CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 119, No. 4 April 2006

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

(Continued)

CONTENTS—*Continued from preceding page*

Polling the effective neighborhoods of spoken words with the verbal transformation effect

James A. Bashford, Jr., Richard M. Warren, and Peter W. Lenz
Department of Psychology, University of Wisconsin—Milwaukee, PO Box 413, Milwaukee, Wisconsin 53201-0413
bashford@uwm.edu, rmwarren@uwm.edu, plenz@uwm.edu

Abstract: Studies of the effects of lexical neighbors upon the recognition of spoken words have generally assumed that the most salient competitors differ by a single phoneme. The present study employs a procedure that induces the listeners to perceive and call out the salient competitors. By presenting a recording of a monosyllable repeated over and over, perceptual adaptation is produced, and perception of the stimulus is replaced by perception of a competitor. Reports from groups of subjects were obtained for monosyllables that vary in their frequency-weighted neighborhood density. The findings are compared with predictions based upon the neighborhood activation model.

PACS numbers: 43.71.An [JH]
Date Received: December 12, 2005 **Date Accepted:** January 28, 2006

1. Introduction

Most modern theories of spoken word recognition [TRACE (McClelland and Elman 1986), SHORTLIST (Norris, 1994), MERGE (Norris *et al.*, 2000), NAM (Luce and Pisoni, 1998), and PARSYN (Luce *et al.*, 2000)] assume that acoustic-phonetic input activates a set of structurally similar verbal representations in memory, which compete with one another in the process of speech perception. These activation-competition models of speech perception are supported by evidence from a variety of paradigms (e.g., shadowing, perceptual identification, and lexical decision), which have shown that processing speed and accuracy for a verbal stimulus are influenced by both its neighborhood density (i.e., the number of structurally similar, lexical neighbors) and by its neighborhood frequency (the sum of the word frequencies of its lexical neighbors). Words having large numbers of high-frequency lexical neighbors are generally processed more slowly and less accurately than words having only a few, low-frequency lexical neighbors (Goldinger *et al.*, 1989; Cluff and Luce, 1990; Luce *et al.*, 1990; Luce and Pisoni, 1998; Vitevitch and Luce, 1998, 1999).

The present study employs a novel procedure for directly accessing the neighborhoods of spoken words. Studies dealing with neighborhood effects typically employ a computational procedure that considers the effective neighbors of a stimulus to be those words in the lexicon that differ from the stimulus by the addition, deletion, or substitution of a single phoneme (Landauer and Streeter, 1973; Luce and Pisoni, 1998). Although this *a priori* method of identifying competitors has proven capable of predicting stimulus differences in word recognition, it likely provides only an approximation to the effective neighborhood of a stimulus (Luce and Pisoni, 1998, p. 16) and hence may benefit from further experimental refinement. The present study pursues this goal, using an auditory illusion known as the verbal transformation effect (Warren, 1961), which can be used to induce listeners to perceive and call out the effective neighbors of the stimulus word or syllable.

When listeners are presented with a clear recording of a syllable or word that is repeated over and over without change, they typically hear an abrupt and compelling illusory change to a different word or syllable (Warren and Gregory, 1958; Warren, 1961). Verbal transformation (VT) appears to result from the operation of two concurrent processes (Warren, 1976): (1) a repetition-induced adaptation effect that progressively lowers the activation level of the initially dominant neural representation corresponding to veridical perception and (2) a

repetition-induced verbal summation[1] that progressively increases the activation level of the most salient of the neural representations that are structurally similar to the stimulus. VT is considered to occur when the diminishing activation level of the perceived stimulus representation is exceeded by that of the next-most highly activated representation. Hence, it appears that VT may offer a means of directly accessing the sets of representations that compete in the process of spoken word recognition. The present study provides a preliminary test of this hypothesis. In order to maximize neighborhood effects upon transformations, neighborhood density and word frequency were manipulated conjointly, with half the stimuli having large numbers of high-frequency neighbors and half having only a few, low-frequency neighbors. The focus of analysis in this initial study is upon listeners' initial transformations reported for each stimulus, categorized in terms of lexicality, word frequency, and structural similarity to the stimulus.

2. Method

The 144 listeners in this study (three groups of 48) were undergraduate student volunteers from the University of Wisconsin—Milwaukee who were paid for their participation. All listeners were native monolingual English speakers who reported having no hearing problems and had normal bilateral hearing, as measured by pure tone thresholds of 20 dB HL or better at octave frequencies from 250 to 8000 Hz.

The test stimuli were 12 monosyllables that were selected to provide three replications of a 2×2 factorial crossing of stimulus lexicality and frequency-weighted neighborhood density (FWND), calculated using the procedure of Newman *et al.* (1997). This selection yielded three exemplar stimuli in each of four stimulus conditions: (1) the high-density lexical monosyllables (i.e., words with many neighbors), "boat," "cane," and "let," having log word frequencies $[\log_{10}(\text{frequency})+1]$ of 2.86, 2.15, and 3.58, respectively, and having neighborhood density scores of 32, 33, and 30, and FWND scores of 67.7, 67.2, and 72.1, respectively; (2) the high-density nonlexical monosyllables, "dake," "leet," and "nane," having density scores of 25, 40, and 25, and FWND scores of 53.3, 82.4, and 53.2, respectively; (3) the low-density lexical monosyllables, "watch," "gouge," and "jibe," having word frequency scores of 1, 1, and 2.91, respectively, and having neighborhood density scores of 5, 3, and 4, and FWND scores of 11.5, 5.7, and 6.4, respectively; and (4) the low-density nonlexical monosyllables, "powsh," "chibe," and "gouk," having density scores of 5, 3, and 1, and FWND scores of 9.4, 4.2, and 1.8, respectively. An additional monosyllabic word, "lean," was produced for use as a warm-up stimulus that preceded the experimental stimuli.

The stimuli were digitally recorded (44.1-kHz sampling, 16-bit quantization) by a male speaker having an average voicing frequency of approximately 100 Hz and no obvious regional accent, who produced the 500-ms monosyllabic stimuli. An additional 250-ms segment of digital silence was then appended to each capture to emphasize syllable boundaries and minimize any tendency for perceptual resegmentation of the stimuli (for example, when the stimulus "ace" repeats without an appreciable silent gap between repetitions, it is readily reparsed perceptually to yield "say"). Although this is an interesting phenomenon in its own right, this type of transformation could lead to the activation of neighborhoods other than those intended for the present study. The 500-ms stimulus, together with the added 250-ms silent gap, was digitally iterated to produce a 4-min test stimulus that provided 320 repetitions of the monosyllable.

