a document explaining acoustical and audio/visual requirements for libraries to library planners, architects, and lay people. Acoustical design issues for libraries discussed on the website include site noise considerations, noise standards for each use space, room acoustics considerations, sound isolation between various use spaces, vibration control for mechanical equipment, and audio/visual system considerations. The completed document forms a part of the Libris Design library facility planning information system, which includes a website, <Librisdesign.org>, with recent information on facility planning topics and a database of recently constructed California public libraries.

2:25

4pAAc5. Wireless technology and library acoustics. Benjamin Markham (Acentech Inc., 33 Moulton St., Cambridge, MA 02138, bmarkham@acentech.com)

The modern library cannot function without modern technology. Printers, copiers, wireless computers, and espresso machines all contribute to the soundscape of the typical municipal library (and many academic libraries, too). While some of these noise sources are stationary and can be isolated acoustically, wireless computing has transformed the acoustic experience for the typical library patron. Study carrels and desks used to define the boundaries of study and work areas in libraries, and architects could design spaces accordingly. With wireless technology, every space is a study space. This paper will discuss the implications of wireless technology on library acoustics. Measurement results at a variety of case study libraries will be presented.

2:45

4pAAc6. Modern library design—architectural directions. Isaac Franco (Moshe Safdie and Associates, 100 Properzi Way, Somerville, MA 02143) and Carl Rosenberg (Acentech, Inc., Cambridge, MA 02138)

The architectural practice of Moshe Safdie and Associates in Boston has been challenged by numerous clients recently to design major public libraries. The clients have stated an interest in making the library a major civic focal point and a place for community activities and identity. Moshe Safdie has responded to these challenges in Philadelphia, Vancouver, and Salt Lake City by designing buildings that encourage community interaction with large light-filled gathering spaces to support public congregating venues, shops and eateries, music recitals, meetings, electronic technologies, and even book club gatherings. This presentation will discuss the architectural context of contemporary library design. As a coda to the presentation, there will be a discussion about the acoustical implications of these new directions in civic library design.

THURSDAY AFTERNOON, 8 JUNE 2006 ROOM 552B, 1:00 TO 5:05 P.M.

Session 4pAO

Acoustical Oceanography and Animal Bioacoustics: Exploitation of Sound by Marine Mammals II

David L. Bradley, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Whitlow W. L. Au, Cochair
Hawaii Inst. of Marine Biology, P.O. Box 1106, Kailua, HI 96734

Chair's Introduction—1:00

Invited Papers

1:05

4pAO1. Anthropogenic noise and baleen whales: A perspective from humpback whale acoustics. Whitlow W. L. Au and Marc O. Lammers (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., Univ. of Hawaii, P.O. Box 1106, Kailua, HI 96734)

The effects of anthropogenic noise on baleen whales is a topic of intense interest not only to marine mammalogists but also to a variety of commercial interests, the military, oceanographers, and researchers that use sounds in the ocean. One of the major difficulties in setting standards for exposure levels and the effective use of anthropogenic sounds is a lack of knowledge of the auditory properties of whales. The auditory sensitivity and frequency range of hearing (audiogram) are not known, as well as how anthropogenic sounds physically affect the auditory system, the psychological well being, and the use of sounds by whales. One approach to gain some insight in this area is to study the acoustics of a baleen whale species, the humpback whale. The source levels and frequency range of humpback whale song units have been measured and can provide insights into what these whales expose each other to. The highest rms source level was 173 dB for a specific unit. The average sound pressure levels of chorusing have been measured and can provide insights into the long-term acoustic environment of humpback whales. The highest daily average of sound pressure level for chorusing sounds was 130 dB.

1:25

4pAO2. Right whales—Social sound production. Susan E. Parks (Bioacoust. Res. Program, Cornell Univ. 159 Sapsucker Woods Rd., Ithaca, NY 14850, sep6@cornell.edu), Douglas P. Nowacek (Florida State Univ., Tallahassee, FL 32306-4320), Mark P. Johnson, and Peter L. Tyack (Woods Hole Oceanograph. Inst., Woods Hole, MA 02543)

Right whales use sound to mediate many social interactions. Examples include females producing screams to initiate the formation of surface active groups, adult males producing gunshot displays in agonistic contexts, and mothers producing contact calls when reuniting with their calves. We are just beginning to understand the more sophisticated abilities of right whales to utilize sound. Playback experiments demonstrate that right whales are capable of precise sound localization over long ranges and can remember these locations after the sound source ends. Further evidence suggests that individual recognition may play an important role in

acoustic communication, from the simple examples from mothers and calves, to more complex evidence that some males preferentially approached playbacks of calls from specific females. Future research investigating the complexity of baleen whale sound use will include studies of individual recognition and the localization and spatial memory abilities of these whales. New technological developments are making it possible to answer these more complex questions and will reveal more details of how baleen whales to use sound to survive.

1:45

4pAO3. Sound production and detection by pinnipeds. Brandon L. Southall (NOAA Ocean Acoust. Program, NMFS Office of Sci. and Technol., 1315 East-West Hwy., Silver Spring, MD 20910), David Kastak, Colleen Reichmuth Kastak, Steven J. Insley, Marla M. Holt, Jason Mulsow, and Ronald J. Schusterman (Univ. of California, Santa Cruz CA 95060)

Pinnipeds (seals, sea lions, and walruses) emit and receive sounds both in air and water. Their amphibious natural history has resulted in a number of fascinating sensory capabilities, including but not limited to hearing, but has apparently precluded the evolution of sophisticated biosonar systems. Pinnipeds produce vocalizations in a variety of social contexts, notably reproduction and aggression. Attraction calls in many species serve to retain or re-establish contact between females and offspring and to draw the attention of potential mates. Repulsion vocalizations warn adversaries and can advertise strength and intent. Cognitive processes, including conditioned learning, are an essential foundation of the sophisticated acoustic communication systems of pinnipeds. Recent advances in understanding these systems and passive listening in the context of predator and prey detection and spatial orientation will be discussed.

2:05

4pAO4. Imaging acoustic centers of the dolphin brain with positron emission tomography. Sam Ridgway, Dorian Houser, Don Carder (SPAWAR Systems Ctr. San Diego, Div. D235, 53560 Hull St., San Diego, CA 92152-5001), and Carl Hoh (Univ. of California, San Diego, CA 92093)

In human subjects, positron emission tomography (PET) has been used in diagnosis of disease, in studies of complex cognitive tasks, and in localizing areas of sensory activation, motor movement, and attention. Now, brain physiology of bottlenose dolphins, *Tursiops truncatus*, has been studied with PET. The dolphin brain is in the human size range but is more highly convoluted. The known acoustic centers of the dolphin brain are greatly enlarged compared with human brains. However, dolphins have a multi-tiered cortex with an extensive extra lobe, the paralimbic lobe, which does not occur in mammals or other animals outside the Order Cetacea. Because of its inaccessible location, there are no functional studies of the paralimbic cortex. PET offers an opportunity to study activity in this unique part of the dolphin brain during various stimulus conditions that may help to reveal functional aspects of the paralimbic cortex. For experimental scans, the dolphin was exposed to acoustic stimuli (tones or pulses) to reveal areas of relative brain metabolism after 18-F-FDG administration. These scans have revealed interesting features of the living dolphin brain including the paralimbic lobe as well as known acoustic areas in the cortex, midbrain, and acoustic nuclei.

Contributed Papers

2:25

4pAO5. Dolphin pods of the Southern California Offshore Range: Localization and behavior. Katherine H. Kim, Paul Hursky, Michael B. Porter (Heat, Light, and Sound Research, Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130, katherine.kim@hlsresearch.com), John A. Hildebrand, E. Elizabeth Henderson, and Sean M. Wiggins (Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0205)

The Southern California Offshore Range, a U.S. Navy training area located near San Clemente Island, is rich in odontocete activity. Pods of dolphins, numbering as many as several hundred individuals, regularly transited the area and frequently vocalized during the 2004 SCORE 23F experiment. Such acoustic data were collected on a high-frequency, four-element, horizontal line array as well as on a high-frequency, single-element buoy. The copious dolphin vocalizations were then processed using automated whistle and click detection techniques and various localization methods which utilized time-of-arrival differences, frequency-domain beamforming, and multipath arrival structure. Combining the acoustic tracking results with concurrent visual observations provides a unique look at large numbers of free-ranging dolphins at and below the sea surface and suggests certain group behavioral patterns.

2:40

4pAO6. Depth/range localization of diving sperm whales using passive acoustics on a single hydrophone disregarding seafloor reflections. Christophe Laplanche, Olivier Adam, Maciej Lopatka, and Jean-François Motsch (Université Paris 12, 61 av. du Gal de Gaulle, 94000 Créteil, France)

Depth and range of sound sources can be estimated using a single hydrophone. Such a passive acoustic technique requires the detection of direct path transmitted, sea surface, and seafloor reflected source signals, so as to measure their time of arrival differences (TOADs). Sperm whales almost continuously emit powerful, directional echolocation sounds (usual clicks) when diving. Sperm whales often dive in deep water, and click seafloor reflections are usually well detected only at the beginning of the dive. Surface echoes may be detected during the entire dive. If the measurement of the surface reflection TOAD of a single click is not enough for estimating the depth/range of the sperm whale at the time when this click was emitted, the joint consideration of delays emitted during the whole dive may provide this estimation. Such delays are indeed the measurements of a single phenomenon: the underwater movements of a single-clicking sperm whale. One can merge such data using a Bayesian technique, as well as use prior information (e.g., range when fluking-up). Such

a method could make the single-hydrophone localization technique usable regardless of the properties of the sea bottom, e.g., in deep or high relief areas.

2:55–3:05 Break

3:05

4pAO7. Linking sound production and dive behavior in feeding humpback whales. Alison K. Stimpert, Whitlow W. L. Au (MMRP/Hawaii Inst. of Marine Biol., P.O. Box 1106, Kailua, HI 96734, stimpert@hawaii.edu), David N. Wiley (Stellwagen Bank Natl. Marine Sanctuary, Scituate, MA 02066), Kenneth A. Shorter, Kira L. Barton, Mark P. Johnson (Woods Hole Oceanogr. Inst., Woods Hole, MA, 02543), Colin Ware, and Roland Arsenault (Univ. of New Hampshire, Durham, NH, 03824)

Acoustic studies of baleen whales are becoming increasingly common. However, a minority of studies combines acoustic data with technologies that allow sound production to be placed in a detailed behavioral context. Noninvasive digital acoustic recording tags (DTAGS) were attached to humpback whales (*Megaptera novaeangliae*) on the western North Atlantic's Great South Channel feeding grounds to study foraging and acoustic behavior. Acoustic records totaling 48.4 data hours from four attachments were aurally and automatically analyzed, and occurrences of several sound types were correlated with body orientation and dive behavior of the tagged subject. Whales produced a wide variety of sound types, which differed in depth and context of production. Long, tonal groans and moans (3-s average duration, 500-Hz peak frequency) occurred more frequently in upper portions of the water column, in contrast to broadband paired bursts (50-ms average duration, 300-Hz peak frequency), which appeared associated with bottom feeding. Virtual behavioral study using GEOZUI4D is used to visualize how sounds may be associated with foraging behavior. Overall results indicate sophisticated, situation-dependent use of sound. Behavioral implications of masking effects by surrounding vessel traffic noise are discussed.

3:20

4pAO8. Naval sonar, marine mammal strandings, and the bubble hypothesis. Mark W. Muller (Marine Mammal Res. Program, Hawaii Inst. of Marine Biol., 46-007 Lilipuna Rd., Kaneohe, HI 96744 and Dept. of Mech. Eng., Univ. of Hawaii, 2540 Dole St., Honolulu, HI 96822), Whitlow W. L. Au (Hawaii Inst. of Marine Biol., Kaneohe, HI 96744), John S. Allen (Univ. of Hawaii, Honolulu, HI 96822), Eryn M. Wezensky (Naval Undersea Warfare Ctr. Newport, RI 02841), and Michael R. Barkley (Bates Creative Group, Silver Spring, MD 20910)

Although spatio–temporal links exist between some military active sonar testing and cetacean mass stranding events, the underlying causal mechanism(s) of the strandings are not well understood. It has been demonstrated that cetaceans may experience *in vivo* gas bubble/emboli development; however, the basis for cavitation formation remains unknown. Acoustically mediated *in vivo* cavitation formation in cetaceans exposed to high-intensity sound sources has been proposed. The distance of these exposure effects from midfrequency sound sources can be influenced by a combination of factors affecting underwater acoustic propagation. To investigate these issues, this experiment considered rectified diffusion as a possible mechanism for acoustically mediated *in vivo* cavitation formation in cetaceans. Gelatin samples were placed under water and exposed to an SPL of 205 dB *re*: 1 μ Pa at 4.5 kHz for up to 1000 s. The sound field was generated by a Massa transducer similar to the transducer element used in the United States Navy AN/SQS–23 shipboard high–power sonar system. Exposure time threshold of cavitation formation was determined to be 600 s for the given sound field. Experimental results were compared with previously developed theoretical models for acoustically enhanced cavitation growth. The applicability and limitations of this approach as well as future research direction are discussed.

3:35

4pAO9. Acoustic analysis for the Haro Strait on 5 May 2003. David M. Fromm (Naval Res. Lab., 4555 Overlook Ave. SW, Washington, DC 20375-5320)

On 5 May 2003, USS SHOUP conducted sonar training for 4 h while transiting through the Strait of Juan de Fuca and the Haro Strait. During the time period when the sonar was configured for long-range operations, the transmissions were alleged by some observers to have influenced the behavior of the resident J-pod of orcas. Additionally, the transmissions were received and recorded on various hydrophones deployed for monitoring marine mammal vocalizations. The U.S.-Navy-commissioned detailed reconstruction of the event that relates the location of SHOUP, the J-pod, and the monitoring hydrophones during the time period of SHOUP's long-range sonar operations is presented. By combining high-fidelity predictions of the acoustic field with the *in situ* acoustic recordings, the reconstruction provides a detailed, moment-to-moment picture of the worst case estimates of the sonar sound levels experienced by the J-pod. [Work supported by the U.S. Navy Pacific Fleet.]

3:50

4pAO10. Are manatees modifying their vocalizations to compensate for noise? Jennifer L. Miksis-Olds (Univ. of Massachusetts SMAST, 706 South Rodney French Blvd., New Bedford, MA 02744, jmiksis@umassd.edu)

Elevated environmental noise levels have the potential to interfere with the acoustic communication system of manatees by masking signals containing biologically important information. Communication theory recognizes several ways a sender can modify its acoustic signal to compensate for noise. These include increasing the source level of a signal, its repetition, its duration, or shifting frequency outside of the noise band. To test whether manatees can modify their vocalizations to compensate for noise, vocalization usage and structure were examined in terms of vocalization rate, duration, frequency, and source level. Results imply that manatees do increase their vocalization energy expenditure, or vocalization effort, as a function of behavior and calf presence. Increases in vocalization rate, duration, and source level were most pronounced when calves were present and during behaviors whereby animals tended to be more dispersed, which may suggest a cohesion function. In conditions of elevated noise levels, manatees increased call duration during feeding and milling behaviors when calves were present, suggesting that ambient noise levels do have a detectable effect on manatee communication and that manatees modify their vocalizations as a function of noise.

4:05

4pAO11. Wavelet-based denoising of manatee vocalizations. Christopher Niezrecki and Berke M. Gur (Dept. of Mech. Eng., Univ. of Massachusetts Lowell, One University Ave., Lowell, MA 01854, christopher_niezrecki@uml.edu)

In the last decade, novel research on the wavelet transform has led to its recognition as a promising tool for denoising signals. Wavelet-coefficient-shrinkage-based denoising algorithms have been applied in various fields of science and engineering and successful results have been reported in Raman spectroscopy and medical imagery. Within this study, the wavelet transform is used to denoise manatee vocalization recordings corrupted with typical underwater noise such as boat, snapping shrimp, and water movement noise. The detection range and performance of acoustically based detection systems depends greatly on the SNR of the acoustic signal. A selective thresholding is performed on the wavelet coefficients and the denoised vocalizations are obtained through the reconstruction of these coefficients. The performance of the method presented is benchmarked against the well-known wavelet-coefficient-shrinkage-based denoising algorithms using the same underwater acoustic data obtained in prior studies. Preliminary results indicate that the suggested wavelet-transform-based denoising algorithm outperforms the shrinkage-based methods by an average of 5 dB.

4:20

4pAO12. Locating manatee position with an acoustic array. Paulin Buaka Muanke and Christopher Niezrecki (Dept. of Mech. Eng., Univ. of Massachusetts Lowell, One University Ave., Lowell, MA 01854, christopher_niezrecki@uml.edu)

This paper presents the effort conducted in determining a manatee's position by using acoustic methods. The time difference of arrival (TDOA) technique developed for the position location is used with the Hilbert envelope peak technique to estimate the time difference of arrival. The time differences obtained are converted into nonlinear hyperbolic equations which are solved using the least square method to perform the manatee localization. To improve the position location accuracy, with four hydrophones, six pairs of time delays are estimated and then combined (three by three) to give 20 calculated locations for each individual vocalization. To get a more accurate location, a statistical analysis test to remove the time delays combinations, which generate outliers, is applied on these 20 location results. Simulated results using the dSpace system for data acquisition and signal processing are presented for the case of an artificial tone signal. This method is applied to a variety of experimental data to test the systems performance.

4:35

4pAO13. Background noise reduction of manatee vocalizations. Zheng Yan, Louis N. Cattafesta, III (Dept. of Mech. and Aerosp. Eng., Univ. of Florida, Gainesville, FL 32611-6250), Christopher Niezrecki (Univ. of Massachusetts Lowell, Lowell, MA 01854), and Diedrich O. Beusse (Univ. of Florida, Gainesville, FL, 32610-0126)

The West Indian manatee (*Trichechus manatus latirostris*) has become an endangered species in the United States partly because of an increase in the number of collisions with boats. A device to alert boaters of the presence of manatees is desired. Previous research has shown that background noise limits the detection range (which is critical for practical implementation). By improving the signal-to-noise ratio of the vocalization signal, it is possible to extend the detection range. The finite impulse response (FIR) structure of the adaptive line enhancer (ALE) can detect and track narrowband signals buried in broadband noise. In this paper, a constrained infinite impulse response (IIR) ALE, called a feedback ALE (FALE), is implemented to reduce the background noise. In addition, a bandpass filter is used as a baseline for comparison. A library consisting of 100 manatee calls spanning ten different signal categories is used to evaluate the performance of the bandpass filter, FIR-ALE, and FALE. The results show that the FALE is capable of reducing background noise by about 6.0 and 21.4 dB better than that of the FIR-ALE and bandpass filter, respectively, when the signal-to-noise ratio (SNR) of the original manatee call is -5 dB.

4:50

4pAO14. Simulation of ranges for acoustic-based manatee detection. Christopher Niezrecki, Berke M. Gur (Dept. of Mech. Eng., Univ. of Massachusetts Lowell, Lowell, MA 01854, christopher_niezrecki@uml.edu), James Cramer (Univ. of Massachusetts Lowell, Lowell, MA, 01854), and Diedrich O. Beusse (Univ. of Florida, Gainesville, FL 32610-0126)

One aspect of the feasibility of an acoustically based manatee detection system relies upon the distance in which a manatee vocalization is detectable. The magnitude of environmental noise, manatee vocalizations, as well as the acoustic spreading properties of the habitat need to be considered to estimate the detection range of a vocalizing manatee. The simulation presented combines measured source levels of manatee vocalizations with the modeled acoustic properties of manatee habitats to illustrate the effective ranges in which manatees are detectable. The simulation indicates that background noise significantly affects the detectable range.

THURSDAY AFTERNOON, 8 JUNE 2006 BALLROOM E, 1:00 TO 5:10 P.M.

Session 4pBB

Biomedical Ultrasound/Bioresponse to Vibration and Physical Acoustics: Celebration Session for Edwin Carstensen II

Frederick W. Kremkau, Cochair
Wake Forest Univ., School of Medicine, Ultrasound Ctr., Medical Center Blvd., Winston-Salem, NC 27157-1039

Leon A. Frizzell, Cochair
Univ. of Illinois, Bioacoustics Research Lab., 405 N. Mathews Ave., Urbana, IL 61801

Invited Papers

1:00

4pBB1. The role of cavitation in therapeutic ultrasound. L. Crum, M. Bailey, V. Khokhlova, O. Sapozhnikov (Ctr. for Industrial and Med. Ultrasound, Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105), B. Rabkin (Univ. of Washington, Seattle, WA 98105), A. Evan, J. McAteer, Y. Pishchalnikov, J. Williams (Indiana School of Medicine, Indianapolis, IN), and R. Cleveland (Boston Univ., Boston, MA 02215)

Ed Carstensen has made many contributions to biomedical ultrasound but among those that are becoming more and more relevant to current clinical practice are those that determine the conditions under which cavitation is induced *in vivo*. For many years, it was assumed that the medical ultrasound devices were unable to induce cavitation in living tissue because either the acoustic conditions were not sufficient or the nucleation sites that are required were too small. With the advent of lithotripters and high-intensity focused ultrasound (HIFU) devices, cavitation generation *in vivo* is commonplace. Our current research at the University of Washington has focused on the role that cavitation plays in stone comminution and tissue damage during lithotripsy, as well as the enhancement or reduction of desirable coagulative necrosis during HIFU application. During HIFU application, we find enhanced heating that results

from nonlinear acoustic wave propagation (a key Carstensen contribution) leads to vapor bubble formation. This presentation will review our recent studies in this area. [Work supported in part by the NIH under CA109557 and DK43881, and the NSBRI SMS00402 under NCC9-58-50.]

1:20

4pBB2. Vessel damage mechanisms by ultrasound or shock wave pulses. Sheryl M. Gracewski, Hongyu Miao (Dept. of Mech. Eng. and the Ctr. for Biomed. Ultrasound, Univ. of Rochester, Rochester, NY 14627, grace@me.rochester.edu), and Diane Dalecki (Univ. of Rochester, Rochester, NY 14627)

Edwin L. Carstensen has been one of the pioneers investigating not only the thresholds for bioeffects of ultrasound, but also determining mechanisms for these bioeffects, especially when air cavities are involved. Previous studies have demonstrated that hemorrhage or endothelial layer damage can occur when the blood vessels are exposed to lithotripter shock waves or to ultrasound in the presence of echo contrast agents. The presence of bubbles is often necessary for the damage to occur. Two main hypotheses describing the mechanism of vessel damage by acoustically excited microbubbles have been proposed. The vessel could be damaged by the high pressures and/or temperatures generated by the violent collapse of cavitation bubbles. Alternatively, the vessel could rupture upon bubble expansion. Here, a simple model is presented for vessel rupture upon bubble expansion and calculations are presented for parameters corresponding to two recently published experimental observations of vessel rupture and of endothelial cell damage to try to distinguish between these two mechanisms.

1:40

4pBB3. Risk factors for tissue injury induced by shock wave lithotripsy (SWL). Andrew P. Evan, Lynn R. Willis, Bret A. Connors, Philip M. Blomgren (School of Medicine, Indiana Univ., Indianapolis, IN 46223), Rajash Handa, Youshi Shao, James A. McAteer, Yuri A. Pishchalnikov (Indiana Univ., Indianapolis, IN 46223), and James E. Lingeman (Methodist Hospital Inst. for Kidney Stone Disease, Indianapolis, IN 46220)

A clinical dose of SWL consistently and specifically disrupts the cellular layers of renal veins and arteries, first in the medulla, but progressing into the cortex with an associated inflammatory response. The resulting hemorrhage within the focal zone produces regions of intraparenchymal and subcapsular hematomas of varying size. Renal blood flow and glomerular filtration rate fall acutely by 50% in the treated kidney and by 20% in the untreated kidney. The sites of acute injury form scars with a permanent loss of renal function and have been associated with apatite/brushite stone formation and hypertension. Our team has shown that the size of the SWL-induced lesion can be dramatically reduced by lowering the kV level, applying fewer shock waves, slowing the shock wave rate, and by pretreating the kidney with 100–500 shock waves at 12 kV prior to standard SWL. The mechanism of tissue injury caused by SWL has not been defined; however, protocols that reduce cavitation activity also reduce the size of the SWL-induced lesion, and injection of bubbles into the renal vasculature, which support cavitation, intensifies and broadens the regions of the SWL-induced injury.

2:00

4pBB4. A possible noncavitational mechanism of ultrasound-accelerated thrombolysis in fibrin clots. E. Carr Everbach (Eng. Dept., Swarthmore College, Swarthmore, PA 19081), Irina N. Chernysh, and John W. Weisel (Univ. of Pennsylvania, Philadelphia, PA 19104)

Blood clots made of human plasma or purified fibrin with and without the addition of rt-PA were insonified with 1-MHz pulsed ultrasound on a microscope slide placed in a spectrophotometer. Clot dissolution times were measured by absorption of light at 350-nm wavelength passing through the clot. Dissolution times were halved for ultrasound-exposed clots versus shams. Clots with no rt-PA present showed a small but significant thrombolysis effect in the presence of ultrasound which could be prevented by the addition of factor XIII, which effectively binds the fibrin. Ultrasound pulse parameter variations showed a broad maximum effectiveness at 0.5-MPa peak negative pressure, 1-ms duration, and 50-Hz PRF. Concentration of protein in solution was elevated as much as a factor of 10 following insonification of clots without rt-PA, suggesting that ultrasound affects the kinetics of fibrin remodeling. A possible explanation that does not involve bubble activity is relative motion between the clot fibers and the interstitial fluid.

2:20

4pBB5. Nonviral gene transfer to murine liver *in vivo* is enhanced by exposure to high-amplitude, 1-MHz ultrasound with Optison®. Andrew A. Brayman (Appl. Phys. Lab., Univ. of Washington, 1013 NE 40th St., Seattle, WA 98105-6698), Zhenping Shen, Liping Chen (Children's Hospital and Regional Medical Ctr., Seattle, WA 98109), and Carol Miao (Univ. of Washington, Seattle, WA 98195)

High-amplitude, 1-MHz ultrasound applied in pulse or burst mode at low duty factor, in conjunction with the injection of a mixture of gas-filled microspheres (Optison®) and plasmid, enhances nonviral gene transfer to surgically exposed murine liver *in vivo*. In this system, neither plasmid + ultrasound, nor plasmid + Optison produced significant gene expression relative to plasmid-only controls. The combination of plasmid + ultrasound + Optison produced a large enhancement of gene expression. The effect has a strong dependence on acoustic pressure. Together with the dependence of the effect on gas bodies, these results suggest that cavitation-induced permeability changes to hepatocytes are involved. Parametrization studies, the goal of which has been to further enhance gene transfer, will be discussed. At this writing, ultrasound treatments comprised by relatively brief pulses (20–40 cycles) at relatively high pulse repetition frequency (50 Hz) have produced nearly an order of magnitude greater gene expression than produced by tone bursts of several hundred cycles at low pulse repetition frequency (3 Hz) when plasmid and Optison are injected into the portal vein. [Project supported by NIH Grant R01 HL69049.]

2:40

4pBB6. Measuring bubble dynamics in lithotripsy using a high-speed camera. Parag V. Chitnis and Robin O. Cleveland (Dept of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, pchitnis@bu.edu)

The primary goal of this study was to examine bubble dynamics in shock wave lithotripsy and experimentally obtain the radius-time (r-t) profile. A piezoelectric array was employed as a shock wave (SW) source. The focal waveform (as measured with a fiber optic probe hydrophone) was double-peaked with two shocks: leading peak pressure 65 ± 3 MPa and second peak 30 ± 3 MPa. The negative phase in between had a peak $p-=-16\pm2$ MPa. The cavitation field was monitored using a dual passive cavitation detector consisting of two 1-MHz focused transducers. The second peak of the SW inhibited cavitation limiting the characteristic time to less than 60 μs. A high-speed camera (up to 10 million frames per second) with a spatial resolution of 2 μm/pixel was used to measure bubble dynamics. The camera could record 14 frames that were timed to measure the growth and collapse phases of the bubbles. The bubbles were initially spherical but became aspherical and exhibited jetting as they collapsed. The bubble radius was estimated from the area of bubble in each snapshot from which r-t curves were constructed and compared to calculations using the Gilmore formulation. The maximum radius ranged from 150 to 300 μm. [Work supported by NIH.]

2:55

4pBB7. Cavitation-induced saturation of the negative-pressure phase of lithotripter shock pulses. Yuri A. Pishchalnikov, James A. McAteer (Dept. of Anatomy and Cell Biol., Indiana Univ., School of Medicine, 635 Barnhill Dr., Indianapolis, IN, 46202-5120, yura@anatomy.iupui.edu), Michael R. Bailey (Univ. of Washington, Seattle, WA 98105-6698), Oleg A. Sapozhnikov (Moscow State Univ., Moscow, Russia, 119899), Irina V. Pishchalnikova, and James C. Williams, Jr. (Indiana Univ., School of Medicine, Indianapolis, IN, 46202-5120)

The tensile stress imposed by the negative-pressure phase of lithotripter shock pulses can cause cavitation. Bubbles continue to grow after the passage of the acoustic pulse; thus, some of the pulse energy is transformed to the kinetic and potential energy of the liquid surrounding the cavitation bubbles and, therefore, no longer belongs to the acoustic field. One might predict that this energy loss should be more pronounced for strong pulses that produce more cavitation. To investigate this, acoustic pulses were measured at the geometric focus of a Dornier HM-3 electrohydraulic lithotripter (water 39°C, dissolved gas ~8% saturation) using a fiber-optic probe hydrophone FOPH-500. Measurements showed that, while the amplitude and duration of the leading positive-pressure phase increased dramatically as charging potential was increased from 12 to 24 kV, the trailing negative-pressure phase of the pulse remained unchanged. This stabilization of the negative-pressure phase could be due to cavitation restricting the amplitude of the negative pressure that can be transmitted through the liquid, such that further increase of the amplitude at the source would not increase the negative amplitude at the target but would only result in stronger cavitation along the acoustic path. [Work supported by NIH-DK43881.]

3:10–3:25 Break

3:25

4pBB8. Mapping stone breakage to the acoustic field of the low-amplitude, broad-focal-zone lithotripter. Yuri A. Pishchalnikov, James A. McAteer, R. Jason VonDerHaar, Irina V. Pishchalnikova, and James C. Williams, Jr. (Dept. of Anatomy and Cell Biol., Indiana Univ., School of Medicine, 635 Barnhill Dr., Indianapolis, IN 46202, yura@anatomy.iupui.edu)

Recent reports describe the design, initial clinical experience, and mechanisms of stone breakage of a wide-focal-zone, low-pressure, self-focusing lithotripter [UMB **27**, 638 (2001); **28**, 769 (2002)]. We undertook mapping of the acoustic field of this device (Xixin lithotripter) using a fiber-optic hydrophone FOPH-500, and looked for correlation between features of the shock pulse, and breakage of gypsum stones at different locations within the field. The efficiency of stone breakage did not correlate with peak positive pressure, shock rise time, duration of the pulse, or positive area under the waveform, but did correlate with energy density and peak negative pressure. Breakage was significantly better at the target point of the imaging system and along the axis 20 mm proximal and distal to this point compared to the geometric focus, located ~4 cm distally. Measured dimensions of focal width (~15 mm) differed only slightly from the published value (18 mm), and stones up to 20 mm off-axis could be broken to completion although it took half as many pulses for breakage at 20 mm than at 15 mm. Thus, stone breakage was most efficient prefocally, and there appears to be considerable latitude for targeting with effective breakage both on-axis and lateral to the acoustic axis with this lithotripter. [Work supported by NIH-DK43881.]

3:40

4pBB9. Customization of acoustic output in lithotripsy using a piezoelectric transducer array. Parag V. Chitnis and Robin O. Cleveland (Dept of Aerosp. and Mech. Eng., Boston Univ., Boston, MA 02215, pchitnis@bu.edu)

A piezoelectric array was employed to examine the role of waveform shaping on stone fragmentation in lithotripsy. The array consisted of 170 elements placed on the inner surface of a 15-cm-radius spherical cap. Each element was driven independently using 170 individual pulsers—each capable of generating 1.2 kV. The acoustic field was characterized using a fiber optic probe hydrophone with a bandwidth of 30 MHz and a spatial resolution of 100 μm. When all elements were driven simultaneously, the focal waveform was a shock wave with peak pressures $p+=65\pm3$ MPa and $p-=-16\pm2$ MPa and the focal region was cigar shaped: 13 mm long and 2 mm wide. The firing of group of elements was adjusted so that each groups focused to different locations in the geometric focal plane. This method allowed the width of the focal region to be broadened albeit with decrease in peak pressure: for a 3-mm-wide region $p+=32$ MPa and for a 4-mm focal region $p+=25$ MPa. Waveform shaping was used to enhance or suppress cavitation effects where cavitation activity was observed using a dual passive cavitation detector consisting of two 1-MHz focused transducers. Stone fragmentation was assessed by treating artificial stones with fixed number of shock waves. [Work supported by NIH.]

3:55

4pBB10. Experimental investigation of linear resonant bubble behavior between parallel plates. Theodore F. Argo, IV and Preston S. Wilson (Appl. Res. Labs., Univ. of Texas, P.O. Box 8029, Austin, TX 78713-8029)

Recent findings suggest that bubble clouds play an active role in the breakup of kidney stones in patients undergoing shock-wave lithotripsy treatment. Understanding the dynamic behavior of the bubble cloud and its component bubbles within confined spaces will increase the efficacy of the treatment, but to date, even idealized model systems are not well understood. The behavior of a single bubble near a rigid plane has been verified; the study of this configuration is one of a bubble and a single bubble image source. A model recently set forth [Cui *et al.*, J. Acoust. Soc. Am. **119** (in press)] gives a description of the dynamics of a bubble between rigid parallel planes by using an infinite string of bubble image sources. This represents a step toward understanding the dynamics of fully enclosed bubble clouds. Experimental measurements of the resonance frequency and quality factor of air bubbles suspended between two parallel stainless-steel plates are presented and compared to the new model. An acoustic spectral subtraction technique was used to isolate the bubble response from that of the experimental tank. The bubbles were independently sized using diffuse backlighting and video microscopy. [Work supported by ARL:UT.]

4:10

4pBB11. Measurement of pressure produced at a solid surface by an acoustically driven cavitation bubble cloud. Teiichiro Ikeda, Shin Yoshizawa, Yoichiro Matsumoto (Dept of Mech. Eng., The Univ. of Tokyo, 113-8656 Tokyo, Japan), Michael R. Bailey, Lawrence A. Crum (Univ. of Washington, Seatte WA, 98105), and John S. Allen (Univ. of Hawaii at Manoa, Honolulu, HI 96822)

A cavitation bubble cloud typically acts as a strong acoustic scatterer; however, at certain frequencies and amplitudes waves transmitted by a cloud are amplified. A long-term goal is to break up kidney stones by using a two-frequency ultrasound forcing method with one frequency that generates the cloud followed by a lower second frequency that excites its violent collapse. The goal of this study is to determine the frequency and pressure amplitudes which produce the largest pressures as measured by a PVDF membrane on the stone surface. Reflection was also measured by a concave PVDF sensor placed 40 mm from the surface, and cloud sizes were determined by high-speed camera images. Transmission is quantified by the force. Reflection and transmission showed a reciprocal relation: peak in transmission corresponded to a minimum in reflection. The largest cloud observed created the largest reflection, whereas the smallest clouds created the largest transmitted signals. Forces generated by the small clouds were five times larger than the amplitude without a cloud. Thus, in using a two-frequency excitation combination, the pressures generated by the cloud cavitation might be optimized for lithotripsy applications. [Work supported in part by NIH DK43881 and NSBRI SMS00402.]

4:25

4pBB12. Numerical study on nonlinear behavior of a collapsing bubble cloud in an ultrasound field. Shin Yoshizawa, Teiichiro Ikeda, Shu Takagi, and Yoichiro Matsumoto (Dept. of Mech. Eng., The Univ. of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan)

Cloud cavitation is known well as one of the most destructive cavitations. High pressure generated by the violent collapse of a bubble cloud causes severe erosion, noise, and vibration in hydraulic machines. In the medical field, many researchers have investigated the behavior of the collapsing bubble cloud to effectively use the highly concentrated energy for HIFU (high-intensity focused ultrasound) applications. In this study, it is assumed that the acoustic cavitation forms a spherical bubble cloud which consists of many microbubbles, and ultrasound focusing in the spherically symmetric cloud is numerically investigated. The compressibility of the liquid, the evaporation and condensation of the liquid at the bubble wall, and heat transfer through the bubble wall are considered in the simulation. The pressure wave focuses to the center of the cloud and the pressure inside bubbles extremely increases when the frequency is near the first mode natural frequency of the cloud. Especially in the cases of relatively high-pressure amplitudes, a shock wave is formed in the cloud and it generates high pressure fluctuation near the center of the cloud even when the frequency is much lower than the first mode frequency.

4:40

4pBB13. Lagrangian description of the interaction of a bubble with a particle. Todd A. Hay and Mark F. Hamilton (Appl. Res. Labs., Univ. of Texas at Austin, Austin, TX 78713-8029, haymaker@mail.utexas.edu)

The Lagrangian formalism used previously to describe pulsations and translations of interacting bubbles [Ilinskii *et al.*, Proc. 17th ISNA, Penn State, edited by Atchley and Sparrow (AIP, New York, 2006)] has been extended to model interactions of bubbles and elastic particles. The bubbles and particles are assumed to be spherical. Of eventual interest is interaction of a cavitation bubble cluster with fragments of kidney stones produced during shock-wave lithotripsy. Here, we discuss only the interaction of a single gas bubble with a single rigid particle. Numerical simulations illustrate change in separation distance in the presence of an applied sound field. Using the averaging procedure employed by Harkin *et al.* [J. Fluid Mech. **445**, 377 (2001)] for two bubbles, we develop approximate equations for the translation under the conditions of weak excitation and weak interaction. Radiation force between the bubble and particle is fifth order in terms of the separation distance, as found previously by Coakley and Nyborg [in *Ultrasound: Its Applications in Medicine and Biology*, edited by F. Fry (Elsevier, Amsterdam, 1978), pp. 77–159]. Numerical and analytical results are compared with results obtained by others using different approaches. [T.A.H was supported by an NSF Graduate Research Fellowship.]

4:55

4pBB14. On the impact of vessel size on the threshold of bubble collapse. Elisabetta Sassaroli and Kullervo Hynynen (Dept. of Radiol., Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115)

Most of the studies investigating the physical properties of ultrasound contrast agents have addressed situations in which the microbubbles are contained in vessels much larger than the bubble size. Since contrast agents are normally injected into blood vessels, the investigation of their behavior in a confined environment may reveal useful information for the optimization of diagnostic and therapeutic applications of ultrasound. A series of experiments involving an ultrasound contrast agent in different sized silica and polyester tubes ranging from 50 to 400 microns in internal radius is presented. It is shown that the threshold for bubble collapse, as recorded by broadband acoustic emission, is not only dependent on the applied pressure amplitude but also on the tube size with an increase for smaller vessels. These data are interpreted and rationalized on the basis of a simple physical model. These findings may have implications for the evaluation of the risk of inertial cavitation in blood vessels. [Work supported by NIH RO1-EB003268 and by the Brigham Radiology Research and Education grant.]

Session 4pEA

Engineering Acoustics: Joe Blue Memorial Session III: Transducers

David A. Brown, Cochair
Univ. of Massachusetts-Dartmouth, Acoustic Research Lab., 151 Martine St., Fall River, MA 02723

Edmund R. Gerstein, Cochair
Leviathan Legacy Inc., 1318 SW 14th St., Boca Raton, FL 33486

Chair's Introduction—2:00

Invited Papers

2:05

4pEA1. Broadband barrel-stave flextensional transducers. Dennis F. Jones (Defence R&D Canada - Atlantic, P.O. Box 1012, Dartmouth, NS, Canada B2Y 3Z7, dennis.jones@drdc-rddc.gc.ca)

The last time I discussed underwater transduction technologies with Joe Blue was at the 125th Meeting of the Acoustical Society of America in May 1993. As we enjoyed lunch on the patio of a downtown Ottawa bistro under a sunny spring sky, the discussion touched on broadband transducers for naval applications. His ideas were insightful, motivating me to redesign and improve my original 1989 broadband barrel-stave flextensional transducer the following year. Over the last decade 30 experimental units were built at DRDC Atlantic, most of them used in marine mammal and coastal surveillance applications [D.F. Jones, J. Acoust. Soc. Am. **117**, 2447 (2005); **118**, 2038–2039 (2005)]. This paper will present electroacoustic measurements made at both the DRDC Atlantic Acoustic Calibration Barge on Bedford Basin near Halifax, Nova Scotia and the NAVSEA Seneca Lake Sonar Test Facility near Dresden in upstate New York. The performance parameters of interest include resonance frequencies, mechanical quality factors, transmitting voltage response versus water depth, and directivity patterns. [Work supported in part by the Office of Naval Research.]

2:25

4pEA2. Piezocomposite array technology for commercial and military sonar applications. Barry J. Doust and Brian G. Pazol (Material Systems Inc., 543 Great Rd., Littleton, MA 01460, bdoust@matsysinc.com)

Piezocomposites are a widely known form of piezoceramic with proven advantages over conventional monolithic piezoelectric ceramics providing improved efficiency and bandwidth. Unique manufacturing processes such as injection molding, curved structures, and layered array assembly provide creative solutions to complex, state-of-the-art designs. In addition, recent advances in piezocomposite technology have addressed the optimization of materials and electroacoustic design to provide devices with improvements in bandwidth and transmit power capability. Several state-of-the-art designs will be presented showing capability of producing complex-shaped transducers and arrays for transmit only, receive only, and two-way (transmit/receive) applications. These commercial and military systems will be presented showing the unique operational capability achieved using curved array structures, unique aperture patterns, and precisely matched phased arrays. Application examples will be presented highlighting communication and navigation, imaging, and parametric systems.

Contributed Papers

2:45

4pEA3. Modal acoustic transducer array. John L. Butler and Alexander L. Butler (Image Acoustics, Inc., 97 Elm St., Cohasset, MA 02025)

An acoustic system composed of an array of transducers, which generates multiple acoustic radiation modes, is presented and shown to yield directional steered beam patterns when the modes are appropriately combined together. A single ring array which uses eight piston transducers with a common tail mass and a cylindrical array of five of these rings are used to illustrate the performance of such a modal acoustic array. A discussion of the operation of the transducer array with a common tail mass as well as the means for combining the modes is presented. Monopole, dipole, and quadrupole array modes are illustrated with finite-element models. These modes are then combined together to generate a true cardioid, a new very wide cardioid type, and a low sidelobe narrow caridioid-type beam pattern. Calculated and FEA results are compared with measurements. [Work supported by ONR.]

3:00

4pEA4. Characterization of an underwater electroacoustic transmit system and the use of real electrical equivalents circuits for approximating loaded projectors. Corey L. Bachand and David A. Brown (BTech Acoust., Acoust. Res. Lab., and ATMC & Dept. of Elect. Eng., Univ. of Massachusetts Dartmouth, 151 Martine St., Fall River, MA 02723)

Many underwater applications using acoustic communication rely upon an effective and efficient system of transmit hardware for adequate signal strength, bandwidth, and long-term sustainability. The focus of this work is to outline a procedure for investigating and optimizing the performance of transmit hardware intended for broadband underwater communication. The transmit hardware considered in this discussion consists of a power supply, power amplifier, matching (step-up) transformer, tuning inductor, and loaded piezoelectric projector. The use of an electrical equivalent circuit of a piezoelectric projector is adopted for performing complex power measurements throughout the transmit hardware for system characterization. Parameters for constructing an electrical projector circuit can be

obtained analytically or experimentally. The complex power measurements are automated with a digital function generator and digitizing oscilloscope under software control. This method allows developers to evaluate the performance of the transmit hardware under realistic operating conditions in an engineering laboratory with standard equipment. One particular transmit hardware configuration, the WHOI Micro-modem, will be used to demonstrate the testing method.

3:15

4pEA5. Electrokinetic acoustic transducer. Nicholas P. Chotiros (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

An electrokinetic (EK) transducer consists of a liquid-saturated porous solid with electrodes at each end. It converts a pressure difference into electrical potential by means of the naturally occurring electrical charge that exists at the interface between certain liquids and solids. In 1964, Collins and Ellis [''Electrokinetic, acoustic pressure transducer,'' J. Acoust. Soc. Am. **36**, 1858–1860 (1964)] published a paper describing a prototype EK acoustic pressure transducer, based on a very simple formula for the EK potential. In 1965, Rice and Whitehead [J. Phys. Chem. **69**(11), 4017–4024 (1965)] published a more elaborate theoretical model of EK flow in a narrow capillary. Although the EK transducer had been presented as a pressure transducer, it is essentially a pressure difference transducer. This paper explores its potential in the light of Rice and Whitehead and later publications. [Work supported by the Office of Naval Research, Ocean Acoustics.]

3:30–3:45 Break

3:45

4pEA6. Investigation of magnetostrictive materials and transducers for the generation of ultrasonic waves. Matthew M. Kropf and B. R. Tittmann (The Penn State Univ., 212 EES Bldg., University Park, PA 16802, mmk230@psu.edu)

The magnetostrictive effect is commonly exploited to generate ultrasonic waves for many applications. A better understanding of the physical mechanism of magnetostriction in terms of material and transducer characteristics could expand its utilization to many new sensor applications. Maximizing the ustable frequency bandwidth for ultrasonic wave generation alone allows for various choices of modes that are individually suitable for different applications. Several different materials were analyzed and compared in terms of theoretically expected magnetostrictive efficiency. Frequency content of ultrasonic waves generated via magnetostriction was examined with reference to the geometry and configuration of a magnetic coil transducer. Static measurements were made and extended in an attempt to quantify magnetostriction in the dynamic sense. A survey of current applications and direction for future research is included.

4:00

4pEA7. An innovative piezoelectric force sensor to determine the integrity of electrical sockets. Elizabeth Blickley and Bernhard Tittmann (The Penn State Univ., 212 EES Bldg., University Park, PA 16802, eab223@psu.edu)

An ultrasonic force sensor was designed to calculate the socket contact force in a cable connector assembly. The force sensor operates through measuring the relationship between the mechanical force and change in resonance frequency in a piezoelectric tube. The sensor's aim is to assess the effectiveness of aging electric sockets to ensure their replacement before failure. A network analyzer was used to perform tests on the repeatability and sensitivity of the resonance change in two sets of piezoelectric tubes of different thickness. The hoop mode proved to be the most consistent as well as most sensitive resonance mode. The results show an increase in the change in resonance frequency corresponding to an increase in applied socket contact force. Socket contact forces are measured through a friction-based method. To refine the testing technique, the 3-D modeling software SolidWorks was used to design a prototype socket testing device to hold the tubes for insertion into the socket. The piezoelectric tubes were adapted to perform in the tube holder and the prototype socket testing device was created out of aluminum.

4:15

4pEA8. Wave propagation in liquid-filled tubes: Measurements and model predictions. Aaron M. Kyle, George R. Wodicka (Weldon School of Biomed. Eng., Purdue Univ., 500 Central Dr., West Lafayette, IN 47907-2022), Sandra L. Ordóñez, Eduardo J. Juan (Univ. of Puerto Rico-Mayaguez, Mayaguez, PR 00681-9042), and J. Stuart Bolton (Purdue Univ., West Lafayette, IN 47907-2088)

Wave propagation in liquid-filled pipes was investigated using time- and frequency-domain techniques. Water- and ethanol-filled copper and PVC pipes having inner diameters of 1.44 and 1.19 cm, and wall thicknesses of 0.0686 and 0.208 cm, respectively, were driven with tone bursts from a submerged hydrophone, and the resultant acoustic pressures in the liquid were measured. Group velocities in the copper pipe were observed to decrease with increasing frequency: velocities were 1124 m/s at 10 kHz and 1039 m/s at 50 kHz in water and 1052 m/s at 10 kHz and 902 m/s at 40 kHz in ethanol. Propagation in the PVC pipe was significantly slower, e.g., the estimated water group velocity at 10 kHz was 593 m/s. The estimated group velocities were then compared to those predicted by an acoustic transmission line model of the pipes. The predicted group velocities were within 13% of the estimated values for the copper tube for both liquids. The pipe systems were also excited with Gaussian noise, and exhibited bands of low spectral energy that could be attributed to the transverse resonances of the tube walls. The model predicted the upper frequency limit of these bands to within 5% of measured values.

4:30

4pEA9. Sonic demultiplexing by inverse designed scattering acoustical elements: An experimental demonstration. José Sánchez-Dehesa, Andreas Håkansson, and Francisco Cervera (Wave Phenomena Group, Nanophotonics Technol. Ctr., Polytechnic Univ., E-46022 Valencia, Spain)

Experiments demonstrating sonic demultiplex devices able to spatially separate several wavelengths will be reported. The devices are based on clusters of cylindrical scatterers of circular section whose positions and sizes are determined by an inverse design approach. Two devices are reported, a first device that spatially separates two wavelengths, 20.0 and 22.7 cm, and a second that separates three wavelengths: 20.0, 21.3, and 22.7 cm. Both prototypes are typical examples of acoustic scattering elements, a name here introduced to define inverse designed devices characterized by their great adaptability in controlling the scattering image, including both the reflected and transmitted wave. [Work supported by MEC of Spain.]

4:45

4pEA10. Band-gap creation using quasiordered structures based on sonic crystals. Vicent Romero Garcia, Elies Fuster Garcia, Juan V. Sanchez-Perez (ACARMA and Dep. de Fisica Aplicada, ETSICCP, UPV, Cno. de Vera s/n, 46022 Valencia, Spain, jusanc@fis.upv.es), Luis M. Garcia-Raffi, and Enrique A. Sanchez-Perez (ETSICCP, UPV, 46020 Valencia, Spain)

It is well known that sonic crystals are periodic structures that present acoustic band-gap attenuation centered at frequencies related with the lattice constant of the structure. In this work we present a new approach based on genetic algorithms to create band gaps in a predetermined range of frequencies. The mechanism used by genetic algorithms to achieve this

objective is the creation of vacancies in a starting sonic crystal, built with air-embedded rigid scatterers. As a result of the optimization process we obtain quasiordered structures whose acoustic properties are based in the multiple scattering theory. [This work has been supported by Universidad Politcnica de Valencia (Interdisciplinary Projects (P.I.I.Ref. 4114 2003), Comision Interministerial de Ciencia y Tecnologia of Spain (Contract N. MAT 2003-04993-03), and by Generalitat Valenciana (Spain) under Grant GV04B-371.]

THURSDAY AFTERNOON, 8 JUNE 2006 ROOM 554AB, 2:00 TO 4:10 P.M.

Session 4pNS

Noise and Architectural Acoustics: Heating, Ventilation and Air-Conditioning Noise Control

Daniel R. Raichel, Cochair
Eilar Associates, 2727 Moore Ln., Fort Collins, CO 80526

Bennett M. Brooks, Cochair
Brooks Acoustics Corporation, 27 Hartford Turnpike, Vernon, CT 06066

Chair's Introduction—2:00

Invited Paper

2:05

4pNS1. Cooling tower noise control. Thomas S. Paige (Kinetics Noise Control, Inc.—Vibron Products Group, 1720 Meyerside Dr., Mississauga, ON, L5T 1A3 Canada, tpaige@kineticsnoise.com)

Cooling towers are typically provided with multiple large-diameter propeller fans mounted on top of the unit and discharging upward. These fans generate a significant level of low-frequency noise. The induced-draft intakes for these fans are typically located on the sides of the unit, so fan noise is also radiated horizontally from these openings. Traditional methods for noise control include sound-barrier walls, silencers, or complete noise enclosures. Barrier walls usually have no adverse effect on the performance of the cooling towers. However, silencers and close-fitting noise enclosures must be carefully designed to minimize resistance to the airflow that is required to maintain proper operation of the cooling tower. This paper discusses the current techniques used to reduce outdoor noise propagation from cooling towers. Some case studies are presented to illustrate successful implementation of these noise-control methods.

Contributed Papers

2:25

4pNS2. Global active control of broadband noise from small axial cooling fans, Part 2. Matthew Green and Scott Sommerfeldt (Dept. Phys. and Astron., Brigham Young Univ., N283 ESC, Provo, UT 84602)

The multi-channel filtered-x LMS algorithm has previously been used for feed-forward control of the tonal noise of a small axial cooling fan [Gee and Sommerfeldt, ''Application of theoretical modeling to multi-channel active control of cooling fan noise,'' J. Acoust. Soc. Am. **115**, 228–236 (2004)]. With much of the tonal noise content removed it becomes desirable to reduce the broadband fan noise. A previous study demonstrated the feasibility of achieving desirable broadband attenuation by means of active feedback control. Measurements will be presented and discussed that have an impact on the control achievable. Experimental control results will be presented.

2:40

4pNS3. Numerical prediction of aerodynamic noise radiated from a cross-flow fan. Song Li, Anbang Chen, and Dongtao Huang (Dept. Eng. Mech., Tsinghua Univ., Beijing 100084, PROC, lisong@tsinghua.edu.cn)

The surface pressure fluctuations of blades, vortex wall, and the rear wall of a cross-flow fan are commonly considered as the aerodynamic noise sources of an air conditioning unit. In order to predict the noise, these pressure fluctuations were at first obtained by solving the 2D unsteady Reynolds-averaged Navier-Stokes (RANS) equations using a computational fluid dynamics software. It was validated by the fan performance experiments that the prediction of static pressure in the channel agreed well with the test data. It means the results of the pressure fluctuations obtained from the numerical simulation are correct. Then, the Ffowcs Williams-Hawkings (FW-H) equation was used to predict the far-field noise caused by these sources. The errors of the prediction of the sound-pressure levels of the blade passing frequency (BPF) and total noise are within 4 and 3 dB, respectively, comparing with the corresponding measured data. It is shown from the numerical prediction that only the surface pressure fluctuations of the rear wall and the vortex wall are main sources, and that the method of numerical prediction of aerodynamic noise radiated from a cross-flow fan is successful.

2:55

4pNS4. An analytical study of the effectiveness in noise control due to combination of mufflers in design of duct systems. B. S. Sridhara (Dept. of Eng. Technol. and Industrial Studies, Middle State Tennessee State Univ., Murfreesboro, TN 37132) and M. G. Prasad (Stevens Inst. of Technol., Hoboken, NJ 07030, mprasad@stevens.edu)

Mufflers are effective noise control devices used in acoustical design of duct systems. The performance evaluation of a single muffler in terms of four-pole matrix approach is well known. The performance evaluated in terms of four-pole parameters and insertion loss describes the system using various components namely source, muffler, and termination. The performance of a reactive muffler is known to be frequency dependent based on its internal geometrical details. This paper presents analytical studies on the effectiveness of combining mufflers on the noise control. The studies deal with both series and parallel arrangement of the reactive mufflers.

The effects of various parameters such as number of mufflers, the length of ducts associated with the system, and the muffler geometry are presented. For a given source model, simple expansion chamber-type mufflers are studied in detail in order to understand the nature of interaction in different types of arrangements. Applications of this study would be to increase the noise reduction by combining mufflers in systems such as automotive, HVAC, etc.

3:10

4pNS5. Subjective and objective evaluation of air-conditioning noise. Jin Yong Jeon, Jin You, and Su Yeon Kim (School of Architectural Eng., Hanyang Univ., Seoul 133-791, Korea)

Noise from an air-conditioning system installed in the ceiling of classrooms was evaluated to investigate the subjective responses to air-conditioning noise. Harmonics of air-conditioning noise was removed and the sound-pressure levels of the frequency bands from 250 to 630 Hz were varied to control sound quality. Evaluation for air-conditioning noise sensation was conducted using a paired comparison method and the results showed that noise without harmonics was preferred to noise with harmonics. The noises which have larger level increases to the noise at 250 to 630 Hz were evaluated better in both of the noises with and without harmonics. Results of multiple regression analysis on psychoacoustic parameters and subjective preferences showed sharpness as a major factor in describing the sound quality of air-conditioning noise. A semantic differential test was also conducted to evaluate the sound-quality characteristics of air-conditioning noise.

3:25

4pNS6. Acoustic performance of sound-attenuating ducts utilizing spatially periodic area changes with absorbing material. Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708) and Pavel Danilov (Continuum-Dynam., Inc., Ewing, NJ 08618)

Sound propagation of one-dimensional waves through a duct with spatially periodic area changes has been studied theoretically and experimentally. The resulting wave behavior is similar to that observed for periodic structures, with Bloch waves, passbands, and stopbands. This configuration substantially reduces sound transmission from noise sources having harmonics that fall in the stop bands, while having little effect in passbands. The present study adapts this concept to configurations with dissipative sound-absorbing material. The periodically spaced regions of the duct with larger cross-sectional area are partially filled with porous material, so that the open cross-sectional area remains constant throughout the duct. The duct is analyzed as a quasi-one-dimensional system using an approach based on periodic structure theory. The porous material makes the propagation wave number in the porous/open subsections complex, thereby changing the nature of the stop- and passbands, and providing significant attenuation even in passbands. Results demonstrate that the best attenuation strategy may be to distribute the porous material in periodically spaced subregions rather than in either a single larger volume, or in a uniform distribution all along the duct, particularly if cancellation of discrete harmonics is desirable.

3:40

4pNS7. Acoustic impedance of perforations in contact with absorbing material. Iljae Lee, Ahmet Selamet (The Ohio State Univ., 930 Kinnear Rd., Columbus, OH 43212, selamet.1@osu.edu), and Norman Huff (Owens Corning, Novi, MI 48377)

Acoustic impedance of perforations in contact with fibrous material is experimentally determined for various hole geometries and filling densities in the absence of mean flow. Two setups are developed to measure the perforation impedance and the acoustic properties of the absorbent, while yielding empirical expressions for both. The experimental results show that both real and imaginary components of the perforation impedance substantially increase as a result of the contact with the fibrous material, whereas they generally decrease with increasing porosity for a given filling density. The empirical expressions developed in this study are implemented in a boundary element method to predict the transmission loss of dissipative silencers, which is then compared with the experimental results.

3:55

4pNS8. Numerical prediction of vortex shedding noise radiated by a centrifugal impeller. Weixiong Wang and Dongtao Huang (Dept. of Eng. Mech., Tsinghua Univ., Beijing PROC, 100084)

Numerical prediction of fan noise is very difficult work because of the restriction of the computer technology. The fluctuating blade force caused by the vortex shedding from a centrifugal impeller is a main noise source but no paper has been published so far to predict this kind of noise. Vortex flow models to predict low-pressure axial flow fans were developed by several researchers since 1964. In this paper a study of numerical prediction of vortex shedding noise radiated by a commercial centrifugal impeller is reported, which is based on a vortex flow model developed by Lee in 1993. Three main research works are carried out by the authors. First, using FLUENT software a numerical simulation of 3D turbulent flow field of the impeller in order to predict the aerodynamic performance is completed and well validated by the experiment. Then, the velocity profile near the wall at the trailing edge of the blade is analyzed in detail to determine the boundary layer thickness both on the pressure and suction sides. Finally, the total SPL of the noise is predicted and the results agree with the experimental data within 3 dB near the design operation.

Session 4pPA

Physical Acoustics: Thermoacoustics and Resonating Systems

Charles R. Thomas, Chair
Roger Williams Univ., 1 Old Ferry Rd., Bristol, RI 02809

Contributed Papers

1:45

4pPA1. An ultrasonic analog for a laser. Richard Weaver, Oleg Lobkis (Dept. of Theoret. and Appl. Mech., Univ. of Illinois, 216 Talbot Lab., 104 So. Wright, Urbana, IL 61801), and Alexey Yamilov (Univ. of Missouri at Rolla, Rolla, MI 65409)

We report measurements on two simple ultrasonic systems analogous to random lasers. One system entails unstable ultrasonic feedback between distinct piezoelectric transducers on an irregularly shaped high-Q solid; another involves a spontaneously emitting piezoelectric device that can also emit by stimulation. Both are found to exhibit some of the behaviors of lasers. Over a wide range of parameters we observe narrow single-emission lines (natural width is estimated at about 0.000 000 000 1 Hz), sensitivity to linear cavity properties, complex multimode emissions, and line narrowing.

2:00

4pPA2. Time-reversal acoustic focusing system based on the liquid-filled acoustic resonator. Alexander Sutin (Artann Labs., 1459 Lower Ferry Rd., West Trenton, NJ 08618, asutin@artannlabs.com), Yegor Sinelnikov (ProRhythm, Ronkonkoma, NY 11779), and Armen Sarvazyan (Artann Labs., West Trenton, NJ 08618)

Numerous time-reversal acoustic focusing systems (TRA FS) based on the use of solid acoustic resonators (acoustic cavities) capable of accumulating acoustic energy in time are described in the literature. We experimentally demonstrated that the liquid-filled resonators can be superior to solid resonators in efficient focusing, especially in applications related to focusing of ultrasound in waterlike media, such as biological tissues. To understand the mechanisms of the observed effects and optimally design the water-filled resonators for biomedical applications, we developed a theoretical model and explored the factors defining the efficiency of TRA FS. The model allows evaluation of the spatial and temporal characteristics of TRA focused field formed as a result of numerous reflections of ultrasound within the resonator. The focused field structure and its dependence on the shape and size of the resonator, ultrasound frequency, and attenuation, and the coefficient of ultrasound reflection from the resonator walls were investigated. Principles of optimization of the resonator shape and size are developed. It is shown that for the resonator filled with an attenuating liquid TRA focusing in binary mode of radiation provides narrower focal spot than conventional analog signals. The experiments confirmed the predictions of developed theory.

2:15

4pPA3. Synchronized-piezoelectric auto-oscillators, a classical analog laser. Richard Weaver and Oleg Lobkis (Dept. of Theoret. and Appl. Mech., Univ. of Illinois, 216 Talbot Lab., 104 So. Wright, Urbana, IL 61801)

We report the behavior of two ultrasonic transducers, each part of an auto-oscillating Wein bridge, in contact with a high-Q elastic body. In particular, we compare the behavior of the one- and two-transducer cases. Each transducer oscillates erratically, and at its own frequency, when isolated and not in contact with the elastic body. When placed in contact with an elastic body, each transducer then stabilizes at a frequency close to the range of the original erratic oscillation, where it excites a "mode" (not necessarily a normal mode of the body) of the system. When both are in contact with the elastic body, a new mode is excited, and both transducers oscillate at the same frequency. They are locked to themselves and to each other by virtue of their mutual access to the same acoustic system, and the power output of the pair is, presumably, approximately 4 times as great as that of either transducer by itself.

2:30

4pPA4. Band gap engineering for two-dimensional periodic systems of fluids. Manvir S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla 72570, Mexico) and Bahram Djafari-Rouhani (Univ. of Sci. and Technol. of Lille I, Lille, France)

Periodic binary systems can give rise to genuine acoustic stop bands within which sound and vibrations remain forbidden. Extensive band structures have been computed for two-dimensional (2D) periodic arrays of air cylinders in water background. Complete, multiple, huge stop bands are found for both square and triangular lattice arrangements. The lowest stop bands are largest for a range of filling fraction $10\% \leqslant f \leqslant 55\%$, with a gap/midgap ratio of 1.8. The most interesting finding of the present investigation is that the low-frequency, flat passbands for a perfectly periodic system correspond to the discrete modes of a *single* airy cylinder. This is attributed to the low-filling fraction and huge density contrast in air and water. It is stressed that such a simple inhomogeneous system as made up of air and water exhibits the largest stop bands ever repoprted for 2D and/or 3D phononic as well as photonic crystals.

2:45

4pPA5. Transmission effects of flexural shell vibration modes in sonic band-gap crystals made of hollow cylinders in air: A theoretical analysis. Yuriy A. Kosevich, José Sánchez-Dehesa (Wave Phenomena Group, Nanophotonics Technol. Ctr., Polytechnic Univ., E-46022 Valencia, Spain), and Cecile Goffaux (CENAERO, B-6041 Gosselies, Belgium)

The transmission properties of periodic arrays of hollow cylinders in air are studied analytically and numerically using a finite-difference–time-domain method. Particularly, transmission spectra across slabs of a few layers of cylinders arranged in a square lattice will be reported. It is shown that $n=2$ and $n=3$ flexural shell vibration modes [A. Leissa, *Vibrations of Shells* (Acoustical Society of America, New York, 1993)] associated with the hollow cylinders play a fundamental role in the band-gap transmission properties of this two-dimensional sonic composite medium. The parameters of the hollow cylinders (thickness and elastic parameters of the material) are chosen in order to have the frequencies of the $n=2$ and $n=3$ flexural shell vibration modes inside, correspondingly, the first and third band gaps of the sonic crystal. In comparison with corresponding sonic crystal made of full cylinders, the transmission properties of this new composite medium present two main physical features: (i) huge transmission resonances with asymmetric Fano-type shape appear in the band gaps; and (ii) widening of the band gap widths occurs and increases with filling fraction of cylinders in the sonic crystal. Analytical expressions for the resonance frequencies and physical picture of both phenomena will be presented. [Work supported by MEC of Spain.]

3:00

4pPA6. Band gap engineering for three-dimensional periodic systems of air bubbles in water. Manvir S. Kushwaha (Inst. of Phys., Univ. of Puebla, P.O. Box J-45, Puebla 72570, Mexico), Bahram Djafari-Rouhani, and Leonard Dobrzynski (Univ. of Sci. and Technol. of Lille I, Lille, France)

Systematic and extensive evidence for the existence of multiple, complete, huge stop bands in the band structure for three-dimensional (3D) cubic arrays of air bubbles in water is reported. A Fourier series expansion method, which does not require matching of the messy boundary conditions, is used to investigate all three important 3D geometries such as face-centered cubic (fcc), body-centered cubic (bcc), and simple-cubic (sc) arrangements. The lowest stop bands are largest for a volume fraction $f \leqslant 10\%$, with a gap/midgap ratio of ≈ 1.8 for all three geometries. Surprising but rigorously justifiable is the fact that the low-frequency flat passbands for the perfectly periodic systems correspond to the discrete modes of a *single* bubble. This is an artifact of the low-filling fraction and huge density contrast in air and water. It is emphasized that such a simple inhomogeneous system made up of air bubbles in water can exhibit the widest stop bands ever reported for elastic and/or dielectric periodic composites.