The 144 listeners were randomly divided into three groups of 48, and each group of listeners received five VT stimuli, which included the initial, warm-up stimulus "lean" and one stimulus drawn from each of the four experimental conditions described above. The assignment of specific experimental stimuli to groups followed the order listed in the stimulus section: That is, the first group of listeners received the experimental stimuli boat, dake, watch, and powsh; the second group received cane, leet, gouge, and chibe; and the third group received the experimental stimuli let, nane, jibe, and gouk. The order of presentation for the experimental stimuli was pseudorandom, with the restriction that each stimulus was presented an equal number of times in each serial position across listeners in each group.

TABLE I. Types of initial verbal transformations reported for lexical and nonlexical stimuli having high- versus low-frequency-weighted neighborhood density (FWND). Percentages are given for the four categories of transformations reported in each experimental condition (see text). Values in parentheses are mean log transformed word-frequency scores for lexical forms reported [$\log_{10}$(frequency)+1]. The values of *N* in the table are the numbers of trials, out of 144, producing at least one verbal transformation for a given type of experimental stimulus.

	Lexical neighbor	Lexical nonneighbor	Nonlexical neighbor	Nonlexical nonneighbor
Words				
High FWND *N*=129	61.2% (3.1)	13.2% (3.1)	15.5%	10.6%
Low FWND *N*=132	9.9% (1.5)	48.5% (3.6)	25.0%	16.7%
Nonwords				
High FWND *N*=125	88.0% (2.6)	0.0% (−)	4.8%	7.2%
Low FWND *N*=133	18.0% (1.3)	45.1% (2.4)	28.6%	8.3%

Testing was performed in a sound-attenuating chamber, with the VT stimuli delivered diotically through Sennheiser HD 250 Linear II Headphones at a slow-rms peak level of 70 dBA SPL, as measured using a Brüel & Kjaer model 2230 precision integrating sound level meter. For each VT stimulus presented, listeners were instructed to call out what the voice was saying at the stimulus onset, and then to call out what the voice was saying any time a change was heard. Listener's responses were transcribed by the experimenter and also audiotaped for subsequent verification.

3. Results

Out of 576 experimental trials (144 listeners × 4 experimental stimuli), a total of 519 trials, or 90.1%, yielded at least one verbal transformation, and hence provided initial illusory forms data for use in this study (median time to first change was about 30 s). The percentages of trials producing at least one change for the four experimental conditions ranged from 86.8% to 92.4% and did not differ significantly by *Z* tests, $Z \leqslant 1.52$, $p \geqslant 0.129$, for the significance of a difference between proportions (Bruning and Kintz, 1977). All transforms reported by our monolingual English listeners were phonotactically legal. For the primary data analysis, transformations were categorized as one of four types: (1) lexical neighbors (words differing from the stimulus by a single phoneme), (2) lexical nonneighbors (words differing by two or more phonemes), (3) nonlexical neighbors, and (4) nonlexical nonneighbors. Table I presents the percentages of these four types of forms reported for the stimuli in each of the experimental conditions. The average log transformed word-frequencies (Kucera and Francis, 1967) for lexical forms reported in each condition are also included in parentheses.

The most frequently reported initial transformations were lexical neighbors of the stimuli (constituting about 45% of responses overall). However, it can be seen in Table I that this percentage varied substantially as a function of stimulus frequency-weighted neighborhood density (FWND). Lexical neighbors comprised about 75% of the transformations evoked by high-FWND stimuli, but comprised only about 14% of those reported for the low-FWND stimuli. This difference in percentages was significant for both the lexical stimuli, $Z=8.69$, $p < 0.0001$, and the nonlexical stimuli, $Z=11.24$, $p < 0.0001$. In contrast, stimuli that were low rather than high in FWND produced a much larger percentage of lexical nonneighbors, with overall averages of about 47% vs. 7%. This difference in percentages also was significant for both the lexical stimuli, $Z=6.16$, $p < 0.0001$, and nonlexical stimuli, $Z=8.57$, $p < 0.0001$. It is

also of interest that the low- and high-FWND stimuli did not differ significantly in the time required to evoke VT, $F(1,143)=0.12$, $p>0.7$. This suggests that the lexical nonneighbors reported predominantly for low-FWND stimuli were equivalent in their salience to that of the lexical neighbors reported for the high-FWND stimuli. Finally, about 19% of reported VT forms were nonlexical neighbors of the stimuli. The majority of these forms (72%) involved the external addition of a vowel or consonant in initial or final position, and thus preserved the component phonemes of the stimulus. About 22% involved the substitution of a consonant or vowel, and these forms were most frequently reported for nonlexical stimuli. Deletion of an individual consonant did occur, but rarely (6%), and then only with nonlexical stimuli.

4. DISCUSSION

The predominance of lexical neighbors obtained as transforms for the high-FWND stimuli is consistent the neighborhood activation model (NAM), which currently considers that words differing from the stimulus by a single phoneme are its competitors during word recognition. However, the large percentage of lexical nonneighbors obtained as transforms for the low-FWND stimuli (typically differing from the stimuli by two phonemic changes) suggests that the "minimum-phoneme-distance rule" may be too restrictive, especially for stimuli having only a few, low-frequency neighbors. The present results also suggest that the dominance of lexical nonneighbors as transforms for the low-FWND stimuli is due to another factor conventionally considered critical in word recognition: competitor word frequency. Not surprisingly, the log word-frequency scores for lexical neighbors reported as transforms for the low-FWND stimuli were correspondingly low, with an average score of 1.37 (standard error=0.11). In contrast, the corresponding score for lexical nonneighbors was 2.93 (standard error=0.13). Lexical nonneighbors reported for the high-FWND lexical stimuli, though constituting a smaller percentage of VTs, showed a similarly high-word-frequency score, with an average value of 3.11 (standard error=0.29). Thus, it appears that word frequency may, under some conditions, override structural similarity in determining the effective neighborhood of a stimulus.

The results of this initial study suggest that the VT effect may indeed serve as a useful tool for directly accessing and identifying the salient competitors for lexical and nonlexical stimuli. In particular, the present study suggests that the number of high-frequency "remote neighbors" of a stimulus (i.e., those differing by two phonemes) may also contribute to measurable differences in processing speed and accuracy, especially for stimuli having sparse, low-frequency lexical neighborhoods, as conventionally calculated.

Acknowledgements

This work was supported by Grant No. DC 000208 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health.

References and Links

[1]Verbal summation was described and named by Skinner (1936), who employed a recording of repeating faint and indistinct speech sounds, which became organized into illusory words with continued listening. A more recent example of verbal summation, based upon clear speech sounds was reported by Warren *et al.* (1990), who employed a recording of repeating loud and clear sequences of three or more brief (30–100ms) isochronous steady-state vowels and reported that listeners could not identify the brief vowels, but after several repetitions of the sequences reported hearing syllables and words. Subsequently, Warren, *et al.* (1996) reported that spectrograms showed a structural similarity between the vowel sequences and listeners″ synchronous productions of the verbal forms being heard.