3:15–3:30 Break

3:30

4pPA7. History effects on the motion of particles subject to acoustic fields. John S. Allen, III, Carlos F. M. Coimbra, and Marcelo H. Kobayashi (Dept. of Mech. Eng., Univ. of Hawaii—Manoa, 2540 Dole St., Honolulu, HI 96822)

Acoustic fields have been used to trap, manipulate, and separate particles suspended in fluids. The separation of solid and gas particles is possible since the acoustic radiation force depends on the particle's size, density, and compressibility. Moreover, numerous devices using acoustic standing waves have been developed for this purpose. The ability to model and accurately predict the particles' trajectories and distributions resulting from these applied acoustic fields is of great importance for a wide range of biomedical and industrial applications. However, limited previous theoretical efforts have been made to account for the unsteady viscous effects on a particle's motion in an acoustic field. We examine these unsteady history effects on a particle in the vicinity of a nodal point in an acoustic standing wave field. The relative importance of the history (Basset) force, which has a fractional derivative dependence of the order 3/2, is demonstrated for acoustic separation device conditions. The inclusion of the history force has the net effect of changing the eigenfunctions of the problem from exponential to Mittag-Leffler functions in time, thus influencing the overall acoustic force balance.

3:45

4pPA8. Possibilities of thin-film bulk acoustic resonators as punctual mass sensors. Humberto Campanella, Jaume Esteve, Josep Montserrat (Centro Nacional de Microelectrnica, Campus UAB, 08193 Bellaterra Barcelona, Spain, humberto.campanella@cnm.es), Arantxa Uranga, Gabriel Abadal, Nuria Barniol (Universitat Autnoma de Barcelona, Barcelona, Spain), and Albert Romano (Universitat de Barcelona, Barcelona, Spain)

In this work, thin-film bulk acoustic resonators (FBARs) are implemented as punctual mass sensors, with the purpose of future integration in biomolecular sensing systems, where selective spatial detection is required. The piezoelectric thin film is a (002)AlN membrane, sputtered over Ti/Pt on a (001)Si wafer, and released by surface micromachining of silicon. The effective lateral area of resonator is $50\times 50\mu$m, and the thicknesses of AlN and Ti/Pt layers are 1000 and 180 nm, respectively. A punctual mass of a C/Pt/Ga composite with a contact surface much smaller compared to that of the resonator's electrode is deposited on the top electrode of the resonators, by means of a focused ion beam. Selected areas of the electrode, especially near the center of the resonator, were chosen to growth the composite with lateral dimensions in the units of micrometers, and thickness on the order of several nanometers. Based on the mass-loading principle, the sensor changed its resonance frequency in a proportional manner to the added mass, achieving an average sensitivity of 2.4^{-22} g/Hz and a mass resolution an the order of femtograms. Regarding these results, FBAR is proposed as a candidate to measure localized particles and, in general, to be applied as biological mass sensor.

4:00

4pPA9. Unconventional thermoacoustic heat engines. Matthew G. Hilt, K. A. Pestka, II, G. D. Mahan, J. D. Maynard (The Penn State Univ., University Park, PA 16802), D. Pickrell, B. Na, and J. Tamburini (Omega Piezo, State College, PA 16801)

During 1988 to 1994, John Strachen and Harold Aspden demonstrated an all-solid-state thermoelastic cooler which was 0.5 cm thick and a few centimeters in length. With 7.2 V dc applied across the device, it drew 6.3 W of electrical power and produced 13.7 W of cooling power with a 20 °C temperature drop. The coefficient of performance was 2.2, or 21% of the maximum possible (Carnot) efficiency. The device operated reversibily; 3.7 W of heat throughput generated 0.15 W of electrical power, for an efficiency of 47% of Carnot. The device operated at a frequency of about 500 KHz. To date, no one has found an explanation for the effect. Recently we have begun a program to develop a theory and to reproduce a working device. This research, as well as some research with other unconventional thermoacoustic devices, will be presented. [Research supported by DARPA and the Aviation and Missle Command.]

4:15

4pPA10. An open-air infrasonic thermoacoustic engine. William Slaton (Dept. of Phys. & Astron., The Univ. of Central Arkansas, Conway, AR 72035)

Wind noise at infrasound monitoring stations is a significant problem especially at the low frequencies of interest. Noise-reducing filters, typically consisting of radial arrays of porous hose, effectively area average the infrasonic pressure wave over a large space. The resulting area-averaged sound wave is then measured by a microbarometer. However, the only method to calibrate both the microbarometer and wind-noise filter is to use atmospheric or land-based man-made explosion. Thus, an infrasound source is needed that can calibrate both the microbarometer and the wind-noise filters in a safe way. A lowfrequency open-air thermoacoustic prime mover in a Helmholtz resonator has been constructed to fulfill this purpose. The system resonates at 10 Hz, is portable, compact, and safe to operate. The engine's experimental and theoretical performance will be presented. [Work supported by a subcontract with the National Center for Physical Acoustics and a grant from the University Research Council at the University of Central Arkansas.]

4:30

4pPA11. Optimizing miniature thermoacoustic coolers. H. El-Gendy, Y. Kwon, L. Lyard, and O. G. Symko (Dept. of Phys., Univ. of Utah, 115 South 1400 East #201, Salt Lake City, UT 84121)

Because of limited sound intensity output from commercial drivers at midaudio and ultrasonic frequency ranges used in thermoacoustic coolers, it is important to optimize their performance. To achieve this, studies were conducted on heat transfer at the cold heat exchanger and at the hot heat exchanger. Measurements were taken on stray heat influx to the cooler by mechanisms of convection, conduction, radiation, and streaming, and from the driver. PIV studies show the contributions of streaming, both from the driver and other parts of the devices. A special 4-layer copper mesh heat exchanger was designed and tested to cope with heat pumped acoustically from the cold heat reservoir as well as the streaming from the piezoelectric driver. For typical sound levels of 155–160 dB achieved inside the resonator by commercial bimorph piezoelectric drivers, a cooling power density of 0.5 watts/cm^2 was achieved with air at 1 atm. Pressurizing the working gas to 17 atm raises the cooling density. The choice of working gas provides a increase in power density. With the above optimizing approaches, a cooling power density of over 50 watts/cm^2 appears feasible for the frequency range of 4 to 24 kHz. [Work supported by a grant from the ONR and the State of Utah.]

4:45

4pPA12. Hot wire anemometry for the measurement of temperature and velocity fluctuations inside a thermoacoustic cooler. Arganthaël Berson, Philippe Blanc-Benon (Ctr. Acoustique, LMFA, UMR CNRS 5509, Ecole Centrale de Lyon, 36 avenue Guy de Collongues, 69134 Ecully Cedex, France), Luc Mongeau, and Insu Paek (Purdue Univ., West Lafayette, IN 47907-2031)

Hot wire anemometry was used to measure velocity and temperature fluctuations inside a thermoacoustic refrigerator. In an oscillating flow with zero-mean velocity, such as in an acoustic standing wave, heat transfer is different from heat transfer in a steady flow. A dynamic calibration of the system is required. This calibration was done inside the resonator of the cooler, with no stack and no heat exchangers. Two thermoacoustic systems were studied to calibrate the anemometer in air at atmospheric pressure and in He/Ar mixtures at a mean pressure of 6 bars. Measurements of temperature and velocity were done at the end of the stack and heat exchangers. Correlations between temperature and velocity fluctuations were obtained, and the enthalpy convective flux was computed. These results were compared with results obtained from numerical simulations recently done at Ecole Centrale de Lyon [D. Marx, AIAA J. **42**(7) (2004)].

THURSDAY AFTERNOON, 8 JUNE 2006 ROOM 555AB, 1:00 TO 4:50 P.M.

Session 4pPP

Psychological and Physiological Acoustics: Characterizing Auditory Attention

Frederick J. Gallun, Chair
Boston Univ., Speech, Language and Hearing Science, 635 Commonwealth Ave., Boston, MA 02215

Invited Papers

1:00

4pPP1. Auditory attention: One term for several concepts? Frederick J. Gallun (Dept. of Speech, Lang. and Hearing Sci. and Hearing Res. Ctr., Boston Univ., 635 Commonwealth Ave., Boston, MA 02215)

The recent rise in the use of spectrally and temporally complex stimuli appears to be responsible for the increasingly frequent use of the term attention as an explanation of results in auditory experiments. It seems timely, therefore, to provide a broad examination of the various meanings often associated with this term. The goal of this session is to explore the ways in which the concept of auditory attention has been applied and perhaps suggest ways in which its application could be refined. An essential first question then is whether auditory attention refers to one concept or many. This talk will begin by examining the historical uses of this term and then proceed to outline some of the contemporary uses of the term, relating them back to the theoretical concepts the term is meant to represent. A specific example will be provided by reviewing a recent series of experiments in which the experimental designs were intended to manipulate listener attention and which were interpreted as providing insight into the mechanisms available for performing a variety of auditory tasks. [Work supported by NIH/NIDCD.]

1:25

4pPP2. Cuing in auditory detection. E. R. Hafter (Dept. of Psych., Univ. of California, Berkeley, CA 94720)

A traditional theory of auditory attention describes it metaphorically as the beam of an attentional searchlight. It underlies, for example, the logic of the auditory critical ratio, where subjects detect a tonal signal in a wideband masker, where the extent to which processing is restricted to a specific frequency-region is inferred through calculation of the critical ratio between signal level at threshold and the spectrum level of the masker. Values so found correlate well with results from other techniques such as restriction of the masker by notched noise and presentation of unexpected probe signals that are presumably attenuated by the skirts of the listening band. Lower performance in response to subjective uncertainty about the signal's frequency has been modeled by assumptions of a wider attentional searchlight, i.e., listening band, an idea supported by alleviation of uncertainty effects by cues that inform the subject about where to listen. The plan here is to review experiments in which various kinds of cues have been used to focus attentional searchlights, including a discussion of the potential for use of this method for examining filters at higher levels of auditory analysis where the effective signal is an emergent property of the basic stimulus.

1:50

4pPP3. Attending to auditory events "takes" time . . . event time. Mari R. Jones (Dept. of Psych., The Ohio State Univ., Columbus, OH 43210)

Commonplace acoustic stimuli in our environment unfold in time over many time scales, ranging from milliseconds to seconds to minutes. At each time level, there is some ability to perceive and/or ignore stimulation. What does this mean for the role of attention? I suggest that auditory attention takes time, but not necessarily processing time. Instead, moment-to-moment attending literally takes hold of (i.e., locks into) event time spans by entraining to event time at one or more levels. Efficient attending involves an internal

activity (a pulse of attentional energy?) that is capable of coordinating in time with an unfolding sound pattern such that attending energy coincides with to-be-attended elements. My main hypothesis is that, to attend efficiently at any time scale, synchronicity of attending with constituent sound elements is a necessary, but possibly not a sufficient, condition. I unpack this idea using illustrative dynamical models that assume an entrainment approach to attending. Several experiments use auditory sequences in which the time structure of constituent tones is systematically varied. In some tasks, listeners judge standard and comparison time intervals in different rhythmic contexts; in others, they identify the pitch of probe tones in sequences where both pitch and time structure vary.

2:15

4pPP4. Multiple mechanisms of auditory attention. Elyse S. Sussman (Dept. of Neurosci., Albert Einstein College of Medicine, 1410 Pelham Pkwy. S., Bronx, NY 10461, esussman@aecom.yu.edu)

Selective attention limits processing so we can handle the vast amount of information that we encounter daily. Actively selecting a subset of sensory information has been shown to bias response properties of neurons within sensory cortices, generally seen as an enhancement of the neural activity associated with the attended stimuli. This talk will present another type of attention effect on neural activity, affecting the representation of sequential sounds in memory. Event-related brain potentials were used to identify the timing of various cognitive processes, the subjects' perception of the sounds, and to assess the stages at which attention interacted with sensory input in perception. In one study, attention modified the grouping and storage of sequential sounds, reflected by the absence or presence of specific ERP components. In another study neural representations were biased by stimulus-driven factors without attentional control. In a third study, attention increased accuracy for discriminating stimulus information but did not change the way the information was organized in auditory memory. These results indicate that there is more than one mechanism of auditory attention operating at different processing stages depending on the listeners task, and supports the view that attention interacts with stimulus-driven processes to facilitate behavioral goals. [Research supported by NIH.]

2:40–2:55 Break

2:55

4pPP5. Spatial auditory attention. Barbara G. Shinn-Cunningham (Hearing Res. Ctr., Boston Univ., 677 Beacon St., Boston, MA 02215)

Historically, most psychoacoustic studies measured performance on tasks limited by bottom-up, sensory factors. However, in the everyday world, sensory sensitivity does not always determine performance. The problem in many everyday settings is not that components of an important sound source cannot be heard, but rather that they are not noticed amidst the clutter caused by competing sound sources. In particular, recent evidence suggests that competition for central resources is often the primary limitation on performance in tasks in complex, uncertain environments. This talk will review results of recent studies demonstrating that spatial cues are one important feature that can enable a listener to focus attention on a sound source of interest and mediate competition for central resources. Psychophysical results will be interpreted in light of physiological measures from visual and auditory tasks illustrating the importance of top-down factors on neural activity. Parallels and differences in the role of spatial attention for visual and auditory tasks will be discussed. Speculations about these similarities and differences will be offered, driven by the hypothesis that similarities derive from common control circuitry, while differences arise from differences in peripheral neural coding. [Work supported by AFOSR, ONR, and NIDCD.]

Contributed Papers

3:20

4pPP6. Selective attention with ambiguous dichotic-pitch images. Frans A. Bilsen (Delft Univ. of Technol. Delft, The Netherlands)

Experiments in the past have shown that several types of dichotic pitch arise from dichotic white noise stimuli having a particular interaural phase relation [e.g., F.A. Bilsen and J. Raatgever, compact disc "Demonstrations of Dichotic Pitch," Delft University of Technology (2002)]. Often such stimuli generate two or more ambiguous images inside the head at different places along the lateralization axis, each image having the same or a significantly different pitch value. Which pitch at which place is heard is a matter of selective attention. Typically, only one image can be heard at a time (ambiguity of place coupled with pitch). New experiments on dichotic repetition pitch will be reported that show the existence of a continuum of different pitches for one parameter setting of the stimulus. Which pitch is perceived depends on the lateralized position of the focus point of attention by the listener. Implications for the organization of the binaural auditory scene and the process of selective attention will be reviewed.

3:35

4pPP7. Frequency-band importance functions for auditory and auditory-visual speech by hearing-impaired adults. Ken W. Grant (Walter Reed Army Medical Ctr., Army Audiol. and Speech Ctr., Washington, DC 20307-5001)

How we listen to speech appears to be fundamentally different in a number of respects when we can simultaneously see the face of the speaker while listening to his or her speech. Not only is our ability to recognize and understand speech far better when we can see as well as hear, as compared to hearing alone, but the way we hear is also different. This is evident in the patterns of activity in primary auditory cortex, which are different for auditory versus auditory-visual speech inputs, as well as data from normal-hearing subjects suggesting that auditory attention is focused on different spectral regions depending on whether visual speech cues are available. The present study examines whether frequency band-importance functions are different for auditory and auditory-visual speech recognition in a group of hearing-impaired listeners. Results show that hearing-impaired listeners perform much like normal-hearing subjects in attending to primarily low-frequency speech cues under auditory-visual conditions. In contrast, under auditory-alone conditions, hearing-impaired

listeners place more importance on mid-to-high-frequency acoustic information, presumably to better recognize place of articulation cues.

3:50

4pPP8. Spatial uncertainty reduces the benefit of spatial separation in selective and divided listening. Antje Ihlefeld, Shawn J. Sarwar, and Barbara G. Shinn-Cunningham (Hearing Res. Ctr., Boston Univ., 677 Beacon St., Boston, MA 02215)

Spatial separation between two talkers can improve speech intelligibility (i.e., cause spatial unmasking) when attending to either one talker (selective listening) or both talkers (divided listening). Such spatial unmasking occurs by (1) allowing listeners to hear otherwise inaudible portions of the signal(s) (energetic masking) and (2) reducing central competition between similar talkers (informational masking, IM). Here, we examine how *a priori* knowledge affects spatial unmasking in selective and divided listening in IM-dominated stimulus conditions. Performance was measured in blocks where the locations of the talkers were either constant (fixed) or varied from trial to trial (spatially randomized). Spatial separation improves performance in both selective and divided tasks when locations are fixed. In the fixed conditions, the acoustic signal-to-noise ratio at the better ear for the less-intense talker partially accounts for the spatial unmasking; however, there is also a large residual spatial unmasking, similar to that seen in previous studies. In the spatially randomized conditions, spatial unmasking is much reduced. In addition, in the divided listening, randomized conditions there is little evidence for any better-acoustic-ear effect. Results suggest that the way in which subjects utilize spatial cues depends on their knowledge of the source locations. [Work supported by ONR Grant No. N00014-04-1-0131.]

4:05

4pPP9. Spatial cues are used differently for localizing and identifying the same attended auditory object. Adrian K.C. Lee (Div. of Health Sci. and Technol., Speech and Hearing Biosci. and Technol. Prog., Harvard-MIT, 77 Massachusetts Ave, Cambridge, MA 02139), Ade Deane-Pratt, David McAlpine (Univ. College of London, London WC1X 8EE, UK), and Barbara Shinn-Cunningham (Hearing Res. Ctr., Boston, MA 02215)

Spatial cues not only affect localization of auditory stimuli, they also influence the formation of perceptual objects (streams). Previous experiments demonstrate that spatial cues influence the grouping of an ambiguous target tone that logically could be part of two different objects in the auditory scene. For such ambiguous stimuli, spatial cues affect how the target tone is grouped, changing the perceived rhythm of a repeating tone stream and the perceived identity of a slowly repeating vowel [Lee *et al.*, ''The missing target: Evidence of tone's inlability to contribute to the auditory foreground'' Proc. Mid-Winter Meeting of the ARO, New Orleans (2005)]. Here, we investigate whether perceived location of tone and vowel streams can be predicted by the grouping results of these previous identification experiments. Results show that the perceived location of the attended object (the repeating tones or vowels) depends on (1) interactions between the perceived locations of the two objects comprising the stimuli and; (2) the spatial cues contained in the target tone. Moreover, the degree to which the target tone affects perceived object location is not predicted by the degree to which the target is perceived as part of the attended object. This suggests that spatial cues affect object formation and object location, but in ways that are independent of one another.

4:20—4:50
Panel Discussion

THURSDAY AFTERNOON, 8 JUNE 2006 ROOM 551B, 1:30 TO 4:30 P.M.

Session 4pSA

Structural Acoustics and Vibration: Radiation and Scattering

Sabih I. Hayek, Chair
Pennsylvania State Univ., Dept. of Engineering Science and Mechanics, 212 Earth Eng. Sci. Bldg., State College, PA 16802-6812

Contributed Papers

1:30

4pSA1. Visualization of the energy flow for guided waves near the surface of an ensonified fluid-loaded elastic cylindrical shell via the Poynting vector field. Cleon E. Dean (Phys. Dept., P.O.B. 8031, Georgia Southern Univ., Statesboro, GA 30460-8031, cdean@georgiasouthern.edu) and James P. Braselton (Georgia Southern Univ., Statesboro, GA 30460-8093)

The Poynting vector field is used to show the energy flow in the near-surface region of an ensonified fluid-loaded elastic cylindrical shell for both forward and backward propagating waves. The present work uses a method adapted from a simpler technique due to Kaduchak and Marston [Kaduchak & Marston, ''Traveling-wave decomposition of surface displacements associated with scattering by a cylindrical shell: Numerical evaluation displaying guided forward and backward wave properties,'' J. Acoust. Soc. Am. **98**, 3501–3507 (1995)] to isolate unidirectional energy flows.

1:45

4pSA2. Nonaxisymmetric acoustic scattering from submerged prolate spheroidal shells. Jeffrey E. Boisvert (NAVSEA Newport, Code 1514 Bldg 1258, Newport, RI 02841) and Sabih I. Hayek (Penn State Univ., University Park, PA 16802)

The equations of motion for nonaxisymmetric vibration of prolate spheroidal shells of constant thickness were derived using Hamiltons principle [S. I. Hayek and J. E. Boisvert, J. Acoust. Soc. Am. **114**, 2799–2811 (2003)]. The shell theory used in this derivation includes shear deformations and rotatory inertias. The shell displacements and rotations were expanded in infinite series of comparison functions. These include associated Legendre functions in terms of the prolate spheroidal angular coordinate and circular functions in the azimuthal angle coordinate. The shell is insonified by a plane wave with arbitrary incident angle. The external (heavy) fluid-loading impedance was computed using an eigenfunction expansion of prolate spheroidal wavefunctions. Far-field backscattered acoustic pressure spectra are presented as a function of incident angle for several shell thickness-to-half-length ratios ranging from .005 to 0.1, and

for various shape parameters, a, ranging from an elongated spheroidal shell ($a=1.01$) to a spherical shell ($a\sim100$). The far-field directivity of acoustic scattering is presented at selected frequencies and plane wave incident angles. [Work supported by the ONR/ASEE Summer Faculty Research Program and the NAVSEA Newport ILIR Program.]

2:00

4pSA3. Acoustic scattering of fluid-filled spheres: Modeled and experimental results. David Deveau (Naval Undersea Warfare Ctr., 801 Clematis St., West Palm Beach, FL 33401, david.deveau@autec.navy.mil)

Investigation into the scattering nature of surfaces or other physical objects often requires the use of measurement systems which cannot always be well controlled. This lack of control can be compensated for by calibrating the resulting measurements against a known target. While these targets can be any object, the goal is to use a target that has a stable scattering response and is independent of angle. The ideal shape is that of a sphere but even this can be improved with the addition of internal fluids that focus and temperature stabilize the scattering response. A scatter response model of a sphere has been developed and used to design four thin-walled spheres, each with a different diameter and filled with a focusing fill fluid (Fluorolube). One pair of spheres was measured in an ocean environment while the second pair was tested in a controlled test tank from 5–120 kHz but using shorter continuous wave pulses. While the ocean-measured spheres closely matched the model, the test tank measurements provided insight to the model's near-field limitation and low-frequency noncontinuous wave effects.

2:15

4pSA4. Liquid-filled focusing cylinders and the caustic merging transition. Christopher Dudley and Philip L. Marston (Phys. and Astron. Dept., Washington State Univ. Pullman, WA 99164-2814)

Thin spherical shells filled with a liquid having a sound speed significantly less than the speed of sound in water are widely used for targets having enhanced backscattering. This enhancement has been analyzed using quantitative ray theory [P. L. Marston, Phys. Acoust. **21**, 1–234 (1992)]. Backscattering enhancements are investigated here for a liquid-filled flat-ended thin fiberglass circular cylindrical shell in water. In addition to partially focused enhanced backscattering for near-broadside illumination, this target exhibits enhancements for specific ranges of tilt angles associated with caustic merging transitions previously demonstrated for blunt plastic cylinders [F. J. Blonigen and P. L. Marston, J. Acoust. Soc. Am. **107**, 689–698 (2000)]. The refractive properties for the tilted cylinder depend on the effective refractive index N that is a function of the tilt angle and the actual acoustic refractive index of 1.68 (relative to water) for the enclosed liquid. Backscattering enhancements are visible from the tilted cylinder associated with two- and four-chord transmitted internally reflected rays when the tilts are such that N is close to 2 and is close to 4, respectively, as predicted by ray theory. The wave-number-radius product was as small as 16. [Work supported by the Office of Naval Research.]

2:30

4pSA5. Combining an energy boundary element analysis with an energy finite element analysis for computing airborne noise inside a flexible structure due to an exterior acoustic source. Geng Zhang (Michigan Eng. Services, 3916 Trade Ctr. Dr., Ann Arbor, MI 48108) and Nickolas Vlahopoulos (Univ. of Michigan, Ann Arbor, MI 48109)

The energy boundary element analysis (EBEA) has been utilized in the past for computing the exterior acoustic field at high frequencies (above ~400 Hz) around vehicle structures and numerical results have been compared successfully to test data [Wang *et. al.*, J. Sound Vib. **126**, 278–283 (2004); Internoise (2002)]. The EFEA has been developed for computing the structural vibration of complex structures at high frequencies and validation has been presented [Zhang *et. al.*, J. Sound Vib. **282**, 679–700 (2005); **41**, 11–12 (2005); Finite Elements Anal. Des. **41**(11-12) (2005); Wang *et. al.*, SAE Paper No. 2005-01-2372]. In this presentation the EBEA is utilized for computing the acoustic field around a cylindrical structure (similar to the NASA aluminum test-bed cylinder) from an external acoustic noise source. The computed exterior acoustic field comprises the excitation for the EFEA analysis. The structural vibration and the interior noise levels are computed, demonstrating how the EBEA and the EFEA can be combined in airborne noise applications for performing a complete vibro-acoustic analysis. [Work supported by ONR Code 334, SBIR program.]

2:45

4pSA6. Scattering of an obliquely incident plane wave from an immersed transversely isotropic rod covered by an isotropic cladding. Sina Sodagar, Farhang Honarvar (NDT Lab, Faculty of Mech. Eng., K. N. Toosi Univ. of Tech, P.O. Box 16765-3381, Tehran, Iran, honarvar@mie.utoronto.ca), and Anthony N. Sinclair (Univ. of Toronto, Toronto, ON M5S 1A4, Canada)

Valuable information is contained within the scattered pressure field from a submerged elastic cylinder. Many theoretical and experimental investigations on the scattering from elastic cylinders have been carried out during the past few years. This includes recent work of the authors on modeling of the scattered pressure field from immersed transversely isotropic cylinders and immersed isotropic clad rods. In the current paper, an analytical solution to the problem of acoustic wave scattering from an immersed transversely isotropic cylinder covered by an isotropic cladding is presented. The mathematical expressions are derived for the far-field backscattered amplitude spectrum, normal modes, frequency spectrum, and phase diagrams resulting from oblique insonification of an infinite, transversely isotropic cylinder covered by an isotropic cladding. The validity of the mathematical model is verified by comparing the results with data available in the literature for simpler cases, such as transversely isotropic rods and isotropic clad rods. The results are also consistent with elasticity theory.

3:00–3:15 Break

3:15

4pSA7. The effect of in-plane forces on sound radiation from convected fluid-loaded plates. Peter L. Schmidt and Kenneth D. Frampton (Vanderbilt Univ., 2400 Highland Ave., Nashville, TN 37212)

It is well established that fluid flow can have significant effects on structural acoustic behavior, along with the fact that induced coupling between discrete modes of vibration becomes significant as flow velocity increases. Work in this area has been confined to flows in air, over unloaded structures, with the effects on sound radiation efficiency, kinetic energy, and sound power radiation quantified and compared for various flow speeds. The purpose of this work is to study the effects that plane stress has on these structural acoustic phenomena. Theoretical development of the equations governing the vibration of a simply supported plate subjected to in-plane forces in an infinite baffle and an aerodynamic system that models a semi-infinite flowing medium along with the method for coupling these systems is included. Computational results are presented illustrating the behavior of the uncoupled coupled effects on the sound power radiation from the plate in both subsonic and supersonic flows, for a variety of loading cases.

3:30

4pSA8. An approximate method for the computation of the response of a line-driven, fluid-loaded plate with edges. Daniel DiPerna and Xingguang DiPerna (Naval Surface Warfare Ctr., 4500 MacArthur Blvd., West Bethesda, MD 20817)

An approximate method for predicting the response of a line-driven, fluid-loaded plate will be presented. The method is based on the free space Green's function using the rational function approximation developed by DiPerna and Feit. The approximte representations of the Green's function and pressure and velocity fields will be shown to satisfy the Helmholtz integral equation for the infinite plate. The Helmholtz integral equation

will then be used in conjunction with these approximate representations to solve the problem of a finit length plate. Results from the approximate solution will be compared to results from finite-element analysis.

3:45

4pSA9. Active control of vibration of fluid-loaded plates at moderately large amplitudes. Chiruvai P. Vendhan and A. N. Suryanarayana Prasad (Dept. of Ocean Eng., IIT Madras, Chennai, Tamilnadu-600036, India)

The problem of active control of vibrations has been studied by many in recent years. Gu and Fuller [J. Acoust. Soc. Am. **93**, 337–345 (1993)] considered the linear vibrations of a simply supported rectangular plate in a fluid half-space following the analysis of Sandman [J. Acoust. Soc. Am. **61**, 1502–1510 (1977)] and studied active control strategies to minimize the far-field sound-pressure levels. They evaluated the fluid loading using the classical solution for a baffled point source. Following a similar approach, Prasad *et al.* [Prasad *et al.*, J. Acoust. Soc. Ind. **33**, 535–540 (2005)] developed a Rayleigh-Ritz analysis capability for the linear vibration of fluid-loaded rectangular plates with a variety of boundary conditions. The present article is concerned with the nonlinear vibration of a simply supported isotropic square plate. The Duffing-type modal equation based on a single transverse mode approximation [AIAA J. **13**, 1092–1094 (1975)] has been chosen. The fluid loading is calculated using the procedure mentioned above. The equivalent linearization technique has been employed to reduce the nonlinear equation and solution in the form of frequency response function (FRF) obtained using an iterative procedure. The effect of piezoelectric patches mounted at the center of the plate is studied.

4:00

4pSA10. Study on the acoustic similitude of the underwater complex shell structure. Sande Wang, Jr. (Dalian Sci. Test & Technol. Inst., P.O. Box 67 Dalian, 116013, China)

The underwater noise that is produced by the underwater complex shell structure is a basic question to be solved for the noise control of an underwater structure. One way to study and control the underwater noise of the underwater complex shell structure is the experiment of vibration and sound radiation of all kinds of scaled-down models for its prototype. This method can predict vibration and sound radiation of an underwater complex shell structure, and provides the necessary parameters for the design of noise control. Using the finite-element method and the spherical function spread method, this paper investigates the similitude of coupled vibration and acoustic radiation for the typical structure of an underwater vehicle (elastic ring-stiffened cylindrical shell with hemispherical caps on the ends), and presents acoustic similitude conditions and relationships of the structures. At the same time, the acoustic scale effect and boundary effect, which are influenced by loss factors, shear and rotatory inertia, and surface sound reflection are investigated in detail. The results show that these effects can be controlled with similitude model designing and experiment design.

4:15

4pSA11. High-frequency broadband radiation from panels and periodic structures including the effect of structural power flow on directivity. Donald Bliss, Linda Franzoni (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708), and Pavel Danilov (Continuum Dynam., Inc., Ewing, NJ 08618)

In the high-frequency limit, a vibrating panel subject to spatially random temporally broadband forcing is shown to have broadband power and directivity properties that can be expressed analytically by a limited set of parameters. The radiated pressure-field power and directivity are characterized in terms of the structural power flow direction, wave-speed ratio, panel damping, and dimensionless frequency. The role of boundary conditions at the panel edge is illustrated, as are the effects of other types of forcing. Corresponding behaviors for panels coupled to form a periodic structure are also shown by approximate analytical solutions and numerical simulation. Overall, relatively simple broadband behaviors are revealed. For a single panel the analytical characterization is shown to be particularly straightforward in the high-frequency broadband limit. The relationship between radiation and structural power flow is potentially useful in energy-intensity-based prediction methods with application to aircraft interior noise prediction.

Session 4pSC

Speech Communication: Acquisition of Speech and Cross-Linguistic Studies (Poster Session)

Cynthia G. Clopper, Chair
Northwestern Univ., Dept. of Linguistics, 2016 Sheridan Rd., Evanston, IL 60208

Contributed Papers

All posters will be on display from 1:30 p.m. to 4:30 p.m. To allow contributors an opportunity to see other posters, contributors of odd-numbered papers will be at their posters from 1:30 p.m. to 3:00 p.m. and contributors of even-numbered papers will be at their posters from 3:00 p.m. to 4:30 p.m.

4pSC1. Rudimentary phrasal parsing in preverbal French-learning infants. David-Etienne Bouchard (Dept. of Linguistics, Univ. of Montreal, Montreal, QC H3T 1N8 Canada) and Rushen Shi (Univ. of Quebec at Montreal, Montreal, QC H3C 3P8 Canada)

The present study examines the initial parsing of noun phrases in infants. Participants were 8-month-old Quebec-French-learning infants. During Familiarization, infants heard two types of trials, one presenting a noun phrase sliced from a sentence (NP trial) and the other presenting the same word sequence (as in the NP trial) but sliced from another sentence in which these words were a part of a noun phrase and a part of a verb phrase, i.e., a non-unit trial. During Test, infants heard the original sentence containing the sliced NP and that containing the sliced non-unit. Two pairs of noun phrases and non-units were used in the experiment across two groups of infants. We predicted that if infants encoded the sliced stimuli in the NP trial as more coherent than those in the non-unit trial during Familiarization, they should listen longer to the sentences containing the NPs than those containing the same words but constituting non-units. Preliminary results show that infants listened significantly longer to the sentences containing the sliced noun phrases, even though the NP trial and non-unit trial during Familiarization involved exactly the same words. This finding suggests that salient prosodic markings enable French-learning infants to parse noun phrases from sentences.