Bruning, J. L., and Kintz, B. L. (**1977**). *Computational Handbook of Statistics* (Scott Foresman, Glenview, IL).
Cluff, M. S., and Luce, P. A. (**1990**). "Similarity Neighborhoods of Spoken Two-Syllable Words: Retroactive Effects on Multiple Activation," J. Exp. Psychol. Hum. Percept. Perform. **16**, 551–563.
Goldinger, S. D., Luce, P. A., and Pisoni, D. B. (**1989**). "Priming lexical neighbors of spoken words: Effects of competition and inhibition," J. Mem. Lang. **28**, 501–518.
Kucera, H., and Francis, W. N. (**1967**). *Computational Analysis of Present-Day American English.* (Brown U. P., Providence, RI).

Landauer, T., and Streeter, L. A. (**1973**). "Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition," J. Verbal Learn. Verbal Behav. **12**, 119–131.
Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.
Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (**1990**). "Similarity neighborhoods of spoken words," in *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*, edited by G. T. M. Altmann (MIT, Cambridge, MA).
Luce, P. A., Goldinger, S. D., Auer, E. T. J., and Vitevitch, M. S. (**2000**). "Phonetic priming, neighborhood activation, and PARSYN," Percept. Psychophys. **62**, 615–662.
McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**, 1–86.
Newman, R. S., Sawusch, J. R., and Luce, P. A. (**1997**). "Lexical neighborhood effects in phonetic processing," J. Exp. Psychol. Hum. Percept. Perform. **23**, 873–889.
Norris, D. (**1994**). "Shortlist: A connectionist model of continuous speech recognition," Cognition **52**, 189–234.
Norris, D., McQueen, J. M., and Cutler, A. (**2000**). "Merging information in speech recognition: Feedback is never necessary," Behav. Brain Sci. **23**, 299–325.
Skinner, B. F. (**1936**). "The verbal summator and a method for the study of latent speech," J. Psychol. **2**, 71–107.
Vitevitch, M. S., and Luce, P. A. (**1998**). "When words compete: Levels of processing in perception of spoken words," Psychol. Sci. **9**, 325–329.
Vitevitch, M. S., and Luce, P. A. (**1999**). "Probabilistic phonotactics and neighborhood activation in spoken word recognition," J. Mem. Lang. **40**, 374–408.
Warren, R. M. (**1961**). "Illusory changes of distinct speech upon repetition—the verbal transformation effect," Br. J. Psychol. **52**, 249–258.
Warren, R. M. (**1976**). "Auditory illusions and perceptual processes," in *Contemporary Issues in Experimental Phonetics*, edited by N. J. Lass (Academic, New York).
Warren, R. M., and Gregory, R. L. (**1958**). "An auditory analogue of the visual reversible figure," Am. J. Psychol. **71**, 612–613.
Warren, R. M., Bashford, J. A., Jr., and Gardner, D. A. (**1990**). "Tweaking the lexicon: Organization of vowel sequences into words," Percept. Psychophys. **47**, 423–432.
Warren, R. M., Healy, E. W., and Chalikia, M. H. (**1996**). "The vowel-sequence illusion Intrasubject stability and intersubject agreement of syllabic forms," J. Acoust. Soc. Am. **100**, 2452–2461.

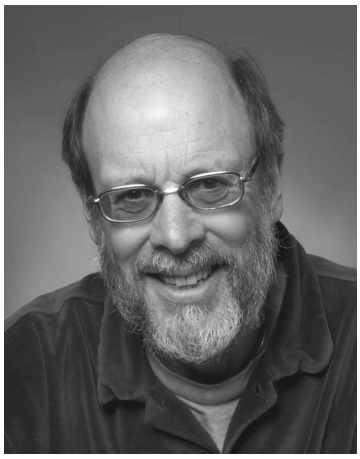
ACOUSTICAL NEWS—USA

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

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

New Fellows of the Acoustical Society of America

Carol Espy-Wilson
For contributions to speech communication and mentoring



Lee A. Miller
For elucidating the mechanisms of insect countermeaures to echolocating bats

Shrikanth Narayanan
For contributions in speech production and imaging

Simon D. Richards
For contributions to high frequency acoustics in littoral waters

Gail ter Haar
For contributions to ultrasound safety and standardization

 0001-4966/2006/119(4)/1897/4/$22.50

F. V. Hunt Postdoctoral Research Fellowship awarded to Erica E. Bowden

The 2006-07 F. V. Hunt Postdoctoral Research Fellowship in Acoustics has been awarded to Erica E. Bowden. Erica Bowden received a B.S. degree in Architectural Engineering from Kansas State University, Manhattan, KS. She is now enrolled in the University of Nebraska and expects to receive a Ph.D. degree in Architectural Engineering in May 2006. Her Ph.D. thesis is titled "Relating indoor noise criteria systems to human performance and noise perception."During her Hunt Fellowship year, Ms. Bowden will undertake a research program in the Department of Environmental Medicine at the Sahlgrenska Academy at Göteborg University in Göteborg, Sweden. The subject of her research will be "Establishing a hospital soundscape through qualitative and quantitative observation."

The Hunt Fellowship is granted each year to an ASA member who has recently received his or her doctorate or will be receiving the degree in the year in which the fellowship is to be granted. The recipient of the fellowship is that individual who, through personal qualifications and a proposed research topic, is judged to exhibit the highest potential for benefiting any aspect of the science of sound and promoting its usefulness to society. Further information about the fellowship is available from the Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. Telephone: 516-576-2360; Fax: 516-576-2377. Electronic mail: asa@aip.org; Web: asa.aip.org/fellowships.html

Reports from F. V. Hunt Postdoctoral Fellows

Report of the 25th F. V. Hunt Post-Doctoral Fellow (2002–03)

Constantin C. Coussios, Magdalen College, Oxford, OX1 4AU, U.K.

As the 2003-2004 F. V. Hunt Postdoctoral Fellow, I had the privilege of working under the guidance of Professor Ronald A. Roy in the Department of Aerospace and Mechanical Engineering at Boston University. The primary objective of my research was to develop a method for monitoring and controlling cavitation induced by high-intensity focused ultrasound (HIFU) in human tissue.

Real-time detection and monitoring of inertial cavitation are becoming increasingly important in the context of therapeutic ultrasound in general and thermal ablation by high-intensity focused ultrasound (HIFU) in particular. Earlier studies at Boston University and elsewhere had indicated that generating and sustaining inertial cavitation could lead to as much as a fivefold increase in the rate of tissue heating during HIFU exposure. However, once nucleated and excited, cavitating microbubbles tend to grow unstably towards the HIFU transducer, shielding the original focus and causing undesirable thermal damage in the prefocal region. Developing an ability to monitor and detect inertially cavitating microbububbles in real time is therefore key to preventing undesired bioeffects and to controlling the extent of the thermally ablated region. By performing simultaneous cavitation detection in tissue-mimicking materials using both a passive cavitation detector (PCD) and a clinical 5 MHz ultrasound imaging head, it was concluded that the appearance of a hyperechogenic region in B-mode images is neither a necessary nor a sufficient condition for cavitation to have occurred at the HIFU focus. In addition, the PCD system developed enabled real-time monitoring of the microbubble field, as well as of the bubble cloud shift towards the transcranial ultrasound therapy system transducer during prolonged exposure. Excellent agreement was found between the peak-detected PCD value and broadband noise emissions in the frequency-domain, which are indicative of inertial cavitation activity: it was therefore concluded that the developed peak-detected PCD system constitutes a cost-effective yet sensitive means of real-time cavitation detection during HIFU exposure. This work was presented at the 147th meeting of the ASA in New York. It is anticipated that this will form the basis of a feedback control system for inertial cavitation during clinical HIFU application.

Two further research activities took place during my Hunt fellowship year, triggered by the opportunity to build two active cavitation detectors (ACD) pretty much from scratch. The first was used not to detect cavitation, but to determine the effect that the presence of red blood cells have on the acoustic response of stabilized gas bubbles (ultrasound contrast agents—UCAs). This study, conducted in collaboration with Paolo Zanetti, led to the conclusion that high concentrations of red blood cells inhibit the response of UCAs to acoustic excitation. Ultrasound attenuation through UCA suspensions in blood was also found to be dominated by red blood cells at low UCA concentrations, and by the gas bubbles at high UCA concentrations. This work was presented by Paolo Zanetti at the 147th Meeting of the ASA in New York, where he was awarded the Best Student Paper Award in Biomedical Ultrasond/Bioresponse to Vibration. The second ACD was taken to the Department of Biomedical Engineering at the University of Cincinnati where, under the guidance of Professor Christy K. Holland, it was utilized in conjunction with a custom-built PCD to assess the relative role played by inertial and stable cavitation during ultrasound-assisted thrombolysis.

I am extremely grateful to Professor Roy for his selfless mentorship and guidance, and for spending countless hours debating theoretical and experimental aspects of acoustics with me. He repeatedly placed my best interests above his own and was a truly exceptional scientific and career mentor. During my time at Boston University, I also greatly benefited from challenging interactions with Professors Glynn Holt, Robin Cleveland, Paul Barbone, Michael Howe, Allan Pierce, and William Carey, to name but a few. I am also extremely grateful to the exceptional graduate students that I worked with throughout my time in Boston, and in particular to Paolo Zanetti, Caleb Farny, and Charles Thomas: They made invaluable contributions to both the quality and quantity of research carried out during my Hunt Fellowship year. Finally, I would like to thank the Acoustical Society for the most formative and intellectually challenging year of my life to date. Given my research interests, I could not have hoped for a better research environment and better conditions under which to carry out my Hunt fellowship, which ended with my appointment to a joint University Lecturership in Biomedical Engineering and a Magdalen College Tutorial Fellowship at the University of Oxford (UK). This has now become the home of the newly established Biomedical Acoustics Laboratory, in which I hope to perpetuate the excellence in research, acoustics education, and graduate mentoring to which I was exposed at Boston University and through the Acoustical Society.

Report of the 26th F. V. Hunt Postdoctoral Fellow (2003-04)

Tyrone Porter, Biomedical Engineering, University of Cincinnati, Cincinnati, Ohio 45267.

My tenure as the 26th Frederick V. Hunt Fellow was spent under the mentorship of Dr. Christy K. Holland at the University of Cincinnati. Dr. Holland had recently received funding for a joint project with scientists at Northwestern University to develop a Transcranial Ultrasound Therapy System (TUTS) for the treatment of stroke. After injection of targeted liposomes specially formulated to encapsulate air and a clot-busting drug, ultrasound would be used to break the vesicles and release the drug at the site of the clot. Additionally, these echogenic liposomes (ELIP) could be used as targeted ultrasound contrast agents for diagnostic ultrasound imaging of atherosclerotic plaques. The primary goals of my project were: (1) To determine the backscatter and attenuation coefficients for echogenic liposomes as a function of frequency, and (2) to identify the pressure threshold(s) for the destruction of these liposomes and Optison®, a commercially available ultrasound contrast agent.

The backscatter coefficient is a parameter that describes the effectiveness with which a given mass scatters acoustic power. Its magnitude is dependent upon the several factors, including the scattering cross section of the insonated media, and, in the case of ultrasound contrast agents, the viscoelastic properties of the outer shell. Therefore, it is possible to calculate the mean volume of gas trapped inside a single liposome. Additionally, the evaluation of the backscatter and attenuation coefficients as a function of frequency provides information on the resonant behavior of the encapsulated microbubbles. This information can be used to guide the selection of acoustic parameters to optimize contrast enhancement in ultrasound images.

I spent the first part of my fellowship learning the measurement and data reduction techniques developed by several research groups over the past twenty years for determining the backscatter coefficient. After spending considerable time carefully characterizing the acoustic output of several transducers for the study, I collected backscatter data at 3.5, 7.5, 10.0, and 15.0 MHz. Using the experimentally derived backscatter coefficients, I was able to theoretically determine the shell properties of the liposomes with the help of Dr. Constantin-C. Coussios, the 25th recipient of the Frederick V. Hunt Fellowship. This work was presented at the 147th meeting of ASA in New York, and a manuscript is currently being drafted for submission to JASA.

I began work on the destruction thresholds of ELIP at various frequencies toward the end of my fellowship year. Reviewing the literature, I learned that there are several mechanisms for ultrasound contrast agent (UCA) destruction. Since then, I have worked with two graduate students, Denise Smith and Saurabh Datta, on designing data collection and processing techniques for distinguishing between acoustically driven diffusion and fragmentation. Using a laboratory assembled pulse-echo system and a clinical diagnostic ultrasound scanner, Denise and I investigated the behavior of Optisonα at varying incident pressures. Plotting the backscattered intensity and mean grey-scale values from a region of interest in the B-mode images as a function of time, we identified the pressure threshold for acoustically driven diffusion of Optison® at 3.5 MHz. The measurement and processing techniques and initial results were presented at the 149th meeting of ASA in Vancouver, British Columbia, and a manuscript is in preparation for journal submission.