4pSC2. Frequency factor in the segmentation of function words in French-learning infants. Rushen Shi, Melanie Lepage, Bruno Gauthier, and Alexandra Marquis (Dept. of Psychol., Univ. of Quebec at Montreal, Montreal, QC H3C 3P8, Canada)

Previous work showed that 8-month-old French-learning infants segment both highly and moderately frequent function words from continuous speech [Shi and Gauthier, J. Acoust. Soc. Am. **117**, 2426 (2005)]. In the present study we examine if high-frequency function words are segmented by even younger infants. Participants were 6-month-old Quebec-French-learning infants. As in the previous study, infants in the high-frequency condition were familiarized with a target functor (des or la) and were then tested with phrases containing the target versus those containing the nontarget. Thus, for infants familiarized to la, la-phrases were the target phrases and des-phrases were the nontarget phrases during Test. The reverse was the case when des was the familiarization target. The lower-frequency condition included a target functor mes versus ta (prosodically analogous to the functors in the high-frequency condition). The results show that 6-month-olds listened significantly longer to phrases containing the familiarized target than those containing the nontarget during Test, suggesting that they segmented the target. No difference was observed in the lower-frequency condition. Thus, while 8-month-olds segmented both high- and lower-frequency functors [Shi and Gauthier, J. Acoust. Soc. Am. **117**, 2426 (2005)], 6-month-olds showed evidence of segmentation only for high-frequency functors. These findings suggest that frequency is a determining factor in the development of the segmentation of function words.

4pSC3. Learning to categorize lexical tones in sentences with variable narrow focus: A neural network approach. Bruno Gauthier, Rushen Shi (Dept. of Psych., Univ. of Quebec in Montreal, C.P. 8888, Succ. Ctr. ville, Montreal, QC, Canada, H3C 3P8, gauthier.bruno@courrier.uqam.ca), and Yi Xu (Univ. College London, London, UK, NW1 2HE)

The surface F0 patterns of lexical tones contain variability from many different sources, which raises questions as to how infants can learn tones based on such highly variable input. A recent modeling study showed that the velocity profiles of F0 (i.e., first derivatives, or D1) enable naive learners to successfully categorize the four Mandarin tones despite cross-speaker and contextual variations [Gauthier, *etal.* ''Recognising tones by tracking moments—How infants may develop tonal catagories from adult speech input,'' in *Proceedings ISCA workshop on plasticity in Speech Perception*, edited by V. Dellwo (UCL Publications, London, UK, 2005) pp. 72–75]. The present study explores the robustness of D1 for categorizing Mandarin tones in sentences said with different focus conditions. As a contrastive communicative function, narrow focus within a multi-word utterance introduces extensive variability to F0, making tone learning a more challenging task. Using multi-speaker productions of utterances with systematically varied tones and focus [Xu, ''Effects of tone and focus on the formation and alignment of contours. J. Phonet. **27**, 55–105 (1999)], self-organizing neural networks were trained with both syllable-sized D1 and F0 as input, with no special treatment for different focus conditions. In the testing phase, novel tokens were tested for tonal categorization. The results revealed that D1 yielded overall excellent categorization, far superior than F0. Detailed analyses showed that with D1, performance of tonal recognition dropped only in post-focus regions. These findings indicate that successful tone learning can be achieved with velocity profiles as input despite variability introduced by focus.

4pSC4. The effect of concurrent serial digit recall on speech processing in children. Sangsook Choi, Dawna Lewis, Brenda Hoover, and Pat Stelmachowicz (Boys Town Natl. Res. Hospital, 555 N 30th St., Omaha, NE 68131, chois@boystown.org)

Developmental changes in speech perception in a dual task paradigm were investigated in 32 children (7–14 years). Monosyllabic words were presented in noise with or without the concurrent serial recall of visually presented digits (3 or 5). Children were instructed to focus primarily on the listening task, placing secondary effort on digit recall. In both single and dual tasks, word recognition and digit recall improved with age. However, word recognition improved and digit recall decreased in the dual task

condition (*re*: the single task) regardless of age. Greater improvements in word recognition were found for younger children during the dual task in the 3 digit condition. Consequently, the age differences in word recognition observed in the single task diminished during the dual task. Results suggest that the competing digit task might have helped children to optimize attention resulting in improved speech processing. In a follow-up experiment, children will be asked to focus attention on the digit task to determine if the instructions were responsible for observed improvements in word recognition in the first experiment. Results of these studies may have clinical implications for perceptual training methods for children with language delays or hearing loss.

4pSC5. Children's production of prosody: Disambiguating sets of compound nouns. Michiko Yoshida and William F. Katz (Univ. Texas—Dallas, Callier Ctr. for Commun. Disord., 1966 Inwood Rd., Dallas, TX 75235)

Studies have suggested that although young children may perceive prosodic patterns marking structural boundaries in ambiguous sentences (e.g., distinguishing ''pink and green, and white'' from ''pink, and green and white''), children do not necessarily produce these prosodic patterns. However, most of these studies did not test perception and production in the same group of subjects. Therefore, the relation between perception and production for a given linguistic structure could not be clearly established. Yoshida and Katz [''Children's use of prosody to identify ambiguous sets of compound nouns,'' J. Acoust. Soc. Am. **116**, 2645 (2004)] had young children and adults listen to resynthesized speech versions of ''sun/flower/pot.'' F0 and duration patterns were parametrically manipulated to provide a continuum of interpretations between ''sun, flowerpot'' and ''sunflower, pot.'' Results indicated that children showed adultlike processing of duration and F0 cues. The current study reports acoustic data for productions made by these same subjects (elicited prior to the perceptual experiment). Subjects played a naming game with pictures that depicted either (a) a sunflower and a pot or (b) a sun and a flowerpot. Acoustic analyses support the idea that children and adults control both duration and F0 to disambiguate their utterances. Results suggest that previous studies may have underestimated young children's capacity for producing prosodic structures.

4pSC6. Consonant-vowel place linkages in the babbling of 6-, 9- and 12-month-old learners of French, English, and Mandarin. Sara Giulivi (Haskins Labs., 300 George St., New Haven, CT 06511 and Univ. of Florence), D. H. Whalen, Louis M. Goldstein (Haskins Labs., New Haven, CT 06511), and Andrea G. Levitt (Haskins Labs., 300 George St., New Haven, CT 06511)

The place of articulation of consonants and vowels in CV syllables in babbling has been claimed to be linked, such that front vowels predominantly occur with labial consonants, mid-vowels with coronals, and back vowels with velars[P. F. MacNeilage, Behav. Brain Sci. **21**, 499–546 (1998)]. Support for this claim has come from a variety of languages, but a further implication has not been tested: If children depend on this frame/content approach less as they develop, there should be a trend away from it as language mastery increases. Here, we examined the babbling of six English-learning, five French-learning, and five Mandarin-learning infants, sampled at 6, 9, and 12 months of age. There was only moderate support for the frame/content account in the overall ratios, with velars occurring noticeably less often than predicted with back vowels. Further, there was no pattern toward a decrease in evidence for frame/content organization across these 6 months. Together with other theoretical considerations, these results cast doubt on the utility of the frame/content account for language acquisition.

4pSC7. Fricative development in English-learning children. Joanna H. Lowenstein and Susan Nittrouer (Dept. of Speech and Hearing, Ohio State Univ., 1070 Carmack Rd., 110 Pressey Hall, Columbus, OH 43210)

Acoustic studies of school-aged children's fricatives have demonstrated that those gestures are not as precisely specified as in adult speech [e.g., S. Nittrouer, J. Acoust. Soc. Am. **97**, 520–530 (1995)]. The current study examined the development of English fricatives at younger ages by audiotaping seven children at 2-month intervals from the onset of first words (around 14 months) to the onset of three-word sentences (around 30 months). Two measures were made: the numbers of words produced with intended /s/ and /sh/ in initial and final word positions were counted, and first through fourth spectral moments were derived for each /s/ and /sh/ noise at each recording session. Fricatives were rare until after 24 months of age. /s/ was produced earlier than /sh/, and /s/ was produced in word-final before word-initial position. The first moments of /s/ and /sh/ were similar, and both resembled those of /sh/ spectra produced by slightly older children. Spectral skewness and kurtosis were produced with the same degree of separation for /s/ and /sh/ as those of adults, but values were more negative. Results suggest that these children were starting to develop mature spectral shapes for /s/ and /sh/. [Work supported by NIDCD Grant No. DC-00633.]

4pSC8. Developmental changes in the acoustic vowel space of Mandarin-learning children before 2 years of age. Li-mei Chen (Dept. of Foreign Lang., Natl. Cheng Kung Univ., 1 Univ. Rd., Tainan City 701, Taiwan)

Vocalic productions of Mandarin-learning infants aged 0–24 months were audiotaped and analyzed. Major findings in this longitudinal study are (1) There is no obvious change in the range of $F1$ and $F2$ values during these 2 years; (2) Midcentral vowels occurs most frequently for infants younger than 6 months. Vowel /a/occurs most frequently in infants 7 to 24 months old; (3) The frequently occurring vowel categories show steadier pattern of vowel formant frequencies in the second year; (4) Other vowels develop rapidly from 18 months on. However, the newly developed /i, u/ are difficult to identify because of the low amplitude and short duration at this stage of development; (5) Diphthongs, triphthongs, and vowels with nasal endings appear. However, they are unstable, weak, abrupt, and incomplete in the transition. These cause great difficulty in perceptual judgment and acoustic analysis. (6) The individual differences found in the first year became less obvious in the second year. [Investigation supported through funds from National Science Council in Taiwan, Republic of China (NSC 90-2411-H-006-015) (NSC 92-2411-H-006-021) (NSC 93-2411-H-006-024) and a faculty grant from National Cheng Kung University.]

4pSC9. Development of phrase boundary effects. Sara Zepralka and Betty Kollia (Commun. Disord., William Paterson Univ., Wayne, NJ 07470)

In this study we investigate the acquisition and development of phrase-final lengthening in children ages 3 to 7 years of age. Fifteen subjects participated in the study. Preliminary data indicate that subjects produced more adultlike phrase-final lengthening at younger ages, with phrase-final

lengthening duration decreasing with increasing age. Statistically significant phrase-final lengthening was observed in 9 of the 15 subjects. Suggestions for this pattern of acquisition and future research are discussed.

4pSC10. Effects of experience on speech perception in noise. Ratree Wayland, Bin Li, LaTosha Csonka (Program in Linguist., Univ. of Florida, Gainesville, FL 32611-5454), and James Harnsberger (Univ. of Florida, Gainesville, FL)

This study will report on an experiment designed to test the hypothesis that the benefit derived from naturally clear speech is commensurate with listeners' increased exposure to the target language. Research participants were 37 native Chinese speakers, 15 native Spanish speakers, 16 native Thai speakers, and 30 native speakers of English served as control subjects. All participants were asked to listen to four lists of English sentences identical to those used in Bradlow and Bent (2002). These sentences were produced in a conversational and a clear speaking style by a male speaker. All sentences were then normalized for their rms amplitude and mixed with white noise to generate two sets of stimuli (for each speaking style) with a signal-to-noise ratio of -4 and -8 dB, respectively. The number of key words ($N=50$ for each of the four lists) correctly identified by each listener were obtained. The statistical analyses to be performed on the data will focus on comparing the performance of subjects, both within and across language group, whose experience with English differs significantly.

4pSC11. Perceptual assimilation in context and isolation. James Harnsberger, Ratree Wayland, and Jenna Silver (Univ. of Florida-Gainesville, 356 Dauer Hall, P.O. Box 117420, Gainesville, FL 32611)

A study was conducted to examine the effects of sentence context and speaking rate in the perceptual assimilation of non-native speech sounds. Native speakers of American English were presented with two voicing contrasts, [k]-[g] and [t\sh]-[d\yog], produced by two Hindi speakers. They appeared word-initially before [a] and within a short sentence frame. These sentences were produced at three speaking rates (slow, normal, and fast). Listeners were administered a categorial AXB discrimination test and a forced-choice identification test. The results were compared to those collected in a previous study that employed the same contrasts appearing in isolated words and produced by the same talkers. The results showed that the discrimination of these speech sounds in a sentence frame was significantly poorer relative to the isolation context. Moreover, the modal native consonant used to classify the Hindi stops varied depending upon the sentence context variable: in isolation, more discriminable uncategorizable assimilation types were elicited, while in a sentence frame, Hindi contrasts tended to be classified using a single English (voiced) consonant. Speaking rate only showed modest effects on proportion of responses represented by the modal native consonants. These results highlight the sensitivity of perceptual patterns elicited in the laboratory to experimental variables.

4pSC12. Temporal aspects of voicing in Russian and English. Roza Zak and Carole Gelfer (William Paterson Univ., 300 Pompton Rd., Wayne, NJ, 07470, zakr@student.wpunj.edu)

The study compares temporal aspects of voicing in Russian and English. Specifically, it examines whether Russian-English bilinguals who have learned English as a second language (L2) in early childhood are able to differentiate Russian and English consonants /p/, /b/, /t/, /d/, /k/, /g/ according to voice onset time (VOT). Monolingual Russian-speaking 3-year-old children's VOTs will be compared to those of monolingual Russian-speaking adults. Bilingual (Russian and English) 9-year-old children's VOTs will be compared to monolingual 9-year-old children's VOTs. It is predicted that 9-year-old bilingual children will demonstrate appropriate voicing boundaries in English; however, their production of Russian VOTs will differ from the Russian monolingual production.

4pSC13. Voice onset time of bilingual English and French-speaking Canadians. Carol Fowler, Sarah Rowland, David Ostry (Haskins Labs. 300 George St., New Haven, CT 06511, sarah.rowland@uconn.edu), Valery Sramko (McGill Univ. Montreal, QC, H3A 1B1 Canada), and Pierre Halle (CNRS/Sorbonne-Nouvelle Paris, France)

There is evidence of voice onset drift when bilinguals are submerged in cultures that are dominated by their second language (L2). Sancier and Fowler [Gestural drift in a bilingual speaker of Brazilian Portuguese and English," J. Phonet. **4**, 421-436 (1997)] found that voiceless stop voice onset times (VOTs) of a bilingual speaker of Brazilian Portuguese and English were longer after months spent in the US than after time spent in the speaker's first language (L1) context. Flege ["The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification," J. Phonet. **15**, 47–65 (1987)] studied French L1 English L2 bilinguals and English L1 French L2 bilinguals who had spent about 12 years living in the L2 dominant culture. The French L1 English L2 bilinguals VOTs were longer than the VOTs of French monolinguals. The English L1 French L2 bilinguals VOTs were shorter than the VOTs of English monolinguals. In the present study the nature of the experience underlying VOT drift was investigated. Specifically, the aim of the current study was to see whether exposure to another language with different VOTs affects VOT production in the first (or only) language or if VOT production drift is dependent on speaking the non-native language. Productions from English and French monolinguals and from English-French bilinguals were compared across varying linguistic contexts and order of acquisition.

4pSC14. Perception of French stops by French and American English speakers. Nassima Abdelli-Beruh (Dept. of Speech-Lang. Pathol. and Audiol., and Dept. of Speech and Hearing, New York Univ., 719 Broadway, Ste. 200, New York, NY 10003) and Steve Chartan (John L. Miller Great Neck North High School, Great Neck, NY)

This study examined how native French and native American English speakers perceive edited and unedited word-initial stops produced by two monolingual Parisian speakers. Most words used in this study exist in both languages, yet are produced differently. Listeners were asked to identify the initial stops of French monosyllabic words recorded in isolation. The words were unedited (condition A), edited (condition B), and whispered (condition C). In condition A, words were produced in isolation at modal vocal intensity. In condition B, voicing lead was removed from words beginning with /b, d, g/, and in condition C, words were naturally whispered. Participants were asked to listen to each stimulus and then identify the initial consonant (b, d, g, p, t, k). Statistical analyses show that native French speakers performed significantly better than native American English speakers on identification of the word initial consonant in condition A (unedited words spoken at modal vocal intensity). In condition B (voicing lead removed from words produced at modal vocal intensity), native American English speakers performed significantly better than the native French speakers. Finally, both language groups were equally impaired in condition C (whispered speech). Errors were analyzed as a function of the listener's language phonetic inventory.

4pSC15. The perception of Mandarin tone contrasts: A cross-linguistic study. Yueh-chin Chang (Grad. Inst. of Linguist., Natl. Tsing Hua Univ., Hsinchu, Taiwan, ycchang@mx.nthu.edu.tw), Pierre Halle (CNRS-Paris V (UMR 8581), Boulogne-Billancourt, France), and Catherine T. Best (Univ. of Western Sydney, Milperra NSW, Australia)

Previous studies of the perception of tone continua found that Mandarin listeners perceive tones more categorically than French listeners, whose performance is essentially psychophysically motivated. In this study, we examined native-language effects on the perception of Mandarin

tone continua by Cantonese, Thai, and Japanese listeners. Cantonese distinguishes six tones: high-rising, low-rising, low-falling, and high-, mid-, and low level. Thai has five tones: rising, falling, and high-, mid-, and low level. Standard Japanese has a word pitch accent with at most one LH and one HL pitch movement per word. Listeners of these languages were tested on three Mandarin tone continua constructed from natural utterances within carrier sentences. Listeners were run on AXB identification and discrimination tests on isolated syllables. Japanese, Thai, and Cantonese listeners, in that order, perceived tones increasingly categorically. Cantonese and Thai listeners were more sensitive to tone-contour differences than Japanese listeners, although less than Mandarin listeners. The Cantonese and Thai data show that both phonological and phonetic properties of the native language affect the perception of the non-native Mandarin tones. The findings are compared with prior cross-linguistic research, and discussed with reference to the suprasegmental phonological differences across languages.

4pSC16. Return of the weckud wetch: Rapid adaptation to a new accent. Neil P. Bardhan, Richard N. Aslin, and Michael K. Tanenhaus (Brain & Cognit. Sci., Univ. of Rochester, Meliora Hall, Rochester, NY 14627)

This study investigated listeners' adaptation to a new accent within a portion of the vowel space of their native language. Previous work has shown that such adaptation requires only brief exposure [Maye *et al.*, "In search of the weckud wetch: Online adaptation to speaker accent," CUNY Conference on Sentence Processing, Cambridge, MA (2003); Clarke & Garrett, "Rapid adaptation to foreign-accented English," J. Acoust. Soc. Am. **116**, 3647–3658 (2004)]. However, neither adaptation involved visual referents and the postadaptation assessment involved either a lexical decision task or a cross-modal lexical matching task. In the present study, English-speaking listeners were exposed to a list of lexical items in which front vowels were lowered (e.g., "witch" became "wetch") but back vowels were not (e.g., moon). During adaptation, subjects selected an object from a visual array after hearing a naturally produced, accented form of the object's name. In a postadaptation lexical decision task subjects were more likely to accept the accented form as a word than before training. Also examined were the effects of generalization across the lexicon to untrained words and throughout the vowel space to include back vowels. This paradigm of lexical adaptation lends itself to further studies using an eye-tracking measure to assess dialect influences on lexical access in spoken word recognition. [Work was supported by NIH 5R01DC005071-05 and NIH 5T32DC000035-12.]

4pSC17. Sequential contrast versus compensation for coarticulation in Japanese versus English. Daniel Mash and Shigeto Kawahara (Ling. Dept., Univ. of Massachusetts, Amherst, MA 01003-9274)

English listeners identify a stop ambiguous between /t/ and /k/ more often as "k" after [s] than after [sh] (Mann and Repp, J. Acous. Soc. Am. **69**, 548–558, 1981). This bias could reflect compensation for coarticulation, which undoes the fronting of a stop expected next to [s], or instead sequential contrast, in which an intermediate F3 value sounds lower and more "k"-like after [s]'s high-frequency energy concentration. Japanese may distinguish these accounts' predictions. The Japanese sibilant fricative is pronounced [s] before [u] but [sh] before [i]. These vowels delete between two voiceless consonants, so [s] or [sh] can occur before [t] or [k] in surface forms. If Japanese listeners treat [s] followed by a stop from a [t-k] continuum as containing an intervening [u] and [sh]-[t-k] as containing an intervening [i], sequential contrast but not compensation for coarticulation predicts that they should identify more of the continuum as "t" after [s] because [u] concentrates energy at lower frequencies than [i]. However, if they instead judge the stop in terms of the fricative that occurs next to it on the surface, then they should respond like English listeners, and their behavior is predicted equally well by sequential constrast and compensation for coarticulation. [Work supported by NIH.]

4pSC18. Perception and production of English vowels by native speakers of Mandarin. Karece Lopez (Dept. Speech, Commun. Sci., & Theatre, St. John's Univeristy, Queens, NY 11439), Yan Helen Yu (CUNY Grad. School, New York, NY), and Fredericka Bell-Berti (St. Johns Univ., Jamaica, NY 11439)

This study examines the relationship between the identification of English vowels by 48 Mandarin ESL speakers and their production of these same vowels, and how factors of AoA, LoR, general and English education, and language spoken at work and home influence that relation. The Mandarin ESL speakers produced eight repetitions of the following words, heed, hid, head, had, hod, hawed, hood, who'd, hud, and heard, and then identified tokens of each of these words produced by two native speakers of Greater Metropolitan New York City English. Native speakers of American English identified the productions of the Mandarin ESL speakers to provide a measure of their ability to produce these vowels. We hypothesize that general and English educational background and language use influence the L2 English speakers' perception and production of English vowels, and that the more accurately the L2 listener distinguishes a vowel contrast, the more distinctly he/she produces that contrast.

4pSC19. Pronunciation of English vowels of native Turkish speakers. Janine Sadaka and Manwa Ng (Commun. Sci. and Disord., Long Island Univ., Brookville, NY 11548-1300)

Foreign accent can be defined as the allophonic variations in the target language (L2) under the influence of the first language (L1). This is mainly caused by the incorrect placement of articulators during production of the vowels and consonants of L2. As direct acoustic correlates of articulatory gestures, formants can be a valuable tool to indicate the amount of misarticulation, and thus amount of accent traits. The present study attempts to investigate the amount of accent in Turkish-English bilingual individuals producing English vowels. The first two formant frequencies (F1 and F2) of eight English monophthongs produced by ten male and ten female native Turkish will be analyzed. The bilingual speakers have been using English and Turkish in their daily communication. All English vowels are produced three times in the /hVd/ context. Five data points will be selected evenly along each vowel segment, and the average F1 and F2 values will be measured. The average F1 and F2 values thus measured will be compared with those of native English speakers as reported by Peterson and Barney (1952). It is hypothesized that the F1 and F2 values of the accented English vowels will be slightly different from those produced by native English speakers.

4pSC20. Vowel errors and sentence intelligibility in Spanish-accented English speech. Jonathan Dalby (Dept. of Audiol. and Speech Sci., Indiana Univ.—Purdue Univ., Fort Wayne, 2101 E. Coliseum Blvd., Ft. Wayne, IN 46805, dalbyj@ipfw.edu)

It has been shown that foreign-accented speech is more difficult for listeners to process than native speech when presented under noisy listening conditions. It has also been shown that errors in vowel productions are more damaging to connected speech intelligibility than are errors or distortions in consonant production. This study compared the intelligibility of Spanish-accented English vowels in isolated words with that of short samples of connected speech. Five native Spanish speakers who had learned English as adults read a list of 42 monosyllabic English words selected so that 14 Spanish-accented English monophthongs could be presented to native listeners in a three-alternative forced-choice identification task. In each trial the target and two nearby (in the vowel space) distracters were presented. Talkers also read a set of 20 phonetically balanced sentences. Ten native English listeners identified the vowels in the isolated word task and judged the intelligibility of the sentences in a simple write-down procedure. While overall sentence intelligibility scores are not very well predicted by word intelligibility scores when the full set of vowel contrasts is included, predictability improves when a subset of the vowels is chosen. [Work supported by Purdue Research Foundation.]

4pSC21. Formant movement and duration in second-language vowel learning. Paul Iverson and Bronwen G. Evans (Dept. of Phonet. and Linguist., Univ. College London, 4 Stephenson Way, London NW1 2HE, UK)

This study tested native speakers of Norwegian, German, French, and Spanish on their recognition and production of L2 English vowels. A goodness-optimization method was used to construct phonetically detailed perceptual vowel space maps for each listener's L1 and L2 (English). These maps comprised best exemplar locations in a multidimensional stimulus space that included $F1$ and $F2$ target frequencies, $F1$ and $F2$ movement, and duration. The maps were then compared to assessments of each individual's recognition accuracy for naturally spoken English vowels and their degree of accent for spoken English. The results demonstrated that all four language groups used phonetic detail to recognize English vowels; individuals had higher recognition accuracy for natural English vowels when their best exemplars had formant movement and duration contrast that were more like those of L1 English speakers. However, their use of phonetic detail was not entirely predictable by their L1 background. It thus appears that individuals with a range of L1 backgrounds form phonetically detailed category representations during L2 learning, and that their representation of phonetic detail has practical implications for recognizing natural L2 speech.

4pSC22. Training Chinese dialect listeners to perceive English /n/ and /l/. Bin Li and Ratree Wayland (Linguist. Program, Univ. of Florida, 4131 Turlington Hall, P.O. Box 115454, Gainesville, FL 32611-5454, bli@ufl.edu)

Dialects and Mandarin subdialects in certain areas of southern China do not distinguish between /n/ and /l/. Native speakers of these dialects usually have difficulty distinguishing these two sounds when learning English, even after years of experience with the language. The goal of the present research is to develop a set of procedures that would be effective in training these speakers to learn this non-native speech contrast in English. Forty listeners were recruited from 3 dialect areas (6 Gan, 11 Jianghuai, and 23 Xinan). Listeners were instructed to identify the initial consonant in an English word spoken by 3 native speakers of American English. A pre- and a post-test design were used to assess the effect of a 4-day training. A generalization test containing the same stimuli spoken by 2 new talkers was also carried out on the same day of the post-test. Overall, subjects' performance improved significantly from the pre-test to the post-test and there was no significant group difference. Additionally, there was no significant difference between their performance of /n/ and /l/ either before or after training. Vowel contexts and syllable structures are also found to have affected listeners' perception in the post-test.

4pSC23. Effects of dialect variation on speech intelligibility in noise. Cynthia G. Clopper and Ann R. Bradlow (Dept. of Linguist., Northwestern Univ., 2016 Sheridan Rd., Evanston, IL 60208, c-clopper@northwestern.edu)

Recent research has shown that listeners can classify talkers based on their regional dialect of American English (AE). The present experiment explored the effects of regional dialect variation on speech intelligibility. Participants from various AE dialects transcribed meaningful English sentences mixed with noise. The sentences were produced by male and female talkers from four different AE regional dialects. At a favorable signal-to-noise ratio (-2 dB), intelligibility of General American and Southern talkers was significantly better than Northern and Mid-Atlantic talkers, and was significantly better for the Northern than the Mid-Atlantic talkers. At a more difficult signal-to-noise ratio (-6 dB), intelligibility was equally poor for the Southern, Northern, and Mid-Atlantic talkers, but remained significantly better for the General American talkers. No effects of listener dialect were observed. Thus, speech intelligibility in noise was generally poorer for more phonologically marked dialects (e.g., North, Mid-Atlantic) than less-marked dialects (e.g., General American). However, noise had differential effects on dialect interference; intelligibility of marked dialects was graded at easier signal-to-noise ratios, but marked dialects were equally difficult to transcribe at unfavorable signal-to-noise ratios, suggesting that dialect information may be conveyed by aspects of the signal that are relatively vulnerable to perceptual disruption by noise.

4pSC24. Consonantal context and experience effects on American L2-learners' perception of French vowels. Erika S. Levy (Teachers College, Columbia Univ., Biobehavioral Sci., Box 180, 525 W. 120th St., New York, NY 10027)

According to Best's [Best, C. T. (**1994**). "The emergence of native-language phonological influence in infants: A perceptual assimilation model," in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. Goodman & H. C. Nusbaum (MIT Press, Cambridge, MA); Best, C. T. (**1995**) "A direct realist view of cross-language speech perception," in *Speech Perception and Linquistic Experience Issues in Cross-Language Research*, edited by W. Strange (York Press, Baltimore, MD)] Perceptual Assimilation Model (PAM), listeners' perceptual assimilation of unfamiliar speech sounds predicts their discrimination accuracy of these sounds. This study applied the PAM to second-language (L2) learners. Three groups of American English (AE) listeners performed a perceptual assimilation (PA) task involving Parisian French (PF) vowels /y-œ -u-o-i-ɛ-a/ and a categorial discrimination task involving French front versus back rounded pairs /y-u/, /y-o/, /œ -o/, /œ -u/, front rounded versus front unrounded /y-i/, /y-ɛ/, /œ -ɛ/, /œ -i/, and front rounded /y-œ /. Vowels were in bilabial /rabVp/ and alveolar /radVt/ contexts. PF front-rounded vowels were perceptually assimilated overwhelmingly to back AE vowels, resulting in more discrimination errors for front-versus back-rounded PF pairs than for front-unrounded versus rounded pairs. Listeners with formal French instruction fared similarly to listeners with no experience. Extensive immersion correlated with fewer errors than formal experience, although /u-y/ discrimination remained relatively poor for this group. More errors occurred in alveolar than bilabial context. Thus, native-language allophonic variation may explain context-specific L2 perception. Using a percent overlap method, a significant correlation was found between PA overlap and discrimination errors for all groups, suggesting that the PAM may be extended to L2-vowel learning. [Work supported by NIH.]

Session 4pSP

Signal Processing in Acoustics and Underwater Acoustics: Optimum and Robust Signal Processing in Uncertain and Random Environments III

R. Lee Culver, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804-0030

Leon H. Sibul, Cochair
Pennsylvania State Univ., Applied Research Lab., P.O. Box 30, State College, PA 16804

Chair's Introduction—1:15

Invited Papers

1:20

4pSP1. Estimating mutual information for the underwater acoustic channel. Paul Gendron (Naval Res. Lab., Acoust. Div., 4555 Overlook Ave. SW, Washington, DC 20375)

Estimates of the upper and lower bounds on the information rates attainable for acoustic communication systems with white Gaussian sources under uncertain channel conditions are provided that account for finite duration signaling and finite coherence time. For the case where the coherence time of the channel exceeds that of the packet duration the estimators are computationally efficient with the Levinson recursion and provide a means of tuning the accuracy of the estimates to fit a given computational budget. The reduction in the information rate relative to the known channel case is quantified in terms associated with the duration of signaling, the posterior covariance of the channel, and the prior covariance of the channel. For scenarios where the prior channel covariance is known exact bounds on the information rates are computable else the estimators provide confidence intervals for these rates. The estimators are tested on 18 kHz at sea data collected in shallow water north of Elba Italy on a 1.8-m aperture vertical array and on very shallow water sites at Keyport Washington. The effects of temporal coherence degradation as well as multi-path spread are quantified. [Work supported by the Office of Naval Research.]

1:45

4pSP2. Robust processing against uncertain reverberation spectra in shallow-water broadband active sonars. Peter G. Cable (BBN Technologies, Old Mystic Mill, 11 Main St., Mystic, CT 06355)

Broadband low-frequency active sonars operating in shallow water must work against reverberation having uncertain spectral color. Under these conditions prewhitening of the background is appropriate for detection and estimation processing. To assess the importance of prefiltering, in this talk signal-to-background performances of energy detected data subjected to no prewhitening, prewhitening, and Eckart filtering are compared for typical observed shallow-water reverberation spectral shapes. Adaptive prewhitening algorithms are described and their performances analyzed in terms of loss associated with spectral estimation errors. Modeled detection performance is compared with measured results obtained from Area Characterization Test III performed in the Korea Strait. Finally, the relevance of adaptive prewhitening to estimation for classification is also described. [Work supported by ONR and DARPA.]

2:10

4pSP3. Measured and modeled multirank signals in a range-dependent environment. Kevin D. Heaney (OASIS, Inc., 11006 Clara Barton Dr., Fairfax Station, VA 22039, oceansound04@yahoo.com)

Environmental variability can take the form of sound-speed fluctuations, range- and azimuthally dependent geoacoustics, and strong bathymetric gradients. In this paper we examine the rank of strong broadband tones in a strongly range-dependent environment and the subsequent effect on adaptive processing. Range-dependent transmissions off the coast of Kauai were recorded during the Basin Acoustic Seamount Scattering EXperiment (BASSEX04) in September of 2004. Broadband transmissions (27-s M-sequence, 75-Hz center frequency) of the North Pacific Acoustic Laboratory (NPAL) source-mounted bottom in 900 m of water off the coast of Kauai were recorded using a 200-m towed line array. The single-element signal-to-noise ratio (SNR) is on the order of 15 dB before matched filtering, so there are no issues of adequate SNR for adaptive beamforming. With such a strong single source, the spread in eigenvalues will be studied as an indication of the effect of motion, environmental variability, and multipath propagation. In this environment, the rank of the CSDM is significantly higher than can be modeled using 2D range-dependent parabolic equation modeling, indicating that environmental variability (scattering, 3D acoustics) dominates over motion and multipath effects (which are well handled in the PE modeling).

2:35

4pSP4. Full-field likelihood ratio detection for multistatic sensors in uncertain environments. Lawrence Stone and Bryan Osborn (Metron, Inc., 11911 Freedom Dr., Ste. 800, Reston, VA 20190, osborn@metsci.com)

As small networked sonobuoys become cheaper and more viable in the submarine warfare community, there is a growing demand for techniques to automatically process detection data from dense fields containing O(100) or more sonobuoys. The presentation describes a Bayesian framework for the incoherent integration of automatic detections from these dense sonobuoy fields. This framework recursively integrates likelihood ratio functions to form a cumulative posterior over a target state space. To be most effective, these methods exploit either empirical sensor performance models or physics-based signal excess sensor performance models. The presentation describes the use of each type of performance model and presents one method for handling uncertainties associated with signal excess predictions. The method presented starts with a prior on a set of possible environmental conditions. When a target is present in the detection data, the cumulative likelihood ratio will peak in the environment most consistent with the true environment.