My year as the Hunt Fellow was fulfilling both scientifically and professionally. Based upon my work during that year, I submitted for and was awarded an NIH supplemental grant. I had the opportunity to gain valuable teaching experience by managing a graduate level course in the Department of Biomedical Engineering. Dr. Holland has been an amazing mentor, and we have developed a professional relationship that I am sure will continue throughout my career. I would like to thank Dr. George Shaw, Dr. Jason Chang, Dr. Lou McAdory, and our collaborators at Northwestern University for stimulating conversation on the various chemical, biological, and medical aspects of liposomes, stroke, and cardiovascular disease. Finally, I would like to thank the Hunt family and the Acoustical Society of America for providing funding and the support as I begin my post-graduate career.

The following is a list of presentations resulting in part from my year as Hunt Fellow:

Porter, T. M., Smith, D. A. B., and Holland, C. K. "Quantification of the pressure threshold for Optison® destruction," (Accepted for oral presentation at the American Institute of Ultrasound in Medicine Annual Convention, March 24, 2006).

Porter, T. M., Vaidya, S. S., Holland, C. K., Huang, S. L., MacDonald, R. C., and McPherson, D. D. **(2005)** "Evaluation of backscattered intensity to quantify the destruction rate of echogenic liposomes," 149th Meeting of The Acoustical Society of America joint meeting with the Canadian Acoustical Association, Vancouver, British Columbia, May 16–20.

Porter, T. M., Holland, C. K., Huang, S., McDonald, R. C., and McPherson, D. D. **(2004)** "*In vitro* characterization of echogenic liposomes by acoustic scattering at 3.5–15.0 MHz," 147th Meeting of The Acoustical Society of America, New York, NY, May 24–28.

Report of the 27th F. V. Hunt Postdoctoral Fellow (2004-05)

Xuedong Zhang, Auditory Perception Group, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

Although it is widely accepted that the responses to sounds in the high-frequency, basal, end of the cochlea are highly nonlinear and sharply tuned, our knowledge of how the cochlea responds to sounds in the low-frequency apical region is sparse and highly controversial due to the technical difficulties of measuring directly from the apex of the mammalian cochlea. As the 2004-2005 F. V. Hunt Postodctoral Fellow, I worked with Dr. Andrew J. Oxenham at the Massachusetts Institute of Technology to study human cochlear mechanics in the low-frequency region, by combining behavioral and quantitative modeling techniques. The specific goal of the research is to estimate the human cochlear tuning at low frequencies using both nonsimultaneous (forward) and simultaneous masking techniques, and to explore whether any differences in the estimates can attribute to the nonlinear interactions between the masker and signal within the cochlea.

In the behavioral study, auditory filter shapes were estimated by measuring signal threshold in the presence of a noise masker with a spectral notch centered around the signal. Twelve normal-hearing subjects participated. The results show that forward masking provides sharper auditory filter estimates than simultaneous masking for all low frequencies tested (250, 500, and 1000 Hz). The filter bandwidth ratios of simultaneous and forward masking are between 1.5 and 2 for all three frequencies, which is comparable with the data from previous studies at higher frequencies. This suggests that, contrary to expectations, nonlinear processing in the low-frequency region of the cochlea may be similar to that found in higher-frequency regions.

Computational modeling was used to simulate the psychophysical results for both forward and simultaneous masking, using the same parameters and tracking procedure as in the behavioral studies. Auditory models with different cochlear nonlinearities were tested in the simulations. The simulation results suggest that suppression could be used to explain the difference of auditory filter estimates between forward and simultaneous masking. The results therefore provide strong constraints on models of cochlear processing and should help in the development of functional models that are valid at all audio frequencies.

Results from our work have been presented in two conferences [Zhang, X., and Oxenham, A. J., "Estimates of auditory filter shapes at low frequencies using forward and simultaneous masking," 28th Annual Mid-Winter Research Meeting of ARO, (2005); Zhang, X., and Oxenham, A. J., "Modeling study of the influence of the cochlear nonlinearity on psychophysical tuning," 149th Meeting Acoustical Society of America (2005)]. A paper with more recent findings is in preparation and should be submitted by the end of the year.

During my tenure as a Hunt Fellow, I was able to learn the behavior techniques used in hearing studies and to collaborate with distinguished scientist Dr. Oxenham. I am most grateful to the Hunt family and the Acoustical Society for this opportunity.

Editor's note: Please note that the spelling error of the name "Klepper" in "Acoustic analysis in Mudejar-Gothic churches: Experimental results" by Miguel Galindo, Teófilo ZamarreZo, and Sara Girón [J. Acoust. Soc. Am. **117**(5), 2783–2888 (2005)] on page 2873 in the fourth paragraph of the Introduction section has been corrected in the online version of JASA. The same misspelling has been corrected in the reference number six within that paper.

USA Meetings Calendar

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

2006

5–9 June	151st Meeting of the Acoustical Society of America, Providence Rhode Island [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; Web: http://asa.aip.org].
17–21 Sept.	INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA [www.interspeech2006.org)]
28 Nov.–2 Dec.	152nd Meeting of the Acoustical Society of America joint with theAcoustical Society of Japan, Honolulu, Hawaii [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; Web: 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, Utah [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; Email: asa@aip.org; Web: http://asa.aip.org].
27 Nov.–2 Dec.	154th Meeting of the Acoustical Society of America, New Orleans, Louisiana (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; Email: asa@aip.org; Web: 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 onBiological Effects of Noise). Foxwoods Resort, Mashantucket, CT [Jerry V. Tobias, ICBEN 9, Post Office Box 1609, Groton CT 06340-1609, Tel. 860-572-0680; Web: www.icben.org. Email 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 NEWS—INTERNATIONAL

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

International Meetings Calendar

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

April 2006

3–4 **Futures in Acoustics**, Southampton, UK (Web: www.ioa.org.uk).

24–27 **French Congress on Acoustics**, Tours, France (Web: cfa06.med.univ-tours.fr).

May 2006

2–5 **International Conference on Speech Prosody 2006**, Dresden, Germany (Web: www.ias.et.tu-dresden.de/sp2006).

5–7 **Sixth International Conference on Auditorium Acoustics**, Copenhagen, Denmark (e-mail: t.j.cox@salford.ac.uk; Web: www.ioa.org.uk/viewupcoming.asp).

10–12 *__Meeting of the Italian Acoustical Association__, Ischia (NA), Italy (Web: www.associazioneitalianadiacustica.it).

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

16–19 **IEEE Conference Oceans'06 Asia Pacific**, Singapore (Web: www.oceans06asiapacific.sg).