Contributed Papers

3:00

4pSP5. The estimated ocean detector (EOD) for Gaussian signal and noise. Jeffrey A. Ballard, Colin W. Jemmott, Leon H. Sibul, R. Lee Culver, H. John Camin, and David L. Bradley (Appl. Res. Lab. and Grad. Program in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804, jab795@psu.edu)

Our overall goal is to develop a method for incorporating knowledge of the environment, and in particular, environmental variability, into the signal processor. This talk begins with the assumption that all available information about the environment has been used to compute the probability density functions (pdfs) of relevant received signal parameters (e.g., amplitude or time spread). For the exponential class of pdfs that arises from applying the maximum entropy method to random signal parameters, the maximum likelihood detectors (MLD) can be implemented as an estimator-correlator [Schwartz, "The estimator-correlation for discrete-time problems," IEEE Trans. on Information Theory. **23**, 93–100 (1977)]. The MLD detector correlates received data with the conditional mean estimate of the signal; hence we have named it the estimated ocean detector (EOD). Here, we present the general structure of the EOD and explore EOD performance for Gaussian signals and noise, which are members of the exponential class. The resulting implementation produces a detection statistic that is the weighted sum of an energy detector and a correlation receiver, with the weights calculated from the variances of the noise and signal. EOD performance bounds are calculated using completely known and completely unknown oceans. [Work supported by The Undersea Signal Processing Program, Office of Naval Research.]

3:15

4pSP6. Performance of the estimated ocean detector (EOD) against non-Gaussian signal and noise. Colin W. Jemmott, Jeffrey A. Ballard, Leon H. Sibul, R. Lee Culver, H. John Camin, and David L. Bradley (ARL and Grad. Program in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804, cwj112@psu.edu)

Our overall goal is to develop a method for incorporating knowledge of environmental variability into a passive sonar detector. An earlier talk presented the derivation of the estimated ocean detector (EOD), which is the maximum-likelihood detector for the exponential class of received signal parameter probability density functions (pdfs). A particular implementation of the EOD for Gaussian signal and noise (members of the exponential class) was shown and performance bounded. This talk evaluates the performance of the Gaussian EOD under nonideal conditions, including inaccuracies in the estimates of signal and noise variance, and examines robustness under received signal distributions that contain nonzero skew and kurtosis. To assess detection performance for non-Gaussian signals, receiver operating characteristics are computed for received signal pdfs obtained by applying the maximum entropy method to fit an exponential class pdf to experimental data. Gaussian EOD performance for non-Gaussian signal and noise conditions is compared to the classical envelope and correlation detectors, which form the limiting cases of the EOD. [Work supported by The Undersea Signal Processing Program, Office of Naval Research.]

3:30

4pSP7. A method to predict active sonar detection range in uncertain shallow water environments. Peter G. Cable (BBN Technologies, Old Mystic Mill, 11 Main St., Mystic, CT 06355)

Knowledge of maximum detection range is important for the appropriate operation of search sonar. For example, effective spacing of multistatic sonar sources and receivers will be influenced by good estimates of achievable detection range. But predicting detection range for low-frequency active sonar in shallow water can be problematic in the face of inadequate environmental characterization. Thus, incomplete knowledge of bottom sediment geoacoustics and bottom scattering strength, and absence of a good understanding of bottom reverberation mechanisms, will greatly hamper the accurate prediction of dependable detection ranges. Given a bistatic sonar configuration in shallow water, a method is proposed and described that exploits the observed transition time from reverberation to noise-limited conditions to simply and robustly predict maximum detection range for specified target strength. Modeling and analysis suggest that estimates of maximum detection range within 25% accuracy can be readily achieved.

Session 4pUW

Underwater Acoustics: High-Frequency Acoustic Propagation and Applications II

Mohsen Badiey, Cochair
Univ. of Delaware, College of Marine Studies, Robinson Hall Rm. 114, Newark, DE 19716

James C. Preisig, Cochair
Woods Hole Oceanographic Inst., Dept. Applied Ocean Physics and Engineering, Bigelow 207, Woods Hole, MA 02543-1053

Chair's Introduction—1:00

Invited Papers

1:05

4pUW1. Spatial diversity in passive time reversal communications. H. C. Song, W. S. Hodgkiss, W. A. Kuperman, T. Akal (MPL/SIO, UCSD, La Jolla, CA 92093-0238), and M. Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

A time reversal mirror exploits spatial diversity to achieve spatial and temporal focusing, a useful property for communications in an environment with significant multipath. This paper presents the impact of spatial diversity on passive time reversal communications between a single probe source and a vertical receive array using at-sea experimental data, while the probe source is either fixed or moving at about 4 knots. The performances of two different approaches are compared: (1) time reversal alone and (2) time reversal combined with adaptive channel equalization. In the presence of source motion, the motion-induced Doppler shift is coarsely estimated using a decision-feedback phase-locked loop with a training sequence and then the received time series is resampled prior to the demodulation process. The time-varying channel responses due to source motion require an adaptive channel equalizer such that time reversal combined with the equalizer outperforms time reversal alone by up to 13 dB as compared to 5 dB for a fixed source case. The experimental results around 3 kHz with a 1-kHz bandwidth illustrate that even two or three receivers (i.e., 2- or 4-m aperture) can provide resonable performance at 4.2- and 10-km ranges in a 118-m deep water.

1:25

4pUW2. Surface reflection caustics and their impact on communications in the littoral zone. Grant Deane (Code 0238, Scripps Inst. of Oceanogr., UCSD, La Jolla, CA 92093-0238)

Gravity waves with wavelengths longer than a Fresnel zone create cusps and caustics in the surface-reflected acoustic field. The properties of these coherent focal regions have been measured in the surf zone and the 20-m-deep littoral waters off Martha's Vineyard. The time-dependent scattering function shows features consistent with a theoretical analysis of the focal region micro-path structure. The rapid amplitude, phase, and Doppler shifts associated with individual micro-paths present challenges to the reliable operation of phase-coherent underwater communications systems that must implicitly or explicitly track the fluctuations in the impulse response. Data from the experiments will be presented and the implications for underwater communications systems will be discussed. [This work has been undertaken in close collaboration with Dr. James Preisig at the Woods Hole Oceanographic Institution and Dr. Chris Tindle at the Auckland University Physics Department, New Zealand, and supported by the Office of Naval Research.]

1:45

4pUW3. High-resolution shallow water tomography using an array of sources and receivers. Philippe Roux, Bruce D. Cornuelle, Jit Sarkar, Tuncay Akal, W. S. Hodgkiss, W. A. Kuperman, Karim G. Sabra, Hee Chun Song (Scripps Inst. of Oceanogr., UC San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0230), and Mark Stevenson (NATO Undersea Res. Ctr., La Spezia, Italy)

During the FAF05 experiment, two vertical arrays of 29 sources and 32 receivers both covering a large part of the shallow water ocean were located 5 km from each other. A complete acquisition of the waveguide impulse responses between each source and each receiver was performed every 20 s during several hours using broadband pulses at 3.5 kHz. Using time-delay beamforming on both the source and receiver arrays, each ray path was isolated and matched to a ray code based on an *a priori* model of the environment. Following this approach, the impulse responses data matrices were then inverted over time to obtain a movie of the range- and depth-dependent fluctuations of the sound speed between the two arrays. Spatial resolution is discussed as a function of the number of rays used in the inversion. [Work supported by ONR.]

2:05

4pUW4. Amplitude statistics for high-frequency underwater acoustic communications. Wen-Bin Yang and T. C. Yang (Naval Res. Lab., Code 7120, 4555 Overlook Ave., SW, Washington, DC 20375, wyang@wave.nrl.navy.mil)

Amplitude statistics is one major channel characteristic of a communication system, particularly when an energy detector is used as an incoherent receiver. The amplitude statistics is usually modeled as a Rayleigh or a Rician distribution for an rf channel. From our recent study, the amplitude statistics of high-frequency underwater acoustic communications channel is found to be a K-distribution based on M-sequence data collected during the TREX04 experiment. The TREX04 experiment was conducted in shallow water ($\sim$70-m depth) off the coast of New Jersey. The source and receivers were located about 35 m deep. The M-sequence signal was operated at 17 kHz with a 5-kHz bandwidth. The K-distribution is one of non-Rayleigh family and may be approximated by a combination of a long-term fading and a short-term fading statistics. The long-term and short-term fading statistics are close to a lognormal distribution and a Rayleigh distribution, respectively. [Work supported by the Office of Naval Research.]

2:20

4pUW5. Single- and multi-channel underwater acoustic communication channel capacity: A computational study. Thomas J. Hayward and T. C. Yang (Naval Res. Lab., Washington, DC 20375, hayward@wave.nrl.navy.mil)

Acoustic communication channel capacity determines the maximum data rate that can be supported (theoretically) by an acoustic channel for a given source power and source/receiver configuration. In this paper, results reported previously [Hayward & Yang, *High Frequency Ocean Acoustics Conference*, AIP Conference Proceedings (American Institute of Physics, New York, 2004) for single-input, single-output (SISO) systems are reviewed. Then, channel capacities of single-input, multi-output (SIMO) and multi-input, multi-output (MIMO) systems are investigated, first in the limit of a dense receiver array spanning the entire acoustic environment and then for arrays consisting of a small number of sensors. For the dense array, the computed values of the MIMO channel capacities for a modeled shallow-water environment are found to be over 2 orders of magnitude higher than for the SISO channel. Channel capacities for smaller arrays are computed using an exact formula and then compared with approximations based on previously published results. Noise modeling issues and implications for underwater acoustic communication systems are discussed. [Work supported by the Office of Naval Research.]

2:35–2:50 Break

2:50

4pUW6. A low-complexity fading simulator for underwater acoustic communication channels. Y. Rosa Zheng (Dept. of ECE, Univ. of Missouri-Rolla 1870 Miner Circle, Rolla, MO 65401)

Shallow water is an extremely complicated place for communications, and realistic channel models can provide an effective tool for system design. This paper proposes a low-complexity simulator that generates channel impulse responses for broadband frequency-selective acoustic channels. It uses a sum-of-sinusoids (SoS) method to generate Doppler fading impulse responses for wide-sense stationary correlated scattering (WSSCS) Ricean fading channels. The line-of-sight component is obtained by $G_0(t,\tau_0)=A_0\exp[2\pi(1+\epsilon_0)ft\cos\alpha_0+\phi_0\delta(\tau-\tau_0)$ and the scatterers by $G_s(t,\tau)=\Sigma_{n=1}^{N}A_n\exp\{j[2\pi ft\Sigma_{m=1}^{M}(1+\epsilon_m)\cos\alpha_{m,n}+\psi_n\delta(\tau-\tau_n)$, where the random parameters A_n, $\alpha_{m,n}$, ψ_n, and τ_n are, respectively, the attenuation coefficients, impinging angles, initial phases, and path delays. The Doppler scaling is $(1+\epsilon_m)f$, where f is the carrier frequency and ϵ_m depends on the speed of sound propagation in water and the velocities of the transmitter, receiver, and wind. Nonuniform probability density functions are selected for impinging angles $\alpha_{m,n}$ to account for the nonisotropic waveguide effect in shallow-water channels. The proposed simulator is a general stochastic model that does not require site-specific information. It only uses $6N+4$ random variables and requires $O(N)$ multiplications and additions, where N is on the order of several tens. This low-complexity simulator can be easily implemented by real-time digital signal processors.

3:05

4pUW7. Ocean variability effects on underwater communication channel. Aijun Song and Mohsen Badiey (College of Marine Studies, Univ. of Delaware, 114 Robinson Hall, Newark, DE 19711)

In the recent high-frequency acoustics experiments KauaiEx and MakaiEx, it is shown that the variability of the oceanographic parameters has direct impacts on the acoustic signal propagation. For example, the variability of the sea surface condition is highly correlated with the changes of energy in the surface bounces. The coherence of the communication channel is also affected by the sea surface wave. To design an efficient communication system, the relevant ocean variability effects should be investigated. Such investigation will also lead to the prediction of the communication performance at dynamic environment. In the presented work, oceanographic parameters, including sound speed profile and sea surface condition, will be correlated with a series of channel property. Their impacts on the communication receiver, especially maximum-likelihood (ML) receivers, will be also presented. [Work supported by the ONR.]

3:20

4pUW8. The effect of medium variability on acoustic variability: Internal waves and thermohaline intrusions (spice). Joy E. Lyons, David L. Bradley, R. Lee Culver, and Kyle M. Becker (Grad. Program in Acoust., The Penn State Univ., P.O. Box 30, State College, PA 16804, jel246@psu.edu)

In July 2002, an experiment was conducted off the coast of southern California in which the two-dimensional temperature and salinity field were measured directly using a towed conductivity, temperature, and depth chain. The spatial scales are approximately 3500 m long by 80 m deep, with a horizontal resolution of 1–2 m. The resulting sound-speed field has been decomposed into fields of internal waves and density-compensated, thermohaline intrusions. The medium variability in these two fields covers many spatial scales. High-frequency acoustic propagation (1–5 kHz) through the derived fields, as well as through the original field, was computed using the PE-based model RAM and the ray-trace model RAY. Using the metric of scintillation index, the contributions of internal waves and thermohaline intrusions to the acoustic variability are examined, and their relative contributions to the total acoustic variability quantified.

3:35

4pUW9. Decompostion of the time reversal operator applied to detection and focusing in a 12-m-deep channel in Brest Bay. Claire Prada, Julien De Rosny, Dominique Clorennec, Jean-Gabriel Minonzio, Mathias Fink (LOA, ESPCI, Universit Paris 7, CNRS UMR 7587, 10 rue Vauquelin, 75231 Paris, France), Lothar Bernire, Sidonie Hibral, Philippe Billant, and Thomas Folgot (Technople Brest Iroise, Brest, France)

We have developed a 12-kHz, 24-element rigid source receiver array [J. Acoust. Soc. Am. **117**, 2462 (2005)]. During sea trials in spring 2005, the array has been deployed in the Brest Bay, in a 12-m-deep channel. Backscattered data from a variety of submerged objects have been recorded for distances up to 1 km. After the description of the experimental setup and the acoustic propagation conditions, results from detection experiments using the DORT method will be presented. In particular, we will focus on separation of target echo from bottom reverberation as well as separation of two targets. Resolution and sensitivity aspects of the method

will be discussed and compared with results obtained with conventional beamforming. Furthermore, selective focusing by transmission of the eigenvectors of the time reversal operator will be shown. [Work supported by the French DGA/SPN under Contract No. 02 77 154 00 470 75 53.]

3:50

4pUW10. Source localization in deep ocean using matched double-path delay method. Yudong Liu (P.O. Box 67, Bin Hai St. No. 16, Dalian City, Liaoning Province, Post Code 116013, PROC, Liu_yudong@tom.com)

The relationship of matched-field processing (MFP) and multipath delay localization (MDL) is described in this paper. MDL is a kind of simplification of MFP. A method of analyzing multipath structure is presented in this paper. The coincidence between the theory prediction and experiment data shows that the method and program could be used to determine multipath time-delay structure. A realistic method called L2 norm match algorithm for MDL is developed. The depth-range ambiguous figures are pretty good. Comparison between the match method of absolute delay and relative delay shows that the performance of the latter is better. In the case of known source depth, a double-path match plan is given here. It is easy to determine the source distance from the range-delay curves. Many experiments have shown that the error is less than 6% within the range of 30 km and less than 10% within the range of 40 km. The influence caused by mismatch of ocean environment parameters and receiver configuration on double-path ranging is analyzed.

4:05–4:35
Panel Discussion

THURSDAY EVENING, 8 JUNE 2006 ROOM 550AB, 4:00 TO 7:00 P.M.

Session 4eAB

Animal Bioacoustics: Urban Bioacoustics

Andrea M. Simmons, Chair
Brown Univ., Dept. Psychology and Neuroscience, Box 1853, Providence, RI 02912

China Blue is an artist and innovator who works in a variety of media that focuses on sound and how it shapes space. Her current works are based on what she terms ''urban bioacoustics,'' recording activities from day-to-day life to examine sonic flow and energy. The current show includes two sound pieces: ''4 Ball Corner Pocket,'' uses spatial recording and post-production processing to capture and manipulate the acoustic element of a billiard game. The manipulated sound fields are presented through a 7.1 surround sound system to segregate the sounds in space. This creates a highly active sound environment that impresses on the viewer both the physicality and the dynamics of the game in a very primal manner. ''Mikey vs. Fabio,'' is a study of the acoustics of a ping-pong game presented in discrete stereo and mimicking the linearity of the two player game. It also immerses the listeners in both the spatialized dynamics of the ball and the human conversations punctuated by the play. She will also display a number of paintings of her visual interpretation of acoustic flow in different environments.

Session 5aAA

Architectural Acoustics: Current Topics in Measurement and Modeling

William M. Hartmann, Chair
Michigan State Univ., Dept. of Physics and Astronomy, East Lansing, MI 48824

Contributed Papers

8:00

5aAA1. An experimental study of how paint affects the specific airflow resistance and visual appearance of nonwoven scrims on acoustical ceiling tiles. William Frantz and Anthony Wiker (Armstrong World Industries Inc., 2500 Columbia Ave., Lancaster, PA 17604, whfrantz@armstrong.com)

This study explores how paint affects the specific airflow resistance and visual appearance of nonwoven scrims used on acoustical ceiling tile. Nine different scrims are included. Each scrim is finished with two application rates of paint. The specific airflow resistance is measured by the ASTM C522 method. The painted samples are also scanned and the images processed by ImagePro software to characterize visual aspects such as number of features, area of features, and roundness. It is found that specific airflow resistance and visual responses can be correlated to the resistance of the unpainted scrim, the surface tension characteristics of the scrim and paint, and the application rate of the paint. The resulting empirical models can be used to understand design tradeoffs between airflow resistance and visual impact.

8:15

5aAA2. Maximum length sequences versus random telegraph noise in measuring impulse responses and transfer functions. Neil L. Aaronson and William M. Hartmann (Dept. of Phys. and Astron., Michigan State Univ., 4230 BPS Bldg., East Lansing, MI, 48824)

Maximum length sequences (MLS) are commonly used to measure impulse responses and transfer functions of linear systems. The remarkable utility of the MLS comes from the property that its autocorrelation is approximately a delta function, and this approximation becomes more valid for higher orders (lengths) of MLS. However, the autocorrelation of a random telegraph noise (RTN) sequence also approaches a delta function in the long-sequence limit. Advances in computer hardware make it possible to use long sequences of RTN. Consequently one wonders if the MLS still confers an advantage. This study presents a quantitative analysis of the accuracy of MLS and RTN for all lengths up to $2^N-1=262{,}143$. RTN sequences were generated by scrambling the elements of an MLS. A special case wherein the approximate length of the impulse response is prior knowledge allows the measured impulse response to be windowed. Even with this advantage applied to RTN (but not to MLS), MLS measurements are clearly more accurate. The numerical experiments conclude than any MLS sequence longer than the impulse response outperforms a RTN of any practical order by an order of magnitude. [Work supported by the NIDCD grant DC 00181.]

8:30

5aAA3. Simulations of head-related transfer functions in wideband acoustics. Eira T. Seppälä, Ole Kirkeby, Asta Kärkkäinen, Leo Kärkkäinen (Nokia Res. Ctr., Itämerenkatu 11-13, FI-00180 Helsinki, Finland), and Tomi Huttunen (Univ. of Kuopio, Kuopio, Finland)

Head-related transfer functions (HRTFs) have been simulated in three dimensions for a head-and-torso model for the entire audio frequency range, from 20 Hz to 20 kHz. As opposed to data acquired through measurements, the results derived from computer simulations are free from the effects of noise and imperfections in the electro-acoustic chain. In addition, the spatial resolution can easily be made better than in any practical experiment. The simulations have been performed using an ultraweak variational formulation method for solving Helmholtz equation. The numerical method with its parallel computing capability is efficient, enabling numerical calculations of large physical systems, e.g., of size $0.4\times0.5\times0.8$ m, at high frequencies. The method uses plane-wave basis functions instead of polynomial basis as in standard finite-element methods (FEM), resulting in the need of much sparser volume mesh than in FEM. For HRTF calculations, a so-called perfectly matched layer has been utilized and the solutions are derived in the far field. Thus, exterior problems often solved using boundary-element method are dealt with. The specific HRTF simulations have been performed using fully reflecting, sound hard, boundaries on the head and torso, and also with complex impedance boundaries on the torso such as clothing.

8:45

5aAA4. Reverberation time, mean-free-path, and sound absorption in concert halls—Numerical examination by computer simulation. Takayuki Hidaka and Noriko Nishihara (Takenaka R&D Inst., 1-5-1, Otsuka, Inzai, Chiba 270-1395, Japan)

Sabine and Eyring equations are commonly used to estimate the reverberation time (RT) in concert halls. Derivations of the two equations are based on well-defined and different physical assumptions, and they are both slightly different from the realistic situation in halls. In many researches so far, comparatively simple room shape has been utilized to study the question of which equation is preferable. In this study, the sound fields in rooms with more complex or wide-ranging shape are numerically analyzed by applying the computer simulation technique and CAD models for architecture. Focusing on the fundamental relation between the mean free path as a physical measure of the room and the RT, validity of the two equations in actual halls is discussed. Next, the influence of the room shape or the choice of the RT equation on the effective sound absorption is examined.

9:00

5aAA5. A high-frequency broadband energy-intensity boundary element method for specular reflection in three-dimensional reverberant enclosures. Jerry W. Rouse, Linda P. Franzoni, and Donald B. Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708, linda.franzoni@duke.edu)

A new method is presented for predicting the spatial variation of mean-square pressure within three-dimensional enclosures having steady-state, high-frequency broadband sound fields. The enclosure boundaries are replaced by a continuous distribution of uncorrelated broadband directional sources, which provide constituent fields expressed in terms of mean-square pressure and time-averaged intensity variables. Boundary conditions for radiating and absorbing surfaces are recast in energy and intensity variables. Superposition of these source fields in a numerical boundary element formulation leads to the prediction of overall mean-square pressure and time-averaged intensity as a function of position. Both specular and diffuse reflection boundaries can be accommodated. In contrast to the traditional boundary element approach, this method is independent of frequency, and each element has multiple unknowns, namely the strengths of directivity harmonics. For verification, exact analytical solu-

tions for the mean-square pressure distribution in several model problem enclosures were obtained by modal sums and frequency integration and compared to the new method. The comparisons show the method is very accurate, and it is extremely computationally efficient in comparison to classical modal and boundary element approaches. Results also show that diffuse field methods do not provide an accurate simulation for specular reflection cases.

9:15

5aAA6. Energy-based field quantities in reverberation chamber measurements. David Nutter, Timothy Leishman, and JohnPaul Abbott (Brigham Young Univ., N247 ESC, Provo, UT 84602, dave_nutter@hotmail.com)

Traditional reverberation chamber measurements, such as sound power and sound absorption measurements, rely on the evaluation of acoustic pressure at many field positions. Other field quantities could be better suited for these applications if they had greater spatial uniformity and required fewer source-receiver positions to produce accurate measurement results. This paper explores the possibility of using energy-based field quantities for this purpose and provides several analytical and experimental results to show their benefits.

9:30

5aAA7. Analysis method for temporal decay of diffuse broadband sound fields in enclosures. Donald Bliss, Ben Manning, and Linda Franzoni (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708)

An analysis method to predict the behavior of time-dependent broadband diffuse sound fields in enclosures is described. A formulation utilizing time-dependent broadband energy-intensity boundary sources, including propagation time delays, is developed as a numerical boundary element method. An interpolation method is used to re-express the actual delays in terms of a discrete set of integer-multiple delays, thereby facilitating the numerical solution. The method is demonstrated for the prediction of temporal decay in interconnected one-dimensional channels and in two-dimensional enclosures. Temporal behavior is expressed in terms of a higher-dimensional eigenvalue/eigenmode problem, with boundary source strength distributions expressed as eigenmodes. In temporal decay from steady state, solutions exhibit rapid short-time spatial redistribution of energy, followed by long-time decay of a predominant spatial eigenmode. Long-time decay depends on behavior of the most slowly decaying eigenmode and the relative source-panel strengths do not depend on initial conditions. Short-time adjustment and decay depends on initial source characteristics and the relative distribution of absorbing material. An interesting feature of the short-time behavior is that the time-decaying eigenmodes are primarily a manifestation of energy redistribution rather than absorption.

9:45

5aAA8. Absorption-based scaling method to predict steady-state broadband sound fields in enclosures. Donald Bliss (Mech. Eng. and Mater. Sci., Duke Univ., Durham, NC 27708), Jerry Rouse (Sandia Natl. Labs., Albuquerque, NM 87185-0847), and Linda Franzoni (Duke Univ., Durham, NC 27708)

An absorption-based analysis method is developed for steady-state broadband sound fields in enclosures having diffuse or specular reflection boundaries. The wall absorption is expressed as an overall spatially averaged value, with spatial variations around this mean. Interior pressures and intensities are expressed in a power series of the overall absorption, treated as a small parameter, thereby giving a separate problem at each order. The first problem has a uniform mean-square pressure level proportional to the reciprocal of the absorption parameter, as expected. The second problem gives a mean-square pressure and intensity distribution that is independent of the absorption parameter and is primarily responsible for the spatial variation of the reverberant field. This problem depends on the location of sources and the spatial distribution of absorption, but not absorption level. Higher-order problems proceed at powers of the absorption parameter, but are essentially small corrections to the primary spatial variation. The primary spatial variation problem is solved using two different methods. Although formally developed based on treating absorption as a small parameter, the method is demonstrated to work remarkably well in practice up to large absorptions, and to provide insight into the behavior and design of acoustic spaces.

FRIDAY MORNING, 9 JUNE 2006 ROOM 551A, 7:55 A.M. TO 12:35 P.M.

Session 5aABa

Animal Bioacoustics and Psychological and Physiological Acoustics: Is Hearing All Cochlear?

Neil P. McAngus Todd, Chair

Univ. of Manchester, Psychology Dept., Oxford Rd., Manchester, M60 1QD, UK

Chair's Introduction—7:55

Invited Papers

8:00

5aABa1. Is all hearing cochlear?—revisited (again). Neil P. McAngus Todd (Faculty of Life Sci., Univ. of Manchester, M60 1QD, UK, neil.todd@manchester.ac.uk)

Over 70 years ago John Tait raised a question with regards to the certainty that the line of division between the hearing and vestibular portions of the VIIIth cranial nerve had been drawn at the proper place [J. Tait, Ann. Otol. Rhinol. Laryngol. **41**, 681–704 (1932)]. His reason for raising this question was the evidence which existed at that time that the otolith organs of anamniotes (fish and amphibians), most commonly the sacculus, were demonstrably organs of hearing in many species. He went on to speculate about a possible auditory role for the otolith organs in amniotes (reptiles, birds, and mammals), including in humans, although at this time there was no direct evidence for this. Since this time, however, there has developed a significant literature documenting evidence, not only for anamniote otolithic hearing, but also for a conserved acoustic sensitivity for the amniote otoliths. In the case of humans, this

5a FRI. AM

sensitivity has been exploited for the development of new acoustic tests of vestibular function. The time is ripe 70 years on, therefore, for us to revisit John Tait's question and explore the possibility that the sacculus may indeed contribute to human hearing. [Work supported by the Royal Society.]

8:25

5aABa2. Structure/function relationships in the saccule of fishes. Arthur N. Popper, Xiaohong Deng, David Zeddies (Dept. of Biol., Univ. of Maryland, College Park, MD 20742, apopper@umd.edu), and Mardi C. Hastings (Office of Naval Res., Arlington, VA 22203)

The saccule is morphologically the most diverse of the otolithic end organs among fish. Each otolithic end organ contains a sensory epithelium with numerous hair cells. The apical ends of the hair cell ciliary bundles contact a solid dense otolith. Relative motion between the sensory epithelium and otolith results in stimulation of the hair cells. Diversity in the saccule includes the overall shape and size of the end organ, the number of hair cells, the length of the ciliary bundles, and the shape and size of the otolith. Functionally, the saccule subserves both hearing and vestibular senses. There is some correlation between the orientation pattern of the sensory hair cells with whether fish are specialized for hearing or not. Yet, there is no real understanding of the relationship between structural diversity and functional differences. In effect, it is not known whether the structural diversity indicates adaptations for different auditory and/or vestibular functions, or whether the structural diversity reflects different ways for taxonomically and ecologically diverse fish to do similar ear functions. This talk will explore these issues using data from highly diverse deep-sea fish and direct measures of responses of the inner ear to sound.

8:50

5aABa3. Acoustically responsive fibers in the mammalian vestibular nerve. John J. Guinan, Jr. (Eaton Peabody Lab., Massachusetts Eye & Ear Infirmary, Harvard Med. School, 243 Charles St., Boston, MA 02114)

There are afferent fibers in the vestibular nerve of cats and guinea pigs that respond to sound at levels within the normal range of hearing. Most acoustically responsive fibers are in the inferior vestibular nerve and have irregular spontaneous activity, although not as irregular as cochlear afferents. Single-fiber labeling shows that acoustically responsive, irregularly discharging (ARID) fibers originate in the saccule. ARID fibers traced centrally arborized extensively in vestibular nuclei and ventromedial to the cochlear nucleus. ARID fibers had broad, V-shaped tuning curves with best frequencies between 500 and 1000 Hz, thresholds of 90 dB SPL or more, and shapes comparable with tuning-curve tails of cochlear afferents. ARID fibers synchronized to a preferred phase of the tone cycle at levels approximately 10 dB lower than their firing-rate threshold. ARID spike rates increased monotonically with sound level without saturating at 115 dB SPL. Acoustic clicks evoked spikes in ARID fibers with minimum latencies less than 1.0 ms. Contraction of the middle-ear muscles decreased responses to sound, consistent with the sound transmission path being through the middle ear. In summary, there is strong evidence that the mammalian saccule responds to sound and sends acoustic information to the central nervous system. [Work supported by NIDCD RO1DC00235.]

9:15

5aABa4. Assessing saccular (otolith) function in man. James G. Colebatch (Dept. of Neurol., Prince of Wales Hospital, Randwick Sydney, NSW 2031 Australia)

Scattered observations have suggest that the vestibular apparatus may have retained sound sensitivity in man and primates [e.g., Young *et al.*, Acta Otolaryngol. **84**, 352–360 (1977)]. In 1964 Bickford *et al.* [Ann. N.Y. Acad. Sci. **112**, 204–218 (1964)] reported the properties of the ''inion response,'' evoked by clicks and which in subsequent publications they deduced was likely to be arising from the saccule. Colebatch *et al.* [Colebatch, *et al.* J. Neurol. Neurosurg. Psychiatry **57**, 190–197 (1994)] reported a new method which reliably demonstrated a short latency vestibulocollic reflex evoked by clicks and recorded from sternocleidomastoid EMG. Consistent with previous evidence, the saccule was thought to be the most likely receptor to mediate the response, a conclusion for which there has been further experimental support. Bone-conducted acoustic stimulation can also evoke a similar reflex and, although saccular receptors are probably excited, this may not be the sole vestibular receptor activated under these conditions. The vestibular apparatus is relatively more sensitive to bone-conducted sound than to air-conducted sound, when compared to the cochlea. [Work supported by NH&MRC Australia.]

9:40

5aABa5. Central connections of anamniote auditory otolith endorgans. Catherine A. McCormick (Dept. Biol. and Dept. Neurosci., Oberlin College, Oberlin, OH, catherine.mccormick@oberlin.edu)

The ascending auditory pathways of anamniotes (fish and amphibians) are organized similarly to those of amniotes (reptiles, birds, and mammals). A common pattern for the flow of acoustic information—inner ear to first-order nuclei to higher-order auditory nuclei in the medulla, midbrain, thalamus, and cerebrum—probably evolved early in vertebrate history and was retained in extant fish and in terrestrial vertebrates (with taxon-specific modifications) despite variability in inner-ear acoustic receptors. Otolith endorgans subserve hearing in jawed fish; auditory fibers segregate from otolith endorgan vestibular fibers and supply auditory divisions of functionally mixed (auditory-vestibular) first-order nuclei. Amphibian auditory otolith endorgans are primarily seismic detectors complementing new sound-pressure receptors of the amphibian and basilar papillae. Otolithic and papillar auditory endorgans supply common areas within functionally mixed (auditory-vestibular) first-order nuclei. Uniquely in anurans, they also supply a discrete first-order auditory nucleus analogous to the amniote cochlear nuclei. Discrete first-order auditory nuclei are correlated with the evolution of a tympanic ear, an independently evolved condition in anurans and among amniotes. Thus, at the first-order level, the amphibian auditory pathway retains primitive organizational features present in fish auditory pathways. Convergent neuroanatomical specializations in anurans and amniotes may reflect common functional requirements.

10:05–10:20 Break

10:20

5aABa6. Evidence for near-field hearing in crocodilian vocal communication: Intensity of the American alligator (*Alligator mississippiensis*) vocal display. Neil P. McAngus Todd (Faculty of Life Sci., Univ. of Manchester, M60 1QD, UK, neil.todd@manchester.ac.uk)

The aim of this study was to carry out an acoustic analysis of alligator vocal displays recorded simultaneously in air and water in order to estimate the role of the sacculus in crocodilian acoustic communication. Crocodilians are interesting animals to study as they are aquatic, highly vocal, and morphologically have a large sacculus. Alligators, in particular, are known to be one of the most vocal species. Previous studies of alligator hearing in air and water have indicated a maximum sensitivity between 800 and 1000 Hz [Higgs *et al.*, J. Comp. Physiol. **188**, 217–223 (2002)]. Analysis of alligator vocalizations, however, indicates that most of the power in adult vocalizations in water is typically between 30–50 Hz. These results leave the very real possibility that the basilar papilla may not be the primary receptor during crocodilian water-borne communication. At 1-m distance peak SPLs of 140 dB can be obtained in water and, in conjunction with existing threshold data, the effective radius of saccular acoustic sensing of the above source can be estimated to be approximately 30 m, which is easily greater than the typical distance between performer and receiver animals during a display.

10:45

5aABa7. Saccular hearing in anurans. Seth S. Horowitz and Andrea M. Simmons (Dept Psych. & Neurosci., Brown Univ., Providence RI 02912)

Acoustic sensing in anurans is mediated by several inner ear organs, including the amphibian papilla, basillar papilla, and lagena. The contribution of the saccule to hearing in adult ranids tends to be constrained to vibration sensitivity and gravity sensing, although there has been some evidence for low-frequency hearing sensitivity as well. The organs that mediate hearing in larval ranids are less well understood. Several studies have postulated a strong role for saccular-mediated hearing in tadpoles, operating by the sensing of particle motion mediated by direct kinetic effects on otolithic organs (the fenestral pathway). Physiological studies show that, in tadpoles, the medial and lateral vestibular nuclei are sensitive to particle motion stimulation, while the dorsal medullary nucleus is more sensitive to pressure stimulation. Anatomical tract tracing with lipophilic cyanocarbanine dyes placed in the saccular branch of the eighth nerve show projections to both auditory and vestibular medullary nuclei, suggesting that input from the saccule may play a role in both particle and pressure sensation.

11:10

5aABa8. The sense of hearing mediated by the saccule in goldfish. Richard R. Fay (Dept. Psychol. and Parmly Hearing Inst., Loyola Univ. Chicago, 6525 N. Sheridan Rd., Chicago, IL 60626)

The saccule mediates the sense of hearing for many fish species, goldfish (*Carassius auratus*) in particular. Motions of the swim bladder in response to pressure fluctuations impinge on the saccule of each ear, where auditory-nerve fibers project to medullar, midbrain, and forebrain nuclei in a circuit analogous to the ascending auditory system of tetrapods. It has been repeatedly shown over the past 40 years that the sense of hearing mediated by the saccule has many fundamental features in common with the sense of hearing of tetrapods. Detection and discrimination thresholds for goldfish are in the vertebrate range quantitatively, and are of the same sorts of simple and complex discriminations that mammals and other tetrapods can perform. Furthermore, goldfish behave as if the perceptual dimensions of spectral pitch, complex musical pitch, timbre, and roughness exist for them. Although goldfish do not vocalize, many fish species that hear using saccular input do vocalize (e.g., mormyrids), and have special adaptations for responding to vocalizations, like many tetrapods. Finally, goldfish have been shown capable of simultaneous stream segregation based on the same principles known for humans and other species. Thus, otolith organs can participate in a sense of hearing that is typical for vertebrates.

Contributed Papers

11:35

5aABa9. Are acoustically induced flows relevant in fish hearing? Charlotte W. Kotas, Peter H. Rogers, and Minami Yoda (Mech. Eng. Dept., Georgia Inst. of Technol., Atlanta, GA 30332-0405, charlotte.kotas@gatech.edu)

The fluid-filled fish ear contains an irregularly shaped, dense, bony otolith that influences the acoustically induced motion of the adjacent hair cells and thus what fish "hear." Because incident sound causes the otolith to oscillate with respect to its surroundings, acoustically induced fluid flows are generated near the otolith. A series of experiments was performed using oscillating spheroids as model otoliths to study the induced steady streaming and periodic flows. The studies focused on how the flows change with spheroid (or otolith) geometry and orientation, as well as oscillation (or sound) frequency ω and amplitude s. Experiments were conducted for Reynolds numbers $\mathrm{Re} \equiv \omega L^2/\nu < 10^2$ and amplitude ratios $\epsilon \equiv s/L = 0.05-0.2$, where L is the product of the spheroid aspect ratio and equivalent radius. Results will be presented for oblate and prolate spheroids oscillating at angles up to 45° relative to the spheroid's axis of symmetry and for flows induced by spheroids driven simultaneously at different frequencies. Understanding these flows may give insight into how fish localize sound sources and lead to new acoustic sensor designs. [Work supported by ONR.]

11:50

5aABa10. Saccular hearing; turtle model for a human prosthesis. Martin L. Lenhardt (Dept. of Biomed. Eng., Otolaryngol., and Emergency Medicine, Virginia Commonwealth Univ., Box 980168 MCV, Richmond, VA 23298-0168)

The saccule is a hearing organ is some vertebrates thought to be responsive to substrate vibration (bone conduction) or low-frequency aerial sound. There was likely some overlap in these functions in the course of evolution after the sensory area to become the cochlea migrated from the saccule. That overlap is preserved in extant turtles by columella (stapes) saccular coupling via fibroelastic strands; thus both organs can respond to air conduction and bone conduction stimulation. Evoked potential data, however, reflect differential AC/BC drive to the inner ear. The columellas inertia provides the force to displace the saccular wall with the tympanum

providing damping. Intense AC stimulation likely stimulates the human saccule. A proprietary stapedial saccular strut is described that serves as a surgically implanted coupling device for humans, allowing more efficient use of AC saccular hearing in clinical deafness. The human saccular resonance is about 350 Hz, which should allow for sufficient speech coding for intelligibility assuming connectivity to the auditory neuraxis. BC stimulation through audible ultrasound also likely activates the saccule in individuals with profound deafness. A stand-alone stapedio-saccular strut or one used in combination with an ultrasonic hearing aid offers the potential of communication through saccular hearing.

12:05–12:35
Panel Discussion

FRIDAY MORNING, 9 JUNE 2006 ROOM 550AB, 8:00 TO 10:00 A.M.

Session 5aABb

Animal Bioacoustics: Classification and Parameter Estimation

John R. Buck, Chair
Univ. of Massachusetts, Dartmouth, SMAST, 706 S. Rodney French Blvd., New Bedford, MA 02744-1211

Contributed Papers

8:00

5aABb1. Online acoustic parameter extraction and annotation for a marine sound collection. Shelagh A. Smith and Jack W. Bradbury (Macaulay Library, Cornell Lab of Ornithology, Ithaca, NY 14850)

After 3 years of archival and with the collaboration of more than 20 researchers and institutions, the online marine sound collection at the Macaulay Library, Cornell Lab of Ornithology, is now in a proof of concept phase. This presentation introduces the new annotation and acoustic parameter extraction and retrieval tools. It is further explained how these tools are used for fine-grain searches and recovery of specific sounds within long recordings while online. Also included is a description of the online open access model and delivery mechanism. [Work supported by ONR and NSDL.]

8:15

5aABb2. Dynamic time warping for automatic classification of killer whale vocalizations. Judith C. Brown (Phys. Dept., Wellesley College, Wellesley, MA 02481 and Media Lab., MIT, Cambridge, MA 02139) and Patrick J. O. Miller (Univ. of St. Andrews, St. Andrews, Fife KY16 9QQ UK)

A large number of sounds from the captive killer whale population at Marineland of Antibes, France have been classified perceptually into call types. [A. Hodgins-Davis, thesis, Wellesley College (2004)]. The repetition rate of the pulsed component of five or more examples of each call type was calculated to give a set of melodic contours. These were compared pairwise using dynamic time warping to give a dissimilarity or distance matrix. The distances were then transformed into a component space using multidimensional scaling, and the resulting points clustered with a *k*-means algorithm. In grouping 57 sounds into 9 call types, there was a single discrepancy between the perceptual and automated methods. Preliminary measurements on a second set of 74 sounds from northern resident killer whales have been made. Both the high-frequency and low-frequency contours have been measured and classified perceptually into 7 groups. Errors for the automatic classification of this group are over 10%. [Sound recording supported by grants from WHOI's Ocean Life Institute and the Royal Society to Patrick J. O. Miller.]

8:30

5aABb3. Probability distributions for locations of calling animals, receivers, sound speeds, winds, and data from travel time differences. John Spiesberger (Dept. of Earth and Environ. Sci., Univ. of Pennsylvania, 240 S. 33rd St., Philadelphia, PA 19104-6316, johnsr@sas.upenn.edu)

A new, nonlinear sequential Monte Carlo technique is used to estimate *posterior* probability distributions for the location of a calling animal, the locations of acoustic receivers, sound speeds, winds, and the differences in sonic travel time between pairs of receivers from measurements of those differences, while adopting realistic *prior* distributions of the variables. Other algorithms in the literature appear to be too inefficient to yield distributions for this large number of variables (up to 41) without recourse to a linear approximation. The new technique overcomes the computational inefficiency of other algorithms because it does not sequentially propagate the joint probability distribution of the variables between adjacent data. Instead, the lower and upper bounds of the distributions are propagated. The technique is applied to commonly encountered problems that were previously intractable, such as estimating how accurately sound speed and poorly known initial locations of receivers can be estimated from the differences in sonic travel time from calling animals, while explicitly modeling distributions of *all* the variables in the problem. In both cases, the new technique yields one or two orders of magnitude improvement compared with initial uncertainties. The technique is suitable for accurately estimating receiver locations from animal calls.

8:45

5aABb4. Automatic call recognition of cricket songs using image processing. T. Scott Brandes (Tropical Ecology Assessment and Monitoring (TEAM) Initiative, Conservation Intl., 1919 M St., NW, Washington, DC 20036), Piotr Naskrecki (Harvard Univ., Cambridge, MA 02138), and Harold K. Figueroa (Cornell Lab. of Ornithology, Ithaca, NY 14850)

An automatic call recognition (ACR) process is described that uses image processing techniques on spectrogram images to detect constant-frequency cricket calls recorded amidst a background of evening sounds found in a lowland Costa Rican rainforest. This process involves using image blur filters along with binary filters to isolate calling events. The binary filters are used to isolate potential calls from background noise, and the blur filters are used to unite discrete call fragments as a single continuous call. Features of these events, notably the events central frequency, duration, and bandwidth, along with the type of blur filter applied, are used with a Bayesian classifier to make identifications of the different calls. Of the 22 distinct sonotypes (calls presumed to be species specific)

recorded in the study site, 17 of them are recorded in high enough numbers to both train and test the classifier. The classifier approaches 100% true-positive accuracy for these 17 sonotypes. The high true-positive accuracy of this process enables its use for monitoring singing crickets in tropical forests.

9:00

5aABb5. Feature normalization for robust individual identification of the ortolan bunting (*emberiza hortulana L*). Kuntoro Adi and Michael T. Johnson (Dept. of Elec. and Comput. Eng., Marquette Univ., Milwaukee, WI 53201, kuntoro.adi@marquette.edu)

This paper presents the use of cepstral moment normalization for the purpose of improving the accuracy of individual identification in the ortolan bunting (*emberiza hortulana L*). The underlying classification algorithm is based on hidden markov models (HMMs) using Greenwood function cepstral coefficients (GFCCs), and the proposed feature normalization is implemented via a combination of cepstral mean subtraction and cepstral variance normalization, which are equivalent to a statistical z-score normalization across each example utterance. Results indicate a significant reduction in error on the individual identification task, improving accuracy from 89% to 96% in initial experiments, using a single call type over a small population of 6 individuals.

9:15

5aABb6. Characteristics of nonlinear phenomena in the tonal vocalizations of a North American canid, *Canis rufus*. Jennifer N. Schneider (Dept. of Psych., Univ. of Buffalo, Park Hall Rm. 206, Buffalo, NY, 14260), Rita E. Anderson, and Edward H. Miller (Memorial Univ. of NF, St. Johns, NF, A1B 3X9 Canada)

This study examines the structure and frequency of occurrence of nonlinear phenomena found in tonal vocalizations produced by red wolves. Spectrograms were obtained from audio tracks of digital video recordings of captive wolves from a breeding facility in Graham, WA. Tonal vocalizations were determined to be composed of 1–30 sound units, arranged in 1–5 phrases. Linear units included squeaks (2600–9600 Hz) and wuhs (100–1600 Hz); nonlinear units accounted for 22% of sounds and included between-type frequency jumps, harmonic and pure-tone biphonations, squeaks with sidebands, and squeak jumps. Five tonal vocalization types were identified based on unit composition: squeaks (48.4%), wuhs (19.3%), and three mixed vocalizations: banded squeaks (13.2%), complex squeaks (6.4%), and squeak-wuhs (12.2%). Unit order within a squeak-wuh vocalization was not random; transitions between units following a structural gradient most likely begin with higher frequency units and end with mixed or lower frequency units. The production of nonlinear sounds varied within and between individuals. The linear and nonlinear structure of red wolf tonal vocalizations are similar to that which has been reported in dholes and African wild dogs, and which have been indicated in reviews of published sonograms of gray wolf vocalizations.

9:30

5aABb7. Equine acoustics: Further analysis of a whinny—The expressive component. David G. Browning (Dept. of Phys., Univ. of Rhode Island, Kingston, RI 02881, decibeldb@aol.com) and Peter M. Scheifele (Univ. of Connecticut, Storrs, CT 06269)

A horse's whinny appears to be a very interesting vocalization. Initial acoustic analysis suggested two principal components, one ''emotional'' and the other ''expressive.'' The first seems to follow Morton's criterion: increasing frequency with emotional state. The second component, since many suggest that a whinny is not a threatening gesture, has the potential for a number of specific expressions for such situations as ''calling'' or ''greeting.'' This paper describes the present indications as the data base slowly increases.

9:45

5aABb8. Discovery of Sound in the Sea website: An educational resource. Kathleen J. Vigness Raposa (Marine Acoust., Inc., 809 Aquidneck Ave., Middletown, RI 02842, kvigness@aol.com), Gail Scowcroft, Jill Johnen, Chris Knowlton (Univ. of Rhode Island, Narragansett, RI 02882), and Peter F. Worcester (Univ. of California—San Diego, La Jolla, CA 92093-0225)

The scientific community and the public have become increasingly aware of, and concerned about, underwater sound. There is increasing interest in learning about sources and uses of sound, and potential effects of sound on the marine environment. The Discovery of Sound in the Sea website (http://www.dosits.org) provides scientific information for the general public and K–16 educators and students. It also includes advanced level content appropriate for high school physics or undergraduate classes. The website has three major sections, (1) Science of Sound in the Sea, (2) People and Sound in the Sea, and (3) Animals and Sound in the Sea, introducing the physical science of underwater sound and how people and animals use sound to accomplish various tasks. The Animals and Sound section also summarizes the current state of knowledge of the effects of underwater sound on marine mammals and fishes. Three galleries contain a collection of underwater sounds (Audio Gallery), video interviews with researchers (Scientist Gallery), and descriptions of acoustic equipment (Technology Gallery). Additions to the website are ongoing. Scientists with recordings and/or equipment that have not been included are encouraged to contact the authors. A limited number of free CDs of the website will be available.

FRIDAY MORNING, 9 JUNE 2006 ROOM 550AB, 10:15 A.M. TO 12:00 NOON

Session 5aABc

Animal Bioacoustics: Acoustic Sensing of Animals

Jennifer L. Miksis-Olds, Chair

Univ. of Rhode Island, Graduate School of Ocean Biology, Narragansett, RI 02882

10:15

5aABc1. Tracking fish motion and behavior with acoustic flow. Srinivasan Jagannathan, Deanelle Symonds, Sunwoong Lee, Nicholas C. Makris (MIT, Cambridge, MA 02139), Ninos Donabed, and Purnima Ratilal (Northeastern Univ., Boston, MA 02115)

Fish populations in the New Jersey continental shelf have been continuously monitored over thousands of square kilometers using acoustic waveguide remote sensing techniques [Makris *et al.*, ''Fish population and behavior revealed by instantaneous continental-shelf-scale imaging,'' Science **311**, 660–663 (2006)]. Here, we explore the approaches for tracking both large shoals and smaller isolated fish groups over space and time from images of fish population density. One method uses density thresholds to segment coherent population centers and track their centroids. These data are then used to compute differential velocities between population centers and to determine the rate at which information may be transmitted across a population center. We also examine the possibility of applying optical flow concepts from machine vision [Horn, B. *Robot Vi-*

5a FRI. AM

sion, MIT Press, 1986], as a global approach for estimating fish velocity at every pixel in the fish density image. The analogous acoustic flow is computed based on changes in acoustic intensity over space and time. Acoustic flow automatically tracks fish motion within a school or shoal and migration across fish bridges connecting shoals segments. We also investigate the possibility of using acoustic flow for navigating the ocean environment and correcting for errors in charting acoustic returns, for instance caused by array-heading sensor.

10:30

5aABc2. The roles of density, geometry, and short-range positional order in low-frequency acoustic scattering from fish schools. Thomas R. Hahn and Tanos Elfouhaily (RSMAS, AMP, Univ. of Miami, 4600 Rickenbacker Cswy., Miami, FL 33149)

Fish can be efficient scatterers of low-frequency underwater sound. This results from the large response of their swim bladders to incoming acoustic waves at frequencies close to their natural resonance frequencies. For large and densely packed fish schools the scattering returns deviate significantly from what could be expected by adding the individual fish returns. These deviations are caused by multiple scattering between the individual fish in the school and concern the target strength as well as the spectral and directional characteristics of the scattering cross sections. Acoustically, a fish school can be viewed as a single object with acoustic bulk parameters determined from the properties of the individual fish and their positional arrangement in the school. This effective medium approach incorporates the acoustic coupling of all fish due to multiple scattering and works for a wide range of fish sizes, -numbers, and -densities. In this study, the effective medium approach as well as numerical computations from first physical principles are employed to investigate the coherent and diffusive low-frequency scattering cross sections of large and dense fish schools as a function of fish density, school geometry, and positional order among the individual fish. Inversions for these parameters are explored.

10:45

5aABc3. Assessing fish populations and enhancing biological detection in continental shelf environments with range-dependent scattering and reverberation models. Mark Andrews, Ameya Galinde, Ninos Donabed, Purnima Ratilal (Northeastern Univ., 360 Huntington Ave, Boston, MA 02115), Tianrun Chen, Deanelle Symonds, Sunwoong Lee, and Nick Makris (MIT, Cambridge, MA 02139)

Fish populations in continental shelf environments can be continuously imaged over thousands of square kilometers using acoustic waveguide remote sensing techniques [Makris *et al.*, Science, Feb. (2006)]. A calibrated range-dependent scattering and reverberation model [Ratilal *et al.*, J. Acoust. Soc. Am. **114**, 2302 (2003)] based on the parabolic equation has been applied to assess population densities of fish by inverting long-range acoustic data collected on the New Jersey continental shelf. This model is now applied to predict the types of fish species and zooplankton that are detectable in a general range-dependent continental shelf environment, including the resolution and accuracy that can be expected in estimating fish population densities and for differentiating fish species. We consider different geometries of the source and receiving array to enhance biological detection and reduce background reverberation in highly range-dependent environments. Using multiple source frequencies, the possibility of distinguishing fish species based on their differing scattering characteristics and resonance frequencies will be examined.

11:00

5aABc4. Integrated marine mammal monitoring and protection system: Gray whale target strength measurements. Irena Lucifredi, Amy Vandiver, and Peter Stein (Scientific Solutions Inc., 99 Perimeter Rd., Nashua, NH 03063)

With a broader objective of establishing a global solution for systematic assessment and protection of the marine environment, one of the central objectives of this work is to distinguish whale backscattered responses from the ones generated by the environmental clutter in a waveguide. Furthermore, it aims to identify and analyze the target signature features that are necessary for enhanced active sonar detection and classification of marine mammals. Over the years there have been very few documented attempts to capture and analyze the backscattering response of whales using an active sonar system. Nevertheless, whales, mostly owing to their size, their motion, and the aspect dependence of their backscattered field, possess desirable properties that help distinguish their scattered response from clutter and other environment related false alarms. As an initial step, data collected during the Marine Mammal Active Sonar Test (MAST04) is presented, and long-range gray whale target strength measurements are obtained. Results are compared to the previously published whale target strengths. Additionally, an investigation is conducted in an effort to provide whale feature identification points suitable for automated detection and classification, as a means of relating gray whale active acoustic signatures to their inherent characteristics and motion. [Work supported by ONR.]

11:15

5aABc5. Initial application of spectrogram-based cross correlation to the localization of Orcas. Martin C. Renken (Naval Undersea Warfare Ctr. Div. Keyport, 610 Dowell St., Keyport, WA 98345-7610, renkenmc@kpt.nuwc.navy.mil)

Localization and tracking of sperm whales (*Physeter macrocephalus*) and pilot whales (*Globicephala macrorhynchus*) has been demonstrated using the widely spaced bottom-mounted hydrophone arrays located at the Atlantic Undersea Testing and Evaluation Center (AUTEC), Andros Island, Bahamas [Moretti *et al.*, US Navy J. Und. Acoustics **52**, 651–668 (2002)]. The marine mammal monitoring utilizes a cross correlation of the spectrogram calculated from acoustic calls received at each hydrophone. The spectrogram information is combined with time of arrival pattern estimates to identify specific animals in order to estimate the animal's position through triangulation. This paper explores the initial application of these techniques to the wide variety of acoustic vocalizations recorded from transient killer whales (*Orcinus Orca*) in Dabob Bay of Washington state's inland waters in February of 2005. The recordings were made from a bottom-mounted linear array of hydrophones during a period of low acoustic interference. While this recording was obtained by Naval Undersea Warfare Center Keyport, active operations are not conducted in the presence of whales according to standard operating policy. Monitoring in the Dabob Bay Naval Operating Area is carried out both visually and acoustically by range operators trained in marine mammal identification by NOAA.

11:30

5aABc6. Modeling of acoustically enhanced bubble growth in supersaturated biological media. Yurii A. Ilinskii, Mark F. Hamilton, and Preston S. Wilson (Appl. Res. Labs., Univ. of Texas at Austin, P.O. Box 8029, Austin, TX 78713-8029)

Recent measurements and predictions of dissolved nitrogen concentration in tissues of dolphins and whales following repeated dives have indicated that the gas concentration may reach 300% of saturation [Houser *et al.*, "Can diving-induced tissue nitrogen supersaturation increase the chance of acoustically driven bubble growth in marine mammals?," J. Theor. Biol. **213**, 183–195 (2001)]. This has generated renewed interest in whether underwater sound may increase the likelihood of decompression sickness in these cetaceans. One potential mechanism is bubble growth due to rectified diffusion. Eller and Flynn ["Rectified diffusion during nonlinear pulsations of cavitation bubbles," J. Acoust. Soc. Am. **37**, 493–503 (1965)] modeled rectified diffusion by solving a Rayleigh-Plesset equation for the acoustically driven bubble dynamics, coupled to a quasi-analytical diffusion model that accounts for increase of the equilibrium bubble radius. Crum and Mao ["Acoustically enhanced bubble growth at low frequencies and its implications for human diver and marine mammal," J. Acoust. Soc. Am. **99**, 2898–2907 (1996)] used similar models to consider dissolved gas saturations up to 223% and demonstrated significantly enhanced bubble growth at sound-pressure levels above 210 dB (re: 1 μPa). We are revisiting this problem by extending these models and considering wider ranges of parameters. A principal difference is that the gas diffusion is modeled by solving the full partial differential equation

numerically. Not only does this approach circumvent analytical approximations, it permits taking into account variations in gas concentration associated with dive-profile-dependent hydrostatic pressure variations. [Work supported by ONR.]

11:45

5aABc7. Physics-based volume clutter from GeoClutter biological distributions. Adam Frankel (Marine Acoustics Inc., 706 Giddings Ave., Ste. 1C, Annapolis, MD 21401), Richard H. Love (12315 Fairway Meadows Dr., Fort Worth, TX 76179), Charles Monjo, Bruce Newhall, Juan I. Arvelo, Jr. (Johns Hopkins Univ., Laurel, MD 20723-6099), and William T. Ellison (Marine Acoustics Inc., Litchfield, CT 06759)

Empirical fish species composition and distribution data obtained during a GeoClutter survey were input into scattering models for marine mammals [Love, J. Acoust. Soc. Am. **49**, 816–823 (1971)] and for schools of fish [Feuillade *et al.*, J. Acoust. Soc. Am. 99, 196–208 (1996)]. These scattering models drew additional inputs from a database of biological parameters to generate volume clutter simulations. Model inputs included distribution of fish size, school population size, depth, orientation, and location. The volume clutter model treated each fish school and individual marine mammal as discrete acoustic targets. Motion of the dominant marine organisms (in addition to source and receiver motion) was taken into account to observe clutter map variability during a series of pings. The volume clutter distribution from this physics-based approach was compared against bottom clutter and against measured clutter-rich reverberation from the GeoClutter experiment. [The Office of Naval Research (ONR) is sponsoring this effort.]

FRIDAY MORNING, 9 JUNE 2006 ROOM 553AB, 9:00 TO 11:30 A.M.

Session 5aBB

Biomedical Ultrasound/Bioresponse to Vibration: Targeted Contrast Agents

Tyrone M. Porter, Chair

Univ. of Cincinnati, Biomedical Engineering, Medical Sciences, 231 Albert Sabin Way, Cincinnati, OH 45267-0586

Invited Papers

9:00

5aBB1. Treatment of ischemic stroke with nanobubbles and ultrasound. Evan C. Unger (ImaRx Therapeutics, Inc., 1635 E. 18th St., Tucson, AZ 85719)

The use of perfluorocarbon-filled microbubbles for ultrasound contrast imaging is well known. Now, the next generation of perfluorocarbon-filled bubbles has reached nanoscale and will be evaluated in clinical trials for the dissolution of blood clots in ischemic stroke and other vascular occlusions. MRX-815 nanobubbles are submicron, lipid coated, perfluorocarbon nanobubbles. Due to their small size they are able to penetrate a thrombus. Since they share a lineage with ultrasound contrast agents, their progress can be monitored on diagnostic ultrasound and once they have reached the site of thrombosis ultrasound can be applied externally to activate the bubbles. This activation consists of the bubbles oscillating in the field of ultrasound and cavitating. This mechanical energy works to dissolve the clot and restore blood flow to the ischemic tissues beyond. This process, SonoLysis® therapy, can be used in tandem with lytic agents or can be used alone in cases where lytic agents are contraindicated. A multicenter, Phase II clinical trial is currently evaluating SonoLysisregistered in the treatment of acute ischemic stroke. Preliminary findings have been encouraging and additional studies are planned.

9:30

5aBB2. Echogenic targeted liposomes for transfection and drug delivery. David D. McPherson (Northwestern Univ., 251 East Huron, Galter 8-230, Chicago, IL 60611, d-mcpherson@northwestern.edu)

We have developed targeted, echogenic immunoliposomes. These agents are lipid based, can be targeted to cells, and have encapsulated air bubbles that render them echogenic, and the air bubbles can also be used as cavitation agents. They have the ability to encapsulate drugs and genes. We have developed stable formulations and optimized attachment of antibodies and peptides for targeted molecular imaging using ultrasound. We are determining the pharmacokinetics of drug and gene entrapment and the ultrasonic destruction threshold for therapeutic delivery. With the enclosed cavitation agent, therapeutic ultrasound (sonoporation) potentiates cellular drug and gene delivery. This presentation will focus on optimizing the liposomes for stability, targeting, drug entrapment and release, destruction threshold, and the additive sonoporation effect of the encapsulated cavitation nucleii. Tissue plasminogen activator and papaverine are prototype drugs and beta-galactosidase is the prototype gene for encapsulation. The objective of this work is to use modifications of our echogenic immunoliposomes to better entrap and release drugs and genes, while utilizing the enclosed cavitation nucleii, which allows ultrasound to not only release the drug/gene, but also improve its delivery.

10:00–10:15 Break

10:15

5aBB3. Gas-filled microbubbles—Contrast agents for targeted molecular imaging. Alexander L. Klibanov, John Hossack, Yun K. Cho, William C. Yang, Joshua J. Rychak, and Klaus F. Ley (Univ. of Virginia, Charlottesville, VA 22908)

Microbubble contrast agents suitable for targeted molecular imaging have been described. Commonly used agents are gas-filled micron-size bubbles. Targeting ligands (antibodies, peptides, or proteoglycans) are coupled to the contrast particles and ensure firm attachment of microbubbles to intravascular regions of interest. Targeting is directed to selectins or integrins upregulated on vascular endothelium in the areas of pathology (inflammation, ischemia-reperfusion injury, or angiogenesis). Real-time imaging has been performed in animal models with sub-millimeter spatial resolution. Detection sensitivity for targeted microbubble contrast is superb: single bubbles (pg mass) can be visualized by ultrasound. Successful retention of microbubbles on the target is dependent on the target receptor surface density as well as targeting ligand surface density on the bubbles. To improve targeting efficacy, microbubbles can be outfitted with fast-binding ligands (such as P-selectin glycoprotein ligand-1 analogs) in addition to slow but firmly binding ligands (such as antibodies). Ultrasound irradiation also improves efficacy of targeting, especially in high shear conditions. In addition to imaging, microbubbles can be applied for drug and gene delivery purposes: interaction of ultrasound and bubbles can be used for in situ targeted deposition, activation and release of pharmaceutical agents [Disclosures: ALK, JJR, and KFL: Targeson stock ownership. ALK: Philips Research grant.]

Contributed Papers

10:45

5aBB4. Nonlinear oscillation of microbubble in microtubes for ultrasound imaging and drug delivery. Shengping Qin and Katherine Ferrara (Dept. of Biomed. Eng., Univ. of California, 451 East Health Sci. Dr., Davis, CA 95616)

Ultrasound-assisted microbubble drug delivery is under intensive investigation due to the potential for enhanced local delivery without systemic adverse effects. The efforts in modeling bubble oscillation have been largely focused on using various modified forms of the Rayleigh-Plesset bubble dynamics equation, of which the cornerstone assumption is that a bubble oscillates in an unbounded liquid field and remains spherical until it collapses. Here, nonlinear oscillation of a bubble confined inside a microtube with diameters comparable with capillaries is studied numerically and experimentally. In numerical simulation, the liquid is treated as incompressible viscid fluid and contrast agent is treated as gas bubble obeying the polytropic law. Microbubble oscillation is shown to decrease in a small vessel as compared with oscillation in an infinite fluid or a larger vessel diameter. For the same bubble size, the natural oscillation frequency of a bubble deceases with the decrease of tube size. The natural frequency of a bubble with an equilibrium diameter of 2.5 μm decreases from 3.56 MHz, in unbounded liquid field, to 1.16 MHz in an 8-μm tube. Decreasing the ultrasound center frequency can increase the amplitude of bubble oscillation and thereby produce a higher pressure on the tube.

11:00

5aBB5. *In vitro* and *in vivo* high-frequency ultrasound response of microbubbles. Orlando Aristizábal, Daniel H. Turnbull (Skirball Inst. of Biomolecular Medicine, New York Univ. School of Medicine, 540 First Ave., New York, NY 10016, oarist@saturn.med.nyu.edu), and Jeffrey A. Ketterling (Riverside Res. Inst., New York, NY 10038)

Recently, there has been a growing interest in utilizing contrast agents with high-frequency ultrasound (HFU $>$ 20 MHz). As a first step to using contrast agents *in vivo* with HFU, the acoustic response of several agents was examined using a 40-MHz ultrasound biomicroscope (UBM, Vevo 770, VisualSonics). Optison, Definity, and near-micron and submicron agents from Point Biomedical were investigated. A small volume of the agents was injected into a petri dish of saline solution, and radio-frequency (rf) backscatter data from the UBM were digitized with a digital oscilloscope. Agents were also injected into the cardiovascular system of mouse embryos and rf data were acquired *in vivo*. The echo signals from individual bubbles were windowed in the rf data and the spectra were calculated. The resulting spectra revealed an increasing harmonic content for each of the agents as the acoustic drive amplitude and number of excitation wavelengths increased. Harmonic generation appeared to be due mainly to nonlinear propagation in the water bath and not from a nonlinear response of the contrast agent because reflected spectra without contrast agent showed similar harmonics. Using a modified parameter imaging method, fundamental (40 MHz) and first-harmonic (80 MHz) images were generated.