23–26 **17th Session of the Russian Acoustical Society**, Moscow, Russia (Web: www.akin.ru).

23–26 *__XXIII Symposium on Hydroacoustics__, Krynica Morska, Poland (Web: www.hydro-etl.pg.gda.pl/sha2006).

30–1 **6th European Conference on Noise Control (EURONOISE2006)**, Tampere, Finland (Fax: +358 9 7206 4711; Web: www.euronoise2006.org).

June 2006

4–6 *__8th International Symposium on Transport Noise and Vibration__, St. Petersburg, Russia (Web: webcenter.ru/~eeaa/tn06).

12–15 **8th European Conference on Underwater Acoustics**, Carvoeiro, Portugal (Web: www.ecua2006.org).

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

26–28 **9th Western Pacific Acoustics Conference (WESPAC9)**, Seoul, Korea (Web: wespac9.org).

26–29 **11th International Conference on Speech and Computer**, St. Petersburg, Russia (Web: www.specom.nw.ru).

July 2006

2–6 *__13th International Congress on Sound and Vibration (ICSV13)__, Vienna, Austria (Web: info.tuwienac.at/icsv13).

17–20 **International Symposium for the Advancement of Boundary Layer Remote Sensing (ISARS13)**, Garmisch-Partenkirchen, Germany (Fax: +49 8821 73 573; Web: imk-ifu.fzk.de/isars).

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

August 2006

22–26 *__9th International Conference on Music Perception and Cognition__, Bologna, Italy (Web: www.icmpc2006.org).

September 2006

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

18–19 *__Greek National Conference on Acoustics__, Heraklion, Crete, Greece (Web: www.iacm.forth.gr/~acoustics2006).

18–20 **Sixth International Symposium on Active Noise and Vibration Control (ACTIVE2006)**, Adelaide, Australia (Web: www.active2006.com).

18–20 **International Conference on Noise and Vibration Engineering (ISMA2006)**, Leuven, Belgium (Fax: 32 16 32 29 87; Web: www.isma-isaac.be).

18–20 **12th International Conference on Low Frequency Noise and Vibration and its Control**, Bristol, UK (Web: www.lowfrequency2006.org).

October 2006

3–6 *__IEEE International Ultrasonics Symposium__, Vancouver, BC, Canada (Web: www.ieee-uffc.org).

11–13 *__Annual Conference of the Canadian Acoustical Association__, Halifax, Nova Scotia, Canada (Web: www.caa-aca.ca/halifax-2006.html).

18–20 *__37th Spanish Congress on Acoustics–EAA Symposium of Hydroacoustics–Iberian Meeting on Acoustics__, Gandia-Valencia, Spain (Web: www.ia.csic.es/sea/index.html).

25–28 **Fifth Iberoamerican Congress on Acoustics**, Santiago, Chile (Web: www.fia2006.cl).

November 2006

2–3 **Swiss Acoustical Society Fall Meeting**, Luzern, Switzerland (Web: www.sga-ssa.ch).

20–22 **1st Joint Australian and New Zealand Acoustical Societies Conference**, Christchurch, New Zealand (Web: www.acoustics.org.nz).

April 2007

10–12 **4th International Conference on Bio-Acoustics**, Loughboro, UK (Web: www.ioa.org.uk).

July 2007

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

August 2007

26–29 **Inter-noise 2007**, Istanbul, Turkey (Web: www.internoise2007.org.tr).

27–31 **Interspeech 2007**, Antwerp, Belgium (Web: www.interspeech 2007.org).

September 2007

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

9–12 **ICA Satellite Symposium on Musical Acoustics (ISMA2007)**, Barcelona, Spain (SEA, Serano 144, 28006 Madrid, Spain; Web: www.isma2007.org).

9–12 **ICA Satellite Symposium on Room Acoustics (ISRA2007)**, Sevilla, Spain (Web: www.isra2007.org).

June 2008

29–4 **Forum Acusticum: Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**, Paris, France (Web: www.sfa.asso.fr; e-mail: phillipe.blanc-benon@ec-lyon.fr).

 0001-4966/2006/119(4)/1901/2/$22.50

July 2008

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

September 2008

22–26 *__INTERSPEECH 2008–10th ICSLP__, Brisbane, Australia (Web: www.interspeech2008.org).

August 2010

23–27 **20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (Web: www.acoustics.asn.au).

Euronoise Conference and Transport Noise Symposium Cooperation

The symposium "Transport Noise and Vibration 2006," to be held in St. Petersburg, Russia, 4–6 June, will be linked with the European Conference "Euronoise 2006" in Tampere, Finland, 30 May to 1 June. Participants who attend both events will be offered a reduced registration fee. Also, transfer from Tampere to Saint Petersburg will be provided. Additional information is available by contacting the organizers of the two conferences at their URL: euronoise2006.org and webcenter.ru/~eeaa/tn06/

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).]*

Sound images of the ocean: In research and monitoring

Peter Wille

472 pp Springer, Berlin, 2005. $129.00. ISBN 3540241221 (hardcover).

This book, quoting the author: "is the first attempt to publish a comprehensive overview of the wide variety of acoustic applications in the fields of research, of utilization, surveillance and protection of the ocean". His goal is achieved admirably with an extraordinary array of images obtained through the use of a number of acoustic devices with varying degrees of resolution. In addition, the text is accompanied by a CD-ROM that allows the reader to manipulate images; zooming in and out, "flying" over landscapes, etc.

The text is divided into two major parts: The basic facts of imaging the ocean and applications of acoustic imaging the ocean. The first part, in turn, has four chapters which contain some useful history of underwater exploration, and also the history of sonar development. It also introduces the acoustic properties of the ocean and the acoustic environment that must be considered when attempting to create acoustic images. This part concludes with detailed descriptions of the types of sonar employed which include multibeam, sidescan, sediment penetrating echo sounder and acoustic Doppler current profiler.

The second part is first divided into three major sections: Natural formations of the sea floor, the ocean volume, and man made matter. Returning to the seafloor, this section further subdivided geographically, considering each major ocean and sea; by feature scale, e.g., tectonic structure, seamounts, escarpment slides, sediment bedforms, etc. The ocean volume is subdivided into water mass transport, detection of gas and submarine fauna, primarily fish stock monitoring. Man made matter considers surveillance and maintenance of coastal sea lanes, submarine construction (e.g., cables and pipelines), archaeology, including wrecks due to wartime events and close-up images made possible with the advent of remotely operated vehicles (ROVs) and autonomous undersea vehicles(AUVs).

For each specific circumstance discussed, an excellent example of an image obtained by a particular sonar is provided, and in many cases, more than a single image, often with a different sonar. Accompanying each image is a detailed explanation of the physical circumstances, which results in rapid understanding of the subject matter. Who will use this book and what is the benefit of ownership? To quote the author a second time; "The book is intentionally written for the non-expert…", but it is intended to reach readers who can technically appreciate its content and develop an understanding of the geophysical complexity of the earth, that cannot be obtained with simple photography.

For this reviewer, it will also be used to introduce students in underwater acoustics to the reality of the environment they have chosen to study.

DAVID BRADLEY
Pennsylvania State University,
Applied Research Laboratory,
State College, Pennsylvania 16804-0030

 0001-4966/2006/119(4)/1903/1/$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*
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*

6,954,175

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 colinearly with a pair of GPS 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

6,954,024

43.30.Yj UNIDIRECTIONAL ACOUSTIC PROBE AND METHOD FOR MAKING SAME

Ngoc-Tuan Nguyen *et al.*, assignors to Thales
11 October 2005 (Class 310/334); filed 9 March 2001

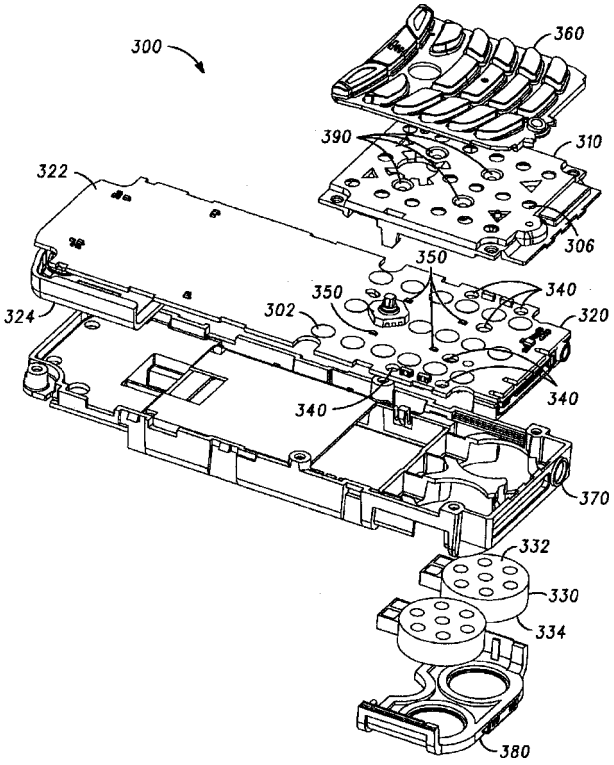
This patent describes a segmented ultrasonic transducer array for NDE testing and underwater use. The concept is nothing new, but the interconnects are made on a flexible substrate, making for a neat assembly and precise alignment.—JAH

6,952,350

43.38.Ar PORTABLE ELECTRONIC DEVICE INCLUDING AN ACOUSTIC CHAMBER

Benjamin M. Finney *et al.*, assignors to Motorola, Incorporated
4 October 2005 (Class 361/748); filed 7 March 2003

This is a detailed look at the acoustical considerations in a cellular



telephone. The addition of aperatures **340** to the frame creates a cavity between the transducers **332** and the top of the frame.—MK

 0001-4966/2006/119(4)/1905/18/$22.50

6,952,965

43.38.Ar VERTICAL MEMS GYROSCOPE BY HORIZONTAL DRIVING

Seok-jin Kang *et al.*, assignors to Samsung Electronics Company, Limited
11 October 2005 (Class 73/504.12); filed in the Republic of Korea 24 December 2002

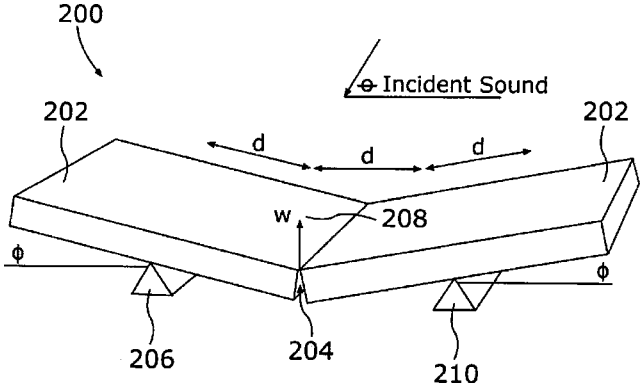
As the art of micromachining gets more mature, it begins to go out of plane. This patent covers a gyroscope that utilizes two silicon wafers so as to achieve sizable out-of-plane motions of a gyroscope mass. This device is apparently unique in that the planes of actuation and detection are not the same, something that is not an advantage in performance, but is an advantage in fabrication.—JAH

6,963,653

43.38.Ar HIGH-ORDER DIRECTIONAL MICROPHONE DIAPHRAGM

Ronald Miles, assignor to The Research Foundation of the State University of New York
8 November 2005 (Class 381/424); filed 22 October 2003

A conventional, first-order microphone achieves its directionality through sensing a pressure gradient between the front and back of the diaphragm. As Olson states, a second-order microphone is one that senses ''the gradient of the gradient'' and thus exhibits even more directionality. Normally, this requires two spaced diaphragms and there have been very few



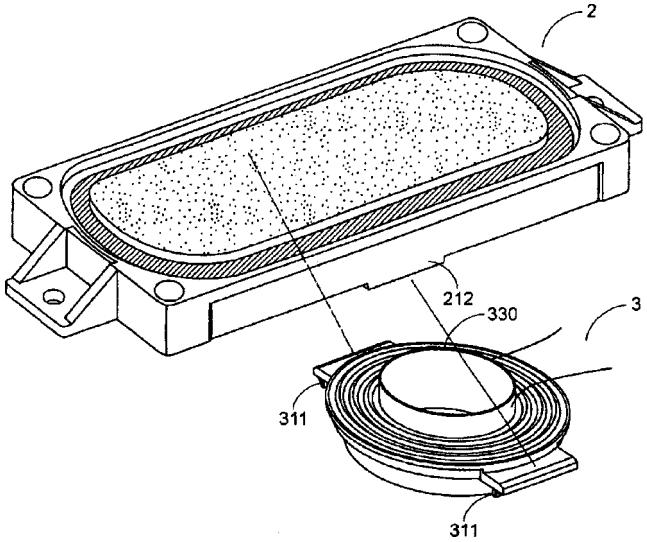
designs in which a single diaphragm structure has been able to accomplish this. The second-order version of this micro-electromechanical-system (MEMS) microphone uses a reticulated diaphragm. Elements **202** move about their individual rocking points and their mutual hinge point. The respective gradient distances are indicated as **d**, and the axis of maximum sensitivity is left-to-right, as seen in the figure.—JME