11:15

5aBB6. Molecular imaging and therapy with perfluorocarbon-based site targeted contrast nanoparticles. Kirk D. Wallace, Michael S. Hughes, Jon N. Marsh, Kathryn C. Crowder, Huiying Zhang, Samuel A. Wickline, and Gregory M. Lanza (Dept. of Medicine, Washington Univ. St. Louis, MO 63130, kirk.wallace@wustl.edu)

Advances in molecular biology and cellular biochemistry are providing new opportunities for diagnostic medical imaging to ''see'' beyond the anatomical manifestations of disease to the earliest biochemical signatures of disease. The emerging field of ''molecular imaging'' encompasses the noninvasive *in vivo* diagnosis of complex pathological processes by detection of unique molecular signatures. Localization of specific biochemical epitopes with targeted contrast agents affords the unique opportunity for both targeted delivery and deposition of therapeutics. Liquid perfluorooctylbromide nanoparticles (approximately 250 nm mean diameter) elicit little acoustic reflectivity in suspension and strong acoustic contrast when bound and concentrated on a surface (e.g., fibrin deposits in a thrombus). This nanoparticle platform may be functionalized with homing ligands, like anti-$\alpha v \beta 3$ integrins, and therapeutic agents. Acoustic imaging of densely distributed biomarkers (e.g., fibrin epitopes) is readily accommodated with fundamental imaging, but for sparse biomarkers (e.g., integrins) we have developed and implemented complementary, nonlinear imaging techniques based upon information-theoretic receivers. Moreover, perfluorocarbon nanoparticles also offer targeted drug delivery, based on contact facilitated lipid exchange, which can be further augmented by ultrasonic insonification.

Session 5aMU

Musical Acoustics, Engineering Acoustics, and Signal Processing in Acoustics: Human-Computer Interfaces

Jonas Braasch, Cochair
Rensselaer Polytechnic Inst., 110 8th Street, Troy, NY 12180

William L. Martens, Cochair
McGill Univ., Faculty of Music, 555 Sherbrooke, St., West, Montreal, PQ, H3A 1E3, Canada

Invited Papers

8:00

5aMU1. Musical coffee mugs, singing machines, and laptop orchestras: New paradigms for musical expression. Perry Cook (Comput. Sci., Princeton Univ., 35 Olden St., Princeton, NJ 08544)

Advances in algorithms, hardware, and sensors now allow us to build new types of expressive musical devices based around small computer systems. These new ''instruments'' can leverage and extend the expertise of virtuoso performers, expand the palette of sounds available to composers, and encourage new ideas and participation from the young or the musically untrained. New musical instuments can take the form of augmentations of traditional musical instruments, whimsical objects made by augmenting everyday nonmusical objects, or entirely new paradigms for making and controlling sound. Wired and wireless networking can provide opportunities, yet pose significant challenges, for new forms of ''orchestral'' expression. This talk will look at a variety of musical technologies, devices, systems, compositions, and performances we have created at Princeton in the last 10 years or so. Lots of audio/video examples, as well as actual live demonstratons, will be presented.

8:30

5aMU2. Mapping strategies for sound synthesis, digital audio effects, and sonification of performer gestures. Vincent Verfaille and Marcelo M. Wanderley (SPCL & IDMIL, Music Technol. Area, McGill Univ., 555 Sherbrooke St. West, Montreal, Canada, H3A 1E3)

Mapping strategies are an essential step when designing realtime musical performance systems, as well as offline digital sound processing. These strategies define how we relate input device parameters to sound synthesis or audio effect parameters. This implies the ability to combine input parameters among themselves (parameter combination) and valid control signals in terms of range, variation type, etc. (signal conditioning). Recent works highlighted the interest of multi-layer mapping strategies in the context of digital musical instruments, which can also be applied in the context digital audio effects. In this presentation, three strategies will be discussed in order to illustrate the role of mapping strategies in various contexts. The first example concerns an additive synthesizer called *Ssynth*, a further development of *Escher*, a prototyping system aiming at studying the effect of mapping strategy in instrument design. The second example is a general mapping strategy for digital audio effects, allowing for both adaptive and gestural control. The final example concerns sonification of gestures, used to provide cues about ancillary movements of performers. For each example, mapping strategies will be explained in terms of their structure and functionality. [Work supported by FQRNT and MDEIE PSR-SIIRI (Québec, Canada), CNRS and PACA (France).]

9:00

5aMU3. On the evolution of controllers for sound spatialization. Jonas Braasch (Program in Architectural Acoust., Rensselaer Polytechnic Inst., Troy, NY 12180)

Dynamic changes in spatial sound attributes have played a role in classical Western music for a long time. It is known that choreographic movements of opera singers were sometimes made for acoustic considerations. Probably the first mechanical spatial sound controller is the so-called wind swell that is found in pipe organs. Already in 1712, Renatus Harris mentions how *swelling* enables the player to project the sound of the pipes *ad libitum* to nearby or further distances. With the invention of electroacoustic music, a number of electromechanical devices were developed to control spatial aspects of sound (primarily positioning sound sources in 3-D space). Typical examples are Stockhausen's rotational table (developed in the 1950s) and Manfred Krause's sound mill (1960). In this presentation, the evolutionary steps in the design of spatial sound controllers will be outlined—beginning with early, purely mechanical devices up to recent approaches including the author's participation in the development of a gestural controllable sound system based on virtual microphone control (ViMiC).

9:30

5aMU4. Deriving individualized control over music synthesis via inversion of psychophysical scaling results. William L. Martens (Schulich School of Music, McGill Univ., Montreal, QC, Canada)

Computer-based interactive music generation systems have an advantage over acoustic instruments with respect to the flexibility of the human interface for their control. Whereas human performers must conform to the relatively invariant constraints of acoustic instruments, a computer interface can be rapidly updated with regard to changes in user preferences and performance requirements.

This paper will describe a software environment that executes miniature psychophysical scaling experiments for a single user of an interactive music generation system in order to derive individualized control over music synthesis via inversion of the obtained psychophysical scaling results. A case study will be presented in which the system has been used to control parametric synthesis of musical timbres using an electronic musical keyboard, providing automatic perceptual mapping of synthesis parameters within their musically useful range. More in-depth exploration of the timbral similarities between a user-selected set of synthesis patches resulted in low-dimensional control structures that could be used to organize musical timbres in preparation for composition or performance.

10:00–10:15 Break

Contributed Papers

10:15

5aMU5. The method for capturing tilt signal of musical instruments using light sensor. Moonseok Kim (CCRMA, Dept. of Music, Stanford Univ., Stanford, CA 94305-8180)

Players of musical instruments used to express their intuition and emotion not only by playing style but also the motion of the instrument's body. Furthermore, on the ensemble, this gesture plays a role as a language to communicate to each other. Substituting some players for a computer is one part of the information to make interactions between human players and computer. To get them, tilt sensors which catch accelerations are used in general. They are used to generate underdamped signal and have to apply LPF for acquiring stable responses. This method results in delay and may affect temporal difficulties on ensemble. If the light source of the performance room is directional, the light intensity signal that is caught at photoresistors on two or more points of the instrument's surface can be used to calculate the amount of tilting. The photoresistor senses the light intensity which is determined by the distance from the light source, and differences on each value represent relative angle according to the light source direction.

10:30

5aMU6. Time marching spectral analysis of a swept sine melody model. Rama B. Bhat (Concordia Univ., 1455 de Maisonneuve West, QC H3G 1M8, Canada)

A melody is a progression of musical notes in a scale. Since the notes have identified frequencies, a melody rendering involves a quick sweep of frequencies between any two neighboring notes, and a dwell at the notes. A melody rendering is modeled as a dwell at the notes and a sinusoidal sweep in between. A time marching spectral analysis shows that the dwell time and the rate of sweep between notes have significant influence on the melody. Any two notes form the not-so-high harmonics of a periodic process which may or may not necessarily contain the fundamental and some of the lower harmonics. If a three-note combination of a first, a fifth, and an octave higher notes are taken, as the drone in Indian classical music, they have ratios of 1:3/2:2. The combination is a periodic process with half of the first note frequency as the fundamental which is missing, while the second, third, and fourth harmonics are present. A melodic rendering of seven notes or less may be construed as a periodic process with some of the harmonics varying or dancing with time. Results on a melody in pentatonic scale is presented and discussed.

10:45

5aMU7. Ignoring irrelevant dimension variability in frequency discrimination. Blas Espinoza-Varas (Commun. Sci. and Disord., OU Health Sci. Ctr., Oklahoma City, OK 73190), Hyunsook Jang (Hallym Univ., Chuncheon, South Korea), and Praveen Jajoria (OU Health Sci. Ctr., Oklahoma City, OK 73190)

The ability to ignore variability in irrelevant dimensions while attending to pitch plays a significant role in music perception. In frequency-discrimination tasks, this study examined how this ability depends on four factors: the presentation order and time proximity of target and irrelevant dimensions, the dimension conveying irrelevant variability, and training. Each observation interval displayed pairs consisting of a target (T) followed by a nontarget sinusoid (N) or the reverse (N leading T), with 20–250-ms interstimulus intervals (ISI). Frequency discrimination thresholds (FDTs) were measured in the presence or absence of irrelevant variability in the frequency, duration, or level of either T or N. The ability to ignore irrelevant variability depended on all four factors studied. Irrelevant variability in frequency or level of trailing N tones induced large and persistent elevations in the initial FDTs, especially at short ISIs. Irrelevant variability in T or N duration, T level, or leading N-tone frequency yielded only small and transient FDT elevations. With T-N pairs, irrelevant variability was harder to ignore when it was displayed after, rather than simultaneously with, the target increment. Prolonged training in irrelevant-variability conditions nullified all FDT elevations. The results are interpreted in terms of memory, attention, and discrimination-strategy constraints.

11:00

5aMU8. Application of cochlear analysis techniques to finding percussive events in electro-acoustic music. Anderson Mills (The Univ. of Texas, 702 W 32nd St., Austin, TX 78705, nodog@mail.utexas.edu)

Electro-acoustic music is a style of music for which often the pressure signal itself is the only consistent, objective representaiton of the music. A project has been undertaken at The University of Texas to automatically extract audio properties from recordings of electro-acoustic music, and the first property chosen to be extracted is a measure of the ''single damped percussive events'' (SPDEs) found in a piece of music. Because electro-acoustic music often contains sounds which are not considered musical in traditional contexts, the analysis of this type of music must begin by approaching the pressure signal without traditional musical foreknowledge, and therefore this analysis is focused on the extraction of any type of percussive event found in the pressure signal. The explored methods of analysis include bandpass filtering, spectrogram, and Patterson's auditory image model (AIM) as front-ends. The results of these different methods are compared. A human experiment to determine the nature of an SPDE and to generate data to use as a measure for the algorithms is also discussed. Self-similarity at different time scales within the measure of SPDEs during a piece of music provides useful information to a music theorist and therefore is the next step in this analysis research.

11:15

5aMU9. Frequency tracking of ecclesiastical Byzantine music frequency intervals. Kyriakos M. Tsiappoutas (Dept. of Psych., Illinois State Univ., Normal, IL 61790, kmtsiap@ilstu.edu), George E. Ioup, and Juliette W. Ioup (Univ. of New Orleans, New Orleans, LA 70148)

Two Byzantine music (BM) pieces performed by two well-established traditional chanters are quantitatively analyzed using modern frequency tracking methods to estimate the tone frequencies of the diatonic scale, one of the three main scales of BM. The two experimental versions of the diatonic scale are compared with each other and with two of many theoretical diatonic scales proposed by traditional BM theorists. This constitutes a measure of how well theoretical scales model performance. Then a direct comparison between Byzantine and common European music scales is attempted. Statistical techniques are employed to reveal any significant differences among the different scales. The attraction effect (AE) or ''elk-

sis,'' i.e., when a specific secondary tone intentionally takes on different frequency values (micro-sharpened or micro-flattened) depending on its direction in the musical script, is measured. This is part of a continuing attempt to quantify the AE empirically. Both main and secondary tones are independently examined to assess their reliance on the AE, but only secondary tones are found subject to it. A connection to the psychophysics of pitch perception illustrates whether the differences among scales are distinguishable by the human ear.

11:30

5aMU10. Idiomatic guitar synthesis. Leandro L. Costalonga and Eduardo R. Miranda (Interdisciplinary Ctr. for Comput. Music Res., Univ. of Plymouth, PL4 8AA, UK, leandro.costalonga@plymouth.ac.uk)

The aim of this research is to add intelligence to the interface between systems for computer-aided composition and synthesizers. In this paper we introduce a system that is able to play guitar music on a synthesized guitar idiomatically. There are a great number of sound synthesis techniques and computer-aided composition systems available. However, there is a general lack of intelligent interfacing between these two types of systems. For example, when one auditions the different parts of an orchestral piece, the software treats these materials identically; that is, the violin part is not played violinistically or the clarinet part clarinetistically. Musicians tend to blame the synthesizers for such poor performance quality, but if the synthesizer was controlled idiomatically the sound synthesis technique itself would be just one of the design aspects to be considered in such systems, and not the single one. Our idiomatic guitar player system includes knowledge about the biophysical constraints of the performer, the performance characteristics associated with different musical styles, and the performance styles of different individuals. [Research sponsored by CAPES Ministry of Education of Brazil.]

11:45

5aMU11. Synthesis and dynamic behavior of diffracted audio spectra. C. Eswaran and V. Vijayakumar (Faculty of Information Technol., Multimedia Univ., Cyberjaya 63100, Selangor, Malaysia)

It is shown in this paper how the intensity and phase of the harmonics of a spectrum containing a discrete set of input frequecies of unit intensity is modified by diffraction through a circular aperture. The resulting spectrum further shows dynamic behavior which is discussed in terms of a modulation parameter. Examples of synthesized audio spectra which show similarity to known musical instruments are provided as illustration of the synthesis method.

12:00

5aMU12. A system for understanding violin bowing technique through measurement of physical dynamics. Diana Young (MIT, E15-443, 20 Ames St., Cambridge, MA 02142, young@media.mit.edu)

Presented here is a custom-built sensor system for the measurement of violin bowing parameters. An improvement on previous designs, the system includes a set of accelerometers, gyroscopes, and force sensors, installed on a carbon fiber violin bow. In addition, another set of inertial sensors, as well as an electric field position sensor, are mounted on the test violin. When taken together, the data from the violin and bow are combined to achieve estimates (relative to the violin) of the three primary bowing parameters: force, speed, and bow-bridge distance. Because the system is relatively compact and lightweight, it is suitable for operation in realistic performance environments outside the laboratory setting. As such, it is possible to obtain detailed physical gesture data from real performances. The potential of the system is explored in several early studies to investigate the interaction of bowing parameters in common strokes. Distinctions between novices and experts, as well as differences between technique and style among experts, are discussed.

FRIDAY MORNING, 9 JUNE 2006 ROOM 554AB, 8:30 A.M. TO 12:00 NOON

Session 5aSC

Speech Communcation: Perception and Modeling of Speech Processes

Douglas H. Whalen, Chair
Haskins Laboratories, 300 George St., New Haven, CT 06511

Contributed Papers

8:30

5aSC1. The recent origin of human speech. Philip Lieberman (Cognit. and Linguistic Sci., Brown Univ., Providence, RI 02912-1978, philip_lieberman@brown.edu), Robert C. McCarthy (Florida Atlantic Univ., Boca Raton, FL 33431), and David Strait (State Univ. of New York at Albany, Albany, NY)

Studies of human development and swallowing show that a supralaryngeal vocal tract (SVT) capable of producing quantal vowels involves (1) facial restructuring yielding a short oral cavity; (2) a tongue that moves down into the pharynx carrying the larynx down with it; and (3) a long neck. The skeletal features of hominid fossils suggest the absence of fully human speech anatomy until 100 000 years ago. With long faces and short necks, Neanderthals and early anatomically modern humans could not have possessed SVTs capable of swallowing and of producing fully human speech. Probable [i] SVT shapes were computer-modeled from perturbations of vocal-tract area functions obtained from MRIs of adult humans. When normalized to the length of an adult human SVT, $F1/F2$ patterns fall outside of the range of Peterson and Barney's [''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **30**, 739–742 (1952)] data used by Lieberman and Crelin, ''On the speech of Neanderthal man,'' Linquistic Inquiry **2**, 203–222 (1971) to assess Neanderthal speech), but at the extreme range for [i] as measured by Hillenbrand *et al.* [''Acoustic characteristics of American-English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111 (1995)]. In neither case are the high-frequency spectra characteristic of human quantal [i]'s. $F1/F2$ patterns for unnormalized SVTs fall outside the human range for both plots.

8:45

5aSC2. Better model and decoding methods for automatic speech recognition. Douglas D. O'Shaughnessy and T. Nagarajan Li (INRS-EMT, 800 de la Gauchetiere west, Ste. 6900, Montreal PQ, H5A 1K6 Canada)

Different ways to improve the accuracy and performance of automatic speech recognition (ASR) systems are examined. Earlier research in the INRS group concentrated on improvements in adaptation techniques, to try to reduce the mismatch between system training and operating conditions. More recently, a concentration was made on improvements to the

5a FRI. AM

modeling and decoding methodology of the ASR systems. A two-level syllable model-based decoding approach is proposed here. At the first level, a syllable model-based decoding is performed to segment the utterances into syllable segments, and to identify which syllable group this segment belongs to. Then, at the second level, each segment is rescored using the phoneme model, where the possible phoneme sequences are constrained by the syllables in that syllable group. A new heuristic score is proposed to be added in hidden Markov model (HMM) decoding that indicates the degree of competition among different HMM states. Speech features are obtained from the posterior of this HMM approach and these features are used to further improve recognition results. The state evolution tracks are incorporated as features in an HMM-based recognizer. Experiments were carried out on a phoneme classification task (TIMIT) and these methods were found to show improvements compared to a baseline performance. [Work supported by NSERC-Canada and Prompt-Quebec.]

9:00

5aSC3. Development of a binaural speech transmission index. Sander J. van Wijngaarden and Rob Drullman (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands)

Although the speech transmission index (STI) is a well-accepted and standardized method for objective prediction of speech intelligibility in a wide range of environments and applications, it is essentially a monaural model. Advantages of binaural hearing to the intelligibility of speech are disregarded. In specific conditions this leads to considerable mismatches between subjective intelligibility and the STI. A binaural version of the STI was developed, based on interaural cross correlograms, which shows a considerably improved correspondence with subjective intelligibility in dichotic listening conditions. The new binaural STI is designed to be a relatively simple model which adds only few parameters to the original standardized STI, and changes none of the existing model parameters. For monaural conditions, the outcome is identical to the standardized STI. The new model was validated on set of 39 dichotic listening conditions, featuring anechoic, classroom, listening room, and cathedral environments. For these 39 conditions, subjective intelligibility (CVC-wordscore) was measured, as well as the binaural STI. The relation between binaural STI and CVC-wordscores in dichotic listening conditions closely matches the STI reference curve (standardized relation between STI and CVC-wordscore) for monaural listening. The monaural STI performs poorly in these cases.

9:15

5aSC4. New directions for a speech-based speech transmission index. Rob Drullman and Sander J. van Wijngaarden (TNO Human Factors, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands)

The standardized method for determining the speech transmission index (STI) involves the use of a specific intensity-modulated test signal. The STI is obtained from measurements on the transmission channel, usually showing reductions of the modulation depths in the received test signal. Instead of using an artificial signal, various approaches have been suggested in the literature to use speech as a test signal [cf. Payton *et al.*, J. Acoust. Soc. Am. **111**, 2431 (2002)]. Such a speech-based STI has several advantages, e.g., predicting intelligibility differences due to speaking style and enabling the evaluation of vocoders. As we encountered shortcomings and inaccuracies in the existing speech-based STI methods, we propose a new procedure for estimating the speech-based modulation transfer function (MTF) which approaches the accuracy of conventional STI implementations. The new procedure uses the cross spectrum between the transmitted and received speech signals, with special phase weighting to address the relative importance of shifted modulations of the temporal envelope. Evaluation of the algorithm on a vocoder database showed promising results, yielding an average correlation coefficient of 0.93 between the subjective CVC scores and the speech-based STI for male speech. Details of this new speech-based STI algorithm will be discussed.

9:30

5aSC5. Statistical characterization of target-to-interferer ratio of co-channel speech. Brett Smolenski (Res. Assoc. for Defense Conversion, 525 Brooks Rd., Rome, NY 13440)

Recent research in speaker identification technology suggests that it can operate in co-channel environments provided the system can have access to only the less-corrupted segments of speech. In order to identify the uncorrupted speech segments as accurately as possible, it is necessary to fully characterize the statistics of the random processes generating the uncorrupted segments. In a co-channel environment the uncorrupted speech segments are produced when one speaker's voiced speech overlaps with the other speaker's silence or unvoiced speech. Hence, if one has a statistical model of voiced, unvoiced, and silence segments, one can use this information to obtain a model of the uncorrupted speech segments. To accomplish this, statistical models that account for the observed voiced, unvoiced, and silence segment lengths are first developed. Markov models are used to account for dependencies between voiced, unvoiced, and silence segments. In addition, a model of the sampling distribution of the segmental target-to-interferer ratio (TIR) is developed and the short- and long-term correlation present in the segmental TIR signal is also explored.

9:45

5aSC6. Local versus distal speaking-rate normalization effects. Matthew Winn and Rochelle Newman (Univ. of Maryland, College Park, 0100 Lefrak Hall, College Park, MD 20742)

Individuals vary their speaking rate, and listeners typically use the speaking rate of precursor sentences to adjust for these changes (Kidd, 1989). This study examines the temporal window over which these effects occur, by comparing the size of rate normalization effects for local, distal + local, and global rate changes in a precursor phrase. A single talker produced the precursor sentence "Sarah brought a bag so Paul could get the _" at three different speaking rates, followed by a final word containing a duration-based contrast (bin versus pin). Listeners heard trials in which either the whole precursor sentence was spoken at a fast, medium, or slow rate, or (via cross-splicing) only portions of the sentence varied in rate. Subjects rated the target phoneme in the final word as either a "b" or a "p" using a six-step rating scale. Local rate changes (changes in the speaking rate of "get the") yielded significant normalization effects, seen as the shifting of the VOT boundary between p and b. The addition of distal rate changes, however, yielded no significant changes in perception on top of this local effect. This is in contrast to earlier work supporting two separate (local and distal) effects of rate changes.

10:00–10:15 Break

10:15

5aSC7. Speaker identification in hybrid synthesis: Implications for speech perception. Susan R. Hertz (NovaSpeech LLC, 215 N. Cayuga St. #48, Ithaca, NY 14850 and Dept. of Linguist., Cornell Univ., Ithaca, NY 14853) and Isaac C. Spencer (NovaSpeech LLC, Ithaca, NY 14850)

As part of an evaluation of hybrid synthesis [Hertz, Proc. IEEE Workshop on Speech Synthesis (2002)], perceptual experiments were conducted that tested the hypothesis that stressed vowels are the primary cues to speaker identification. Hybrid sentences were constructed for eight voices, including child and adult and male and female, in which stressed vowels were taken from a single human speaker, but other segments were replaced by surrogates from different sources. Some surrogates were natural speech segments; others were formant synthesized. Some matched the age or gender of the stressed vowel speaker; others did not. After being trained on six human target voices, listeners were asked to identify the hybrid stimuli, and also fully synthetic and natural stimuli (for target and nontarget voices), in terms of age, gender, and whether and how much they matched a target voice. For all categories, hybrid stimuli, in contrast to synthetic ones, were identified as accurately as natural speech, both by listeners familiar with a target voice (e.g., family members) and those

unfamiliar. The experiment demonstrated how little attention is paid to consonants and reduced vowels in determining speaker identity, and supported our hybrid approach to modeling voices. [Work supported by NIH Grant 1 R43 DC006761-01.]

10:30

5aSC8. On the relation of apparent naturalness to phonetic perceptual identification of voicing. Robert E. Remez, Claire Landau, and Daria F. Ferro (Dept. of Psych., Barnard College, 3009 Broadway, New York, NY 10027-6598)

Does the qualitative experience of speech sounds influence phonetic perception? Our studies of consonant place perception have revealed a dichotomous relation between phonetic sensitivity and naturalness. Although natural quality and phonetic sensitivity sometimes co-vary, in other conditions phonetic sensitivity is indifferent to huge variation in naturalness. New tests are reported here extending the research to the dimension of voicing, a contrast correlated with qualitatively distinct acoustic constituents in normal production. Two acoustic methods were used to create naturalness variants: (1) variation in the excitation of a synthetic voicing source and (2) variation in the bandwidth of the formant centers. A naturalness tournament was composed of items drawn from the test series, and the sensitivity of perceivers to the voicing contrast was estimated with the cumulative d' across the series in identification tests. Together, the findings show how intelligibility and naturalness can be either orthogonal or contingent aspects of consonant perception for the dimensions of place and voicing. These measures offer a tool to understand the contribution of normative functions in the perception of speech. [Research supported by NIH (DC00308).]

10:45

5aSC9. Garden-path phenomena in spoken word recognition: Gradient sensitivity to continuous acoustic detail facilitates ambiguity resolution. Bob McMurray (Dept. of Psych., E11 SSH, Univ. of Iowa, Iowa City, IA 52242), Michael K. Tanenhaus, and Richard N. Aslin (Univ. of Rochester, Ithaca, NY 14850)

Studies have demonstrated that lexical activation is gradiently sensitive to continuous acoustic detail [McMurray *et al.*, "Gradient effects of within-catagory phonetic variation on lexical access," Cognition **86**, B33–B42 (2002)]. This study investigates the consequences of this for online recognition, hypothesizing that such processes help maintain lexical alternatives and reactivate them if early interpretations are disfavored by subsequent input. Stimuli were pairs like barricade/parakeet, where, if the initial segment was misperceived, disambiguation would occur quite late. The effect of VOT on recognition was assessed for tokens favoring the competitor (barakeet with VOTs between 15 and 0). A categorical system would categorize these as equally [b]; garden path to barricade; and show difficulty reactivating parakeet after /eet/. A gradient system would activate parakeet more for VOTs near the boundary, and recover from the garden path faster. Subjects heard ten 8-step continua of this form and selected the target from screens containing the target (parakeet), competitor (barricade), and unrelated objects, while eye movements were recorded. Fixations revealed a gradient pattern of recovery time to switch from the competitor to the target was linearly related to VOT. Subsequent experiments verified that this was not due to stimulus range, nor to the visual presence of the competitor. Thus, lexical activation is sensitive to continuous detail, and this facilitates ambiguity resolution.

11:00

5aSC10. Not all feedback is created equally: The role of multisensory feedback in the perceptual learning of speech under a novel acoustic transformation. Kevin T. Webster and Lorin Lachs (California State Univ., Fresno, 2576 E. San Ramon, M/S ST11, Fresno, CA 93740, llachs@csufresno.edu)

Previous research has investigated the perceptual learning of speech modified by the frequency-inversion transformation, in which the spectral content of an acoustic pattern is rotated about an arbitrary frequency. The resulting transformed speech is almost completely unintelligible. However, given appropriate training, individuals can learn to perceive the linguistic content of the transformed spoken message. Recent work by Webster and Lachs has indicated that providing multisensory feedback during training appears to enhance the rate at which individuals adapt to the transformation. Furthermore, different types of multisensory feedback give rise to different patterns of learning. For example, learners presented with concurrent visual displays of a talkers lip movements (i.e., articulatory information) demonstrate better learning of training materials than learners presented with concurrent visual displays of text. The goal of the present experiment was to refine the technique for training perceivers in order to exaggerate differences between feedback groups in overall learning. A new control condition was also included to test the hypothesis that nonarticulatory but dynamic visual displays yield qualitatively different patterns of learning that those exhibited by perceivers trained with articulatory displays. The results provided interesting insights into the methodology necessary to test perceptual learning of frequency-inverted speech under multisensory conditions.

11:15

5aSC11. Lexical influences on the progressive facilitation during perception of assimilated speech. Cheyenne Munson, Bob McMurray (Dept. of Psych., Univ. of Iowa, E11 SSH, Iowa City, IA 52242, cheyenne-munson@uiowa.edu), and David Gow (Massachusetts General Hospital, Boston, MA 02114)

Phonological processes such as place assimilation, in which coronal sounds partially adopt the place of a subsequent noncoronal (e.g., green boat becomes green/m boat), may create ambiguity during speech comprehension, but may also paradoxically facilitate word recognition. Gow (2001) demonstrated that partial assimilation facilitates perception of postassimilation context. Two eye-tracking experiments investigated whether this progressive effect is influenced by lexical processes. Assimilated and nonassimilated adjectives (e.g., green) were spliced onto coronal or noncoronal nouns (e.g., boat or dog) to create phonologically plausible and implausible assimilation. In experiment 1, assimilation resulted in nonwords; in experiment 2, words assimilated into other words, cuing lexical competition (e.g., cat box became cat/p box). Subjects viewed a screen showing 4 pictures: a coronal noun, 1 noncoronal, and 2 fillers. Eye movements were monitored as subjects heard instructions to select a picture with a mouse. In both experiments subjects were reliably faster to look at the noncoronal target following the assimilated adjective than the nonassimilated one. This progressive effect occurred approximately 100 ms later when assimilation created competitors (experiment 2). These results support an interactive view of word recognition in which lexical processing interacts in real time with phonological and perceptual grouping and integration processes.

11:30

5aSC12. The effect of tonal information and word frequency on spoken word processing. Yao-ju Lin (Dept. of English, Linguist. Div., Natl. Taiwan Normal Univ., 162, HePing East Rd., Sec. 1, Taipei, Taiwan, 69221021@ntnu.edu.tw) and Janice Fon (Natl. Taiwan Univ., Taiwan)

This paper aims to examine the role of tonal information and word frequency in processing spoken Mandarin lexicon. Experiment 1 uses a lexical decision paradigm in which subjects are to hear high- and low-frequency bisyllabic word pairs whose segmental makeups are similar (e.g., [high frequency] jing4-zheng1 competition vs. [low frequency] jin4-sheng1 promotion). The F0 information of the stimuli are either uncontaminated, transformed for the second syllable, or transformed for both syllables. Experiment 2 uses a shadowing task with the same set of stimuli in which subjects are asked to repeat the words as quickly as possible. Preliminary results showed that tonal information is essential for lexical retrieval, especially for low-frequency words. Based on the TRACE model, it is also expected that subjects are more likely to substitute similar-sounding words of higher frequency for low-frequency words whose F0 information is manipulated.

11:45

5aSC13. Auditory, but perhaps not visual, processing of Lombard speech. Eric Vatikiotis-Bateson, Victor Chung, Kevin Lutz, Nicole Mirante, Jolien Otten, and Johanna Tan (Dept. of Linguist., Univ. of British Columbia, E270-1866 Main Mall, Vancouver, British Columbia V6T 1Z1, Canada)

Perception results for three studies are presented that address the role of Lombard speech in auditory, visual, and auditory-visual speech perception. Predictably, when presented in auditory-only conditions with masking noise, listeners recover speech recorded in noise (Lombard speech) better than speech recorded in quiet and presented with the same level of masking noise. However, there is almost no difference in listener performance when Lombard and quiet speech are presented audiovisually with masking noise. Both conditions are enhanced compared to auditory-alone conditions, but there is no indication that the facial motion correlates, demonstrated previously for quiet speech [H.C. Yehia, *et al.*, Speech Commun. **26**, 23–44 (1998)], play as strong a role in enhancing auditory-visual processing of Lombard speech, even though Lombard speech is accompanied by larger facial motions. Perhaps it is no surprise that, at a cocktail party, one leans in with an ear rather than with the eyes. [Research supported by CFI and NSERC.]

FRIDAY MORNING, 9 JUNE 2006 ROOM 556AB, 7:55 A.M. TO 12:00 NOON

Session 5aSP

Signal Processing in Acoustics, Underwater Acoustics, and Engineering Acoustics: Processing of Acoustic Vector Sensors

Paul Hursky, Cochair
Heat Light and Sound Research, 12730 High Bluff Dr., Suite 130, San Diego, CA 92130

Gerald L. D'Spain, Cochair
Scripps Inst. of Oceanography, Marine Physical Lab., 291 Rosecrans St., San Diego, CA 92106-0701

Chair's Introduction—7:55

Contributed Papers

8:00

5aSP1. Radiated noise measurements with vector sensor arrays. Joseph Clark and Gerald Tarasek (NSWCCD, Code 7340, 9500 MacArthur Blvd., West Bethesda, MD 20817-5000, joseph.a.clark1@navy.mil)

The use of arrays of vector sensors to measure source levels of radiated noise from submarines is currently being studied. Ambient noise limits the accuracy of these measurements. The directional character of the ambient noise field (anisotropy) has been found to affect measurement accuracy differently for alternative kinds of processing. In this talk, two types of vector sensor processing, linear matrix and intensity processing, will be compared theoretically and experimentally. Theoretical comparisons of achievable measurement accuracies will be developed in terms of the fourth moment of the measured signal amplitude (the variance of the sound power). Experimental comparisons will be based on data obtained during recent testing at SEAFAC, the U.S. Navy acoustic signature measurement facility in Behm Canal, Alaska.

8:15

5aSP2. Changes in ocean infrasonic energy density and vector intensity with changes in wind and ocean surface wave conditions as measured by a freely drifting vector sensor. Gerald L. D'Spain and Scott A. Jenkins (Marine Physical Lab., Scripps Inst. of Oceanogr., 291 Rosecrans St., San Diego, CA 92106)

During an experiment in 1.3-km-deep water in the Southern California Bight, the wind speed decreased from about 7 to 1 m/s over a 10-h period, and then increased to 9 m/s over the next 8-h period. Measurements made by a neutrally buoyant, freely drifting vector sensor at 920-m depth show that the corresponding changes in the potential and kinetic energy density levels of the infrasonic sound field have a frequency-dependent delay. The levels in a 0.1-Hz-wide bin at 3 Hz respond to wind-speed changes with approximately a 2-h delay, and the delay in response increases with decreasing frequency at a rate of about 1 h per hertz down to 1 Hz. The corresponding evolution in directionality of the active and reactive acoustic intensity vectors in these same frequency bins and their relationship to the surrounding bathymetry will be presented. The directional spectra of the ocean surface wave field during the experiment were reconstructed using back-refraction modeling with wave data from shore stations and offshore platforms archived by the Coastal Data Information Program to obtain estimates of the temporal variability of the nonlinear wave-wave interactions generating the ocean infrasound. [Work supported by ONR.]