6,934,399

43.38.Dv PISTON-TYPE PANEL-FORM LOUDSPEAKER

Tai-Yan Kam, Hsin Chu, Taiwan, Province of China
23 August 2005 (Class 381/152); filed in Taiwan, Province of China 19 May 2003

Despite some fanciful ideas about how an electrodynamic motor works and what the spider in such an assembly actually does and how it is typically



implemented, the patent describes an oval, flat-panel diaphragm that may indeed reduce the cost of materials and speed assembly of the unit.—NAS

6,954,121

43.38.Fx METHOD FOR CONTROLLING PIEZOELECTRIC COUPLING COEFFICIENT IN FILM BULK ACOUSTIC RESONATORS AND APPARATUS EMBODYING THE METHOD

Paul D. Bradley *et al.*, assignors to Agilent Technologies, Incorporated
11 October 2005 (Class 333/187); filed 9 June 2003

This patent describes film bulk acoustic resonators for the GHz frequency range fabricated in Si. It is argued that the use of layers of metal and dielectric film in the resonators makes it possible to alter the effective coupling constant of the structure, thereby changing the filter characteristics of the resonator. This does not involve any new discoveries, although the application of the idea to the μm scale may be new.—JAH

6,955,787

43.38.Fx INTEGRATED BIOLOGICAL AND CHEMICAL SENSORS

William Paynter Hanson, Carlisle, Pennsylvania
18 October 2005 (Class 422/50); filed 11 October 2003

This patent describes the use of arrays of piezo sensors to measure chemical and biological processes by sensing bonding and chemisorption of the analytes to the transducer surfaces. The authors show sketches of several different arrays, most of them using a differential sensing approach in which pairs of sensors are placed on the substrate with one being exposed to the analyte and the other not exposed, so as to counter drift problems. It is not clear what is novel about this.—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

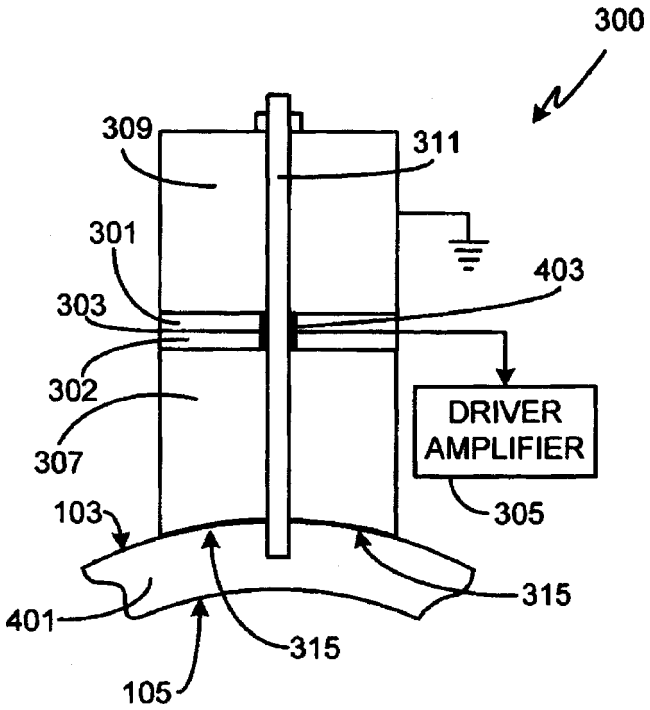
The solid shell of either a flextensional transducer or a slotted shell cylindrical transducer is replaced by a honeycomb shell. The honeycomb shell provides high specific stiffness with much less weight than a solid shell, thus increasing the transducer's bandwidth, because the honeycomb shell can be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

6,960,869

43.38.Fx ACOUSTIC DRIVER ASSEMBLY FOR A SPHERICAL CAVITATION CHAMBER

Ross Alan Tessien *et al.*, assignors to Impulse Devices, Incorporated
1 November 2005 (Class 310/325); filed 17 September 2004

An acoustic driver assembly for exciting a spherical sonoluminescence chamber is described. It consists of a basic longitudinal vibrator **300** driven by a pair of oppositely polarized piezoceramic discs **301** and **302**, tail mass **309**, head mass **307**, and stress bolt **311**. In various embodiments, the head mass can be curved to match the curvature **315** of the chamber wall, or can have a smaller radius of curvature so that the head mass only touches the chamber wall in a circle, or be flat, or have an inverted curvature so that there is minimal contact between the two surfaces. The transducer element is attached to the external wall **401** of the chamber in a variety of ways. The figure illustrates the case where the threaded stress bolt **311** also engages a tapped hole in the wall. Alternatively the head mass may be brazed, diffusion bonded, or epoxy bonded to the wall.—WT

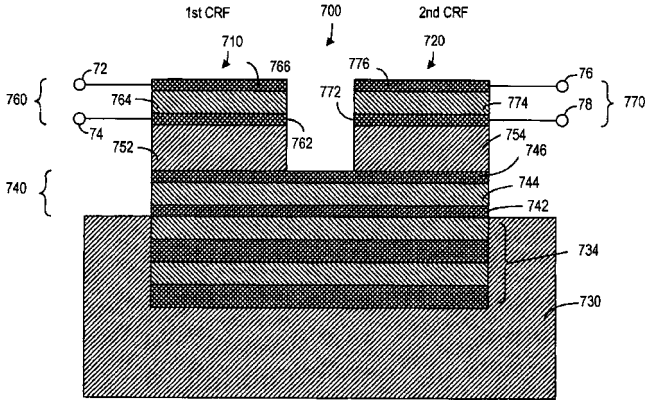


6,963,257

43.38.Fx COUPLED BAW RESONATOR BASED DUPLEXERS

Juha Ellä and Robert Aigner, assignors to Nokia Corporation
8 November 2005 (Class 333/133); filed 19 March 2004

This patent describes the use of an acoustically coupled pair of FBAR filters to act as a diplexer, viz., a device that allows you to talk both ways through the single input port. This filter operates at cellular-phone



frequencies and is designed to sit between the antenna and the electronics of the phone. Apparently, the novel features of this are just the disposition of the two FBARs in a sort of U shape, as shown in the diagram.—JAH

6,941,246

43.38.Hz METHOD FOR THREE-DIMENSIONAL POSITION CALIBRATION OF AUDIO SENSORS AND ACTUATORS ON A DISTRIBUTED COMPUTING PLATFORM

Vikas C. Raykar *et al.*, assignors to Intel Corporation
6 September 2005 (Class 702/186); filed 18 September 2003

Certain sound-producing events, particularly conferences and the like, are advantageously recorded using an array of microphones. However