8:30

5aSP3. Optimal detection with vector sensors and vector sensor line arrays. Gerald L. D'Spain (Marine Physical Lab., Scripps Inst. of Oceanogr., 291 Rosecrans St., San Diego, CA 92106), James C. Luby (Univ. of Washington, Seattle, WA 98105), Gary R. Wilson, and Richard A. Gramann (Univ. of Texas, Austin, TX 78113)

The detection performance of single vector sensors and vector sensor line arrays is degraded by nonacoustic self-noise and spatial coherence of the noise between vector sensor components. Results based on optimizing the directivity index for a single vector sensor show that the particle motion channels should always be included in the processing for optimal detection, regardless of self-noise level, assuming these levels are properly taken into account. The vector properties of acoustic intensity can be used to estimate the levels of nonacoustic noise in ocean measurements. Application of conventional, minimum variance distortionless response, and white-noise-constrained adaptive beamforming methods with ocean acoustic data collected by a vector sensor illustrates increase in spatial resolution but corresponding decrease in beamformer output and introduction of bias with increasing beamformer adaptivity. Expressions for the spatial coherence of all pairs of vector sensor components in homogeneous, isotropic noise show that significant coherence exists at half-wavelength spacing between particle motion components. For angular intervals about broadside, an equal spacing of about one wavelength for all components

provides maximum directivity index, whereas each of the component spacings should be different to optimize the directivity index for angular intervals about endfire. [Work supported by ONR.]

8:45

5aSP4. An electronic windscreen for microphones. Gary W. Elko and Jens Meyer (mh acoustics LLC, 26 Blackburn Pl., Summit, NJ 07901)

Wind-noise pickup in microphones can be highly disturbing, especially for directional microphones that attain their directivity by using velocity or spatial derivatives of the acoustic pressure. Since small directional microphones are common in audio communications systems, the problem of wind noise is well known and the typical solution is to use a windscreen. Conventional windscreens that block the wind but do not significantly alter the sound field can be very effective, but need to be relatively large to attain a good reduction in wind-noise pick-up. Small directional microphone arrays are more commonly used in hearing aids since they have been clinically shown to improve hearing in noise for hearing-impaired users. Since hearing aids are required to be compact, an effective conventional windscreen is not practical. This talk describes a novel multi-microphone wind-noise reduction algorithm that we refer to as an ''electronic windscreen.'' This new algorithm and its variants have low computational cost and are therefore applicable to commercial digital hearing aids that use multi-microphone arrays to attain directionality.

9:00

5aSP5. Vector sensor processing results using data from the Makai experiment. Paul Hursky, Michael B. Porter, Martin Siderius (Heat, Light, and Sound Res. Inc., 12730 High Bluff Dr., Ste. 130, San Diego, CA 92130), and Bruce M. Abraham (Appl. Physical Sci. Corp., New London, CT 06320)

Interest in small mobile platforms such as AUVs and gliders, and fixed autonomous systems, whose small size and power budget severely limit their ability to deploy larger arrays, has stimulated interest in vector sensors (that sense both pressure and particle velocity). Such combined sensors provide performance that can only be duplicated by larger pressure-only hydrophone arrays. In September 2005, during the Makai experiment, we deployed a four-element vector sensor array in waters off the coast of Kauai in Hawaii. We will present comparisons of processing a towed source and acoustic communications packets recorded on the four-element vector sensor array and a comparable pressure-only sensor array. We will also discuss modifications to propagation models based on Gaussian beams and normal modes to predict particle velocity, showing comparisons between model predictions and our experiment data.

9:15

5aSP6. Imaging marine geological environments with vector acoustics. Dennis A. Lindwall (Marine Geosciences Div., Naval Res. Lab., Stennis Space Ctr., MS, 39529)

Using vector acoustic sensors for marine geoacoustic surveys instead of the usual scalar hydrophones enables one to acquire 3D survey data with instrumentation and logistics similar to current 2D surveys. The addition of a collocated hydrophone allows the signal to be rectified, removing the ghost images. This concept was tested by a scaled experiment in an acoustic water tank that had a well-controlled environment with a few targets. Using vector acoustic data from a single line of sources, the three-dimensional tank environment was imaged by directly locating the source and all reflectors. [Office of Naval Research program element 61153N.]

9:30

5aSP7. Sediment sound speed measurements using buried vector sensors. Anthony P. Lyons (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804-0030, apl2@psu.edu), John C. Osler, David M. F. Chapman, and Paul C. Hines (Defence R&D Canada—Atlantic, Dartmouth, NS, Canada, B2Y 3Z7)

As part of the SAX04 sediment acoustics experiment conducted off the coast of Florida in the Gulf of Mexico, vector sensors containing three-axis accelerometers and pressure sensors were buried in the seabed. These sensors were used as the receivers to measure sediment sound speed dispersion from 300 to 3000 Hz a band in which it has been particularly difficult to make these measurements. The acoustic intensity measured by the vector sensors was combined with the experimental geometry and propagation conditions, to relate the angle of arrival to sediment sound speed as a function of frequency. Lower frequency measurements, 300 to 1200 Hz, used noise radiated from a moored research vessel as the acoustic source. Higher frequency measurements, 800 to 3000 Hz, used transmissions from an acoustic projector in a three-point mooring in the water column that could be adjusted to change the angle of ensonification. Results from the two techniques are compared with each other and previous measurements of dispersion in sandy sediments. [Work performed under ONR Grants N00014-04-1-0013 and N00014-03-1-0883.]

9:45

5aSP8. Development of a triaxial pressure-velocity probe for deep submergence applications. J. A. McConnell, P. D. Lopath, S. C. Jensen (Acoustech Corp., 4900 South Broad St., Philadelphia, PA 19112), D. H. Trivett, V. B. Biesel, J. J. Caspall (Georgia Inst. of Technol., Atlanta, GA 30332), J. F. McEachern, and R. Coughlan (Navmar Applied Science Corp., Warminster, PA 18974)

A triaxial pressure-velocity probe has been developed for the purpose of measuring the ambient acoustic noise field in the ocean at frequencies below 500 Hz. The probe consists of an omnidirectional hydrophone, a compliantly suspended sphere containing a triaxial geophone, and an electronics housing that contains a four-channel preamplifier and a digital compass. All of these components are packaged in a frame which in turn is attached to a deployment cable (i.e., a strength member) and submerged to a prescribed depth. The performance specifications for the transducers include a submergence depth rating of 6 km and a noise floor of 40 dB *re*: 1 mPa^2 per hertz. These two parameters essentially drive the design of the probe and the paper will address these topics from the standpoint of the design evolution that took place in order to meet the specifications. Other aspects of the probe's development include efforts to mitigate flow-induced noise and shielding the compass from the magnetic field emitted by the geophones. Test data obtained on a prototype unit will be presented. [Work supported by C. M. Traweek at the Office of Naval Research Code 321MS.]

10:00–10:15 Break

10:15

5aSP9. Forming first- and second-order cardioids with multimode hydrophones. J. A. McConnell (Acoustech Corp., 4900 South Broad St., Bldg. 6, Ste. LL00, Philadelphia, PA 19112), R. C. Haberman, T. A. McCormick (BBN Technologies, Mystic, CT 06355), and J. Rudzinsky (Appl. Physical Sci. Corp., New London, CT 06320)

Most everyone in sonar signal processing is familiar with forming a first-order cardioid using scalar and vector sensors, which are defined as the reciprocal analog of a monopole and dipole radiator, respectively. The second-order cardioid can be formed using scalar, vector, and dyadic sensors, in which the latter sensor is defined as the reciprocal analog of a quadrupole radiator. Multimode hydrophones configured as piezoelectric cylinders and spheres with segmented electrodes have the ability to form the both types of cardioids. This paper compares and contrasts the use of multimode hydrophones to produce each beam pattern and presents the results of several sensors that were developed. Topics relating to the sen-

sitivity, bandwidth, noise floor, beamwidth, beam steering, and directivity index will be covered. [Work supported by C. M. Traweek at the Office of Naval Research Code 321MS and L. Shumway at the Department of Homeland Security.]

10:30

5aSP10. Development of a miniature uniaxial pressure-acceleration probe for bioacoustic applications. J. A. McConnell and S. C. Jensen (Acoustech Corp., 4900 South Broad St., Bldg. 6, Ste. LL00, Philadelphia, PA 19112)

The bioacoustics community has recently expressed an interest in using collocated measurements of the acoustic pressure and the particle velocity to categorize how certain species of marine life transmit and receive sound. For most applications the frequency range of interest spans from tens of hertz to a few kilohertz. Because the test specimens and the test facilities can be relatively small, a miniature pressure-acceleration ($b-a$) probe was developed. The probe conforms to the geometry of a cube having a principal side length of 31.75 mm (1.25 in.) and contains flexural mode piezoelectric transducers to measure the acoustic pressure and one component of the acoustic particle acceleration vector. The cube also houses a low-noise two-channel preamplifier and is compliantly suspended inside a light frame. This paper highlights the design evolution of a few prototypes and addresses issues associated with the electronic noise floor of the accelerometer, mechanical cross talk between the transducers, damping of the external suspension system and transducers, and the dynamics of the signal cable. [Work supported by Acoustech Corporation IR&D.]

10:45

5aSP11. Position-invariant close-talking microphone array. Jens Meyer and Gary W. Elko (mh acoustics LLC, 26 Blackburn Pl., Summit, NJ 07901)

Close-talking microphones are commonly used for speech pick-up in high-noise environments since they can significantly improve the signal-to-noise ratio (SNR) relative to conventional pressure microphones. Increased SNR is based on the assumption that unwanted noise sources are in the far field. However, close-talking microphones have to be carefully positioned or otherwise they can have reduced SNR improvement. A new close-talking microphone array is proposed that is orientation invariant with respect to the attenuation of far-field noise sources. It is shown that by appropriate processing of the close-talking microphone array signals, one can adaptively compensate for the distance and orientation of the microphone for a near-field source. The proposed close-talking array is based on a spatially orthonormal decomposition of the sound field for a near-field source. It consists of four or more pressure microphones that are mounted in the surface of a small rigid sphere. Finally, it is shown that decomposing the sound field into spatially orthonormal modes leads to a computationally efficient implementation. The orientation compensation is applicable not just to close-taking microphones but to directional microphones in general, e.g., podium microphones or hand-held vocal microphones.

11:00

5aSP12. Intensity processing of vector sensors in the bistatic regime. Nathan K. Naluai, Gerald C. Lauchle, and Thomas B. Gabrielson (Grad. Program in Acoust., The Pennsylvania State Univ., 217 Appl. Sci. Bldg., University Park, PA 16802)

Acoustic intensity processing of signals from directional frequency analysis and recording (DIFAR) acoustic subsystems is used to enhance the detection of submerged bodies in bistatic sonar applications. In some directions, the scattered signals may be completely dominated by the incident blast from the source, depending upon the geometry, making the object undetectable by traditional scalar pressure measurements. Previous theoretical derivations suggest that acoustic vector intensity sensors, and the associated intensity processing, are a potential solution to this problem. Deep-water experiments conducted at Lake Pend Oreille in northern Idaho are described. A large body of revolution is located between a source and a number of SSQ-53D DIFAR sonobuoys positioned from 5 to 30 body lengths away from the scattering body. Scalar pressure measurements change by less than 0.5 dB when the scattering body is inserted in the field. However, the phase of the acoustic intensity component formed between the omnipressure and directional channel orthogonal to the direction of incident wave propagation varies by as much as 75. This metric is shown to be a repeatable and strong indicator of the presence of the scattering body. [Work supported by ONR, code 321MS.]

11:15

5aSP13. Processing of vector fields in underwater waveguides. Brian R. Rapids (Appl. Phys. Lab., Johns Hopkins Univ., 11100 Johns Hopkins Rd., Laurel, MD 20723-6099, brian.rapids@jhuapl.edu)

The ability to measure the scalar and vector properties of underwater acoustic fields has improved over the past decade with advancements in transducer technology. Investigations regarding the processing of such sensors have principally focused upon the exploitation of a vector sensor as a directional hydrophone and the ability to achieve significant spatial gains in an unusually small form factor (i.e., superdirectivity). These superdirective elements may then be examined in the framework of standard linear array processing techniques. Alternatively the scalar and vector sensors can be combined multiplicatively to generate estimates of the acoustic intensity vector. Several investigations regarding the detection of signals in noise with these multicomponent sensors have found that optimal linear processing produces a superior signal-to-noise ratio at the back end of the processor when compared to the intensity processor. While the intensity-based processing may not provide an optimal detector, it does provide insight into the fundamental nature of the underwater acoustic field by characterizing the energy flux density. This talk will discuss the effects of signal-model mismatch on the linear and multiplicative process by examining cases in which the assumed phase difference between pressure and particle velocity is not zero.

11:30

5aSP14. Comparative beamforming studies employing acoustic vector sensor data. Kevin B. Smith (Dept. of Phys., Naval Postgrad. School, Monterey, CA 93943), Roger T. Richards, and Philip V. Duckett, Jr. (Naval Undersea Warfare Ctr., Newport, RI 02841-1708)

Data obtained from a vector-sensor line array (VSLA), while deployed vertically at Lake Pend Oreille during calibration, are examined in the context of propagation phenomena. Both broadband (FM sweep) and cw signals are analyzed. Only linear plane-wave beamforming is considered, based on well-established array processing techniques for vector sensors [Cray and Nuttall, J. Acoust. Soc. Am. **110**, 324 (2001)]. Normalization is addressed, as is the determination of the vertical plane of propagation for proper steering of the vector components. Results show that the vector sensor array can be properly scaled and steered using a strong, single, cw source. Peak-to-sidelobe improvements of several dB were observed using vector beamforming rather than pressure-only beamforming. Arrival direction resolution using cardioid null is also presented. The vertical configuration allowed testing of predicted multipath interference behavior. As expected, arrival angle determinations with a single vector sensor is possible with the broadband signal. The difficulty of similarly resolving multipath arrival angles with one vector sensor and a single cw tone is shown. An array of vector sensors will resolve these angles. [Work completed with the support of Michael Traweek, Thomas Curtin, and Ellen Livingston, all from the Office of Naval Research, and by Daniel Deitz of NAVSEA-PMS403.]

11:45

5aSP15. Acoustic source separation via particle velocity vector measurement. Kenbu Teramoto and M. D. Tawhidul Kahn (Dept. of Mech. Eng., Saga Univ., Saga, Japan)

A novel blind source separation of a mixture of two or more voice signals has been proposed in the present paper. The separation system has been focused based on the spatio-temporal blind source separation algorithm. The proposed algorithm utilizes the linearity among the signals of sound pressure and the components of three-dimensional (x, y, and z directional) particle velocity, all of which are governed by the equation of motion. Principally, as the mechanism of blind source separation uses no *a priori* information about the parameters of convolution, filtering, as well as mixing of source signals, some simple assumptions such as the statistical independency of the linearly combined (mixed) observed signals containing zero mean as well as unit variance have been implied in the present separation algorithm. Therefore, the proposed method has successfully simplified the convoluted blind source separation problem into an instantaneous blind source separation problem over the spatio-temporal gradient spaces. An acoustic experiment with two female voices has been carried out to compare the simulated data as well. A Micro-Flown system has been adopted to evaluate the voice signals efficiently.

FRIDAY MORNING, 9 JUNE 2006 ROOM 552A, 8:00 TO 9:45 A.M.

Session 5aUWa

Underwater Acoustics: Acoustic Propagation in Sediments

Nicholas P. Chotiros, Chair

Univ. of Texas, Applied Research Lab., P.O. Box 8029, Austin, TX 78714

8:00

5aUWa1. Acoustic mode at low frequencies in marine sediments: Derivation from first principles using the method of homogenization. Allan D. Pierce, William M. Carey, and Paul E. Barbone (Boston Univ., Boston, MA 02537)

The original low-frequency Biot model factors into three characteristic modes, one being the acoustics mode, having only three significant aggregate parameters: apparent density and bulk modulus, and a numerical coefficient relating attenuation to ω^2. This prediction, although from a heuristic multiparametered model, finds an independent basis in a modification of the derivation by Burridge and Keller ["Poroelasticity equations derived from microstructure," J. Acoust. Soc. Am. **70** 1140–1146 (1981)] of the poroelasticity equations from microstructure, using only the equations of elasticity and of those for a fluid with viscosity, with no exotic mechanisms. The overall mathematical procedure is the method of homogenization. Isolation of the acoustic mode is achieved by introducing a reference velocity scale, c_{ref}, the intrinsic sound speed in the fluid alone, with the macroscopic length scale H taken as c_{ref}/ω. Mass densities of fluid and elastic media are assumed comparable, and the elastic modulus is assumed comparable to bulk modulus of fluid. The nature of sediments suggests that one needs two inner scales, the larger of which is a representative grain dimension, and the smaller of which is associated with the contact areas of adjacent grains caused by gravitational compression.

8:15

5aUWa2. Broadband sound velocity and attenuation in sandy seabottoms in shallow water. Ji-Xun Zhou and Xue-Zhen Zhang (Georgia Inst. of Technol., Atlanta, GA 30332-0405 and Inst. of Acoust., Chinese Acad. of Sci., Beijing 100080, China)

A previous paper analyzed and reviewed low-frequency (LF) field measurements conducted at 17 locations in different coastal zones around the world, and indicated that the effective acoustic attenuation curves for sandy seabottoms, derived from different field characteristics at these locations, exhibited similar magnitude and nonlinear frequency dependence in a frequency range of 50–2000 Hz. [Zhou and Zhang, "Nonlinear frequency dependence of the effective seabottom acoustic attenuation from low-frequency field measurements in shallow water," J. Acoust. Soc. Am. **117**, 2494 (2005)]. This paper aims to report the sound velocity in sandy bottoms, derived from the LF field measurements. The resultant LF velocity ratio in the sandy bottom and in the water (at the water/bottom interface) is in a range of 1.04–1.08. Combining the LF field-derived sound velocity and attenuation with the benchmark data of bottom acoustic parameters at mid- to high frequency, obtained from the SAX99 [Williams *et al*, IEEE J. Ocean. Eng. **27**, 413–428 (2002)], offers a reference data set of sound velocity and attenuation in sandy bottoms for a broadband of 50–300 000Hz. The broadband sound velocity data show a significant amount of dispersion. The attenuation data are robust enough to cast doubt on a universal assumption of a linear dependence of attenuation on frequency. [Work supported by ONR and CAS.]

8:30

5aUWa3. Dispersion and attenuation due to scattering from heterogeneities in the frame bulk modulus of sand sediments. Brian T. Hefner and Darrell R. Jackson (Appl. Phys. Lab., Univ. of Washington, Seattle, WA 98105)

A model that predicts the sound speed and attenuation of Biot fast compressional waves when there is scattering into the slow compressional wave has been developed. The scattering is due to small scale heterogeneities in the sediment properties such as the porosity or the bulk modulus of the frame that arise due to the random packing of the sediment grains. This model may account for the deviation of measured fast wave attenuation from the predictions of Biot theory that have been observed in both ocean and laboratory sediments. To compare the predictions of this theory with sound speed and attenuation measurements, the spatial statistics of the sediment properties must be known. Despite work in the granular physics community to understand the distribution of the contact forces between the grains, no direct measurements have been made on the spatial variations in the bulk modulus for dry or fluid-saturated sediments. In order to estimate these variations, we have performed molecular dynamics simulations of spherical bead packs which may approximate simple glass bead sediments. The progress of these simulations and the techniques used to estimate the bulk modulus on scales approaching the grain diameter will be presented. [This work is supported by ONR.]

8:45

5aUWa4. Seabed sediment properties from single-channel seismic profiling data. Klaus C. Leurer, Brian O'Connell, and Colin Brown (Dept. of Earth and Ocean Sci., Natl. Univ. of Ireland, Galway, Galway, Ireland, klaus.leurer@nuigalway.ie)

An acoustic model is presented for the calculation of velocity and attenuation of compressional waves in unconsolidated marine sediments. The model uses the concept of a standard sediment structure that assumes spherical grain shapes and a bimodal grain-size distribution in which the

small particles represent the clay fraction. The description of the acoustic response of this bimodal sediment model is based on Biot's poroelastic theory and includes two viscoelastic extensions to the original formulation to enable the resulting combined model to cover a wide range of unconsolidated marine sediments. The first extension, an effective-grain model, is introduced to include the fine-grained spectrum of these sediments and the second, a viscoelastic grain-contact model, to account for their unconsolidated granular nature. The acoustic model provides the basis for a neural-network (NN) inversion scheme to extract sediment physical properties from measured single-channel seismic reflection profiling data. Input parameters for the NN are travel times, attenuation, and reflectivities, which are determined from the seismograms. Preliminary application to real reflection data has provided reasonable results within the framework of our standard sediment model for sediment thickness, density, porosity, sound velocity, and concentration of fine-grained material.

9:00

5aUWa5. Gassy sediment acoustics: Low-frequency sound-speed measurements paired with computed x-ray tomography imaging in reconstituted natural sediments. Preston S. Wilson (Mech. Eng. Dept., Univ. of Texas at Austin, 1 University Station C2200, Austin, TX 78712-0292, pswilson@mail.utexas.edu), Allen H. Reed, Warren T. Wood (Stennis Space Ctr., MS 39529-5004), and Ronald A. Roy (Boston Univ., Boston, MA 02215)

Various models of sound propagation in gas-bearing sediments have appeared in the literature but they are largely unverified by experiment. This is primarily due to the difficulty of assessing the bubble-size distribution and global void fraction in optically opaque sediments. We successfully measured the sound speed in reconstituted natural mud and kaolinite sediments containing varying fractions of biogenic gas bubbles. We used a one-dimensional acoustic resonator technique to determine the low-frequency (100–2000 Hz) sound speed of sediment samples contained within cylindrical acrylic core liners. High-frequency (400 kHz) time-of-flight sound-speed measurements were obtained on a subset of the samples. The sediment samples were contemporaneously imaged with a computed X-ray tomography system that yielded independent estimates of the void fraction and bubble-size distribution. Measured sound speeds ranged from 1520 m/s for gas-free mud to as low as 280 m/s for gassy mud and 110 m/s for gassy kaolinite. CT-derived mud void fractions ranged from 0.001 to 0.002, and the kaolinite void fraction was 0.005. The results of these measurements will be compared to existing theory, but a preliminary analysis shows reasonable agreement between measurement and Wood's suspension model. [Work supported by NRL and the UT College of Engineering.]

9:15

5aUWa6. An experimental laboratory investigation of the low-frequency acoustics of seagrasses. Preston S. Wilson (Mech. Eng. Dept., Univ. of Texas at Austin, 1 Univ. Station C2200, Austin, TX 78712-0292, pswilson@mail.utexas.edu) and Kenneth H. Dunton (Univ. of Texas, Port Aransas, TX 78373-5015)

The acoustics of seagrass beds can impact sonar operation in shallow water and may also be exploited for remotely assessing the health of estuarine and littoral ecosystems. The acoustics are dominated by air channels (aerenchyma) within the plants and air bubbles produced by the plants during photosynthesis. *In situ* studies of seagrass beds have confirmed their unique acoustic properties with respect to side-scan sonar backscatter and propagation effects. The results of these studies suggest a richness of exploitable phenomena, but corresponding forward theoretical models have yet to appear. To address this question, low-frequency acoustic laboratory experiments were conducted on three freshly collected seagrass species, *Thalassia testudinum*, *Syringodium filiforme*, and *Halodule wrightii*. A one-dimensional acoustic resonator technique was used to assess the effective acoustic properties and gas content of individual plants, leaves, and rhizomes. We successfully tracked bubble formation and transport from above- to below-ground tissues and present our results with respect to model predictions based on independent optical gas content estimates obtained via microscopic cross-section imagery. [Work supported by the UT College of Engineering and the Texas Sea Grant College Program Grant R/ES-87.]

9:30

5aUWa7. Ultrasonic velocity measurements for synthetic gas-hydrate samples. Hailan Zhang, Dong Wang, Haibo Zhao, Weijun Lin (Inst. of Acoust., Chinese Acad. of Sci., Beijing 100080 China), and Xiuming Wang (CSIRO Petroleum, Bentley WA 6102, Australia)

Laboratory ultrasonic methods offer a way of studying acoustic velocity of a gas-hydrate bearing formation. By measuring ultrasonic velocities of the gas-hydrate samples in various temperature and pressure conditions, more effective inversion techniques can be developed to quantitatively evaluate gas-hydrate concentration and distributions. Low-temperature laboratory measurements of compressional velocities in compacted samples are conducted. These gas-hydrate samples are synthesized by using various densities at various pressures and temperatures. At $-10°$C, the compressional velocities of the compacted gas-hydrate samples are from 2440 to 3570 m/s with the density range from 475 to 898 kg/m^3. Compressional velocity measurements are made where the temperature and pressure can be controlled. When the pore pressure increases from 10 to 40 MPa, the compressional velocities of the sample increases from 2340 to 2600 m/s at 1.5°C. When the temperature decreases from 10° to $-13°$C, the compressional velocity will increase from 3600 to 3800 m/s at a pore pressure of 6 MPa. Our experimental results are qualitatively in agreement with those of weighted average model and the Biot-Gassmanns model when the gas-hydrate concentration in a sediment bearing sand is about 20%. [Work supported by National Natural Science Fundation of China, No. 10534040.]

Session 5aUWb

Underwater Acoustics: Acoustic Backscattering and Clutter

Nicholas P. Chotiros, Chair
Univ. of Texas, Applied Research Lab., P.O. Box 8029, Austin, TX 78713

Contributed Papers

10:30

5aUWb1. Statistical characterization of sonarlike clutter observed on the STRATAFORM during the 2003 Acoustic Clutter Experiment in the 400–1500-Hz region. John R. Preston (Appl. Res. Lab., The Penn State Univ., P.O. Box 30, State College, PA 16804)

In 2003 ONR sponsored the Acoustic Clutter Experiment to study shallow-water scattering and clutter in the STRATAFORM area off New Jersey. Sources were bistatically received coherent pulses. The receiver was the five octave research array (used horizontally). The STRATAFORM has benign surface morphology but contains many subsurface features. MIT researchers have shown fish to be a primary source of the observed clutter and reverberation. *K* distributions, with their shape and scale parameters, are used to describe non-Rayleigh behavior. Statistical characterization is presented versus location. The ''bandwidth'' effect is shown where the shape parameter first decreases inversely proportional to bandwidth but then increases back toward the Rayleigh distribution at higher bandwidths. The shape parameter estimates for the 2003 data are well fit by an elongated patch model of Abraham and Lyons. It is shown that shape parameter estimates are about the same for the 2003 and 2001 data taken in the same area. The main differences between the data sets are that the typical scatter sizes/strengths seem to have increased. This is consistent with observations by Nero of NRL and by MIT researchers of higher and more concentrated fish populations observed in 2003. [Work supported by ONR Code 32, Grant N00014-05-1-0156.]

10:45

5aUWb2. The effect of multi-path on synthetic aperture sonar image statistics. Shawn F. Johnson, Anthony P. Lyons, Douglas A. Abraham (Appl. Res. Lab., The Penn State Univ., 117 Appl. Sci. Bldg., State College, PA 16804), and Kevin Williams (Univ. of Washington, Seattle, WA 98105)

The K-distribution shape parameter has been shown to be a useful metric of seafloor reverberation statistics. Using this metric, synthetic aperture sonar (SAS) data collected during the Office of Naval Research sponsored SAX04 experiment in the Gulf of Mexico during October 2004 has been processed to examine image statistics as a function of range from the sonar. SAS images are unique in that the resolution cell size does not increase with range as for traditional sonar. Analysis of the SAX04 data has shown that in spite of the fixed resolution cell size, shape parameter estimates were found to increase with range from the sonar. The increase can be attributed to multi-path propagation. At increasing distance from the sonar, scattered returns arising from additional propagation paths arrive in a specified time window, with the result that two or more resolution cells are contributing, which leads to a larger estimate of the shape parameter than would be expected when only the direct path contributes. Model results using the sonar system geometry and seafloor scattering theory will be used to demonstrate this effect. [Work supported by ONR Codes 321 and 333. First author supported by a National Defense Science and Engineering Graduate Fellowship.]

11:00

5aUWb3. Time evolution of acoustic scattering from a bioturbated seafloor. Anthony L. Gerig, Anthony P. Lyons (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA, 16804, alg14@psu.edu), and Eric Pouliquen (NATO Undersea Res. Ctr., La Spezia, Italy)

Knowledge of the structure, composition, and time evolution of the seabed is essential to modeling acoustic seafloor scattering. The purpose of this work is to quantify the change in a bioturbated seafloor profile with time, and use the results to predict concomitant changes in a backscattered acoustic wave. To accomplish this task, digital stereo photograph pairs of an approximately 0.5-$-m^2$ patch of bioturbated seafloor were taken every 10 min over the course of several days at several different locations off the Florida coast as a part of the Office of Naval Research sponsored SAX04 experiment. Photographs were captured, calibrated, and processed according to standard photogrammetric methods to produce surface height fields of the seafloor. The time sequence of fields was then used to estimate height decorrelation with time. Fields were also used to calculate height spectra for the seafloor, and to estimate spectral decorrelation with time. These estimates were subsequently compared for differences in decorrelation time between frequencies. Finally, quantitative results were used as input for the small perturbation surface scattering model to predict the time evolution of acoustic backscatter. Preliminary results indicate that seafloor decorrelation occurs on the scale of hours. [Work supported by ONR.]

11:15

5aUWb4. Statistical and change-detection analyses using quantitative side-scan-sonar data. Jerald W. Caruthers (The Univ. of Southern Mississippi, 1020 Balch Blvd., Stennis Space Ctr., MS 39529)

This paper discusses recent basic, high-frequency, bottom-backscattering research conducted in conjunction with sediment-classification projects and the ONR bottom scattering research projects, KauaiEx and SAX04. This work has a bearing on statistical analyses and bottom change detection relevant to the detection of objects on or near the bottom in shallow water. The research discussed here includes the analyses of data from three surveys made between 2002 and 2004. The primary contributions of this work are based on the use of probability density functions (PDFs) of the backscatter signals to specify the various regimes of bottom sediments and to detect spatial changes in these regimes based on chi-square and other tests applied to these PDFs. A simple, quantitative test for non-Rayleighness of a PDF was also applied. [This work was supported by the ONR.]

11:30

5aUWb5. Low-frequency scattering from a mud volcano located offshore Sicily. Anthony L. Gerig and Charles W. Holland (Appl. Res. Lab., Penn State Univ., P.O. Box 30, State College, PA, 16804)

Mud volcanoes are objects that form on the seafloor due to the emission of gas and fluidized sediment from the earths interior. They vary widely in size, can be exposed, proud, or buried, and are of interest to the underwater acoustics community as potential sources of active sonar clutter. Coincident seismic reflection data and low-frequency bistatic scattering data were gathered from one such volcano buried in the Straits of Sicily. The bistatic data were generated using a pulsed piston source and a 64-element horizontal array, both towed over the top of the volcano. The purpose of this work was to appropriately model low-frequency scattering from the volcano using the bistatic returns, seismic bathymetry, and knowledge of the general geoacoustic properties of the area's seabed to guide understanding and model development. Ray theory, with some approximations, was used to model acoustic propagation through overlying layers. Due to the volcano's size, scattering was modeled using geometric acoustics and a simple representation of volcano shape. Modeled bistatic data compared relatively well with experimental data, although some features remain unexplained. Results of an inversion for the volcanos reflection coefficient indicate that it is acoustically softer than expected. [Work supported by ONR, NURC.]

11:45

5aUWb6. Backscattering by layered elastic seafloors: The equivalent input backscattering strength approach. Laurent Guillon (Ecole Navale, IRENav, BP 600, 29240 Brest Armees, France, guillon@ecole-navale.fr)

In low-frequency experiments, an acoustic signal may penetrate with great depth into the seafloor and interact with submerged structures. Consequently, for many applications, there is need for a model to predict the backscattering strength (BS) of a layered seafloor containing elastic layers. A few models exist; however, two problems exist for their use: first, as most of them are based on specific configurations they are poorly adaptable, and second, very few deal with both layered and elastic seafloors. This paper proposes to extend to the case of elastic layers the equivalent input backscattering strength (EIBS) approach defined for a fluid layered seafloor [L. Guillon and X. Lurton, ''Backscattering from buried sediment layers: The equivalent input backscattering strength model,'' J. Acoust. Soc. Am. **109**, 122–132 (2001)]. The BS is calculated as follows: each layer BS is defined with a specific model, then each local BS is modified to take into account the various effects of stratification, and finally the different modified BS are added to provide the total BS. Two different local models for the individual BS are tested within this general frame. The numerical results obtained show the flexibility of the EIBS approach and its ability to reveal specific effects of layering on the global BS. [Work supported by IFREMER.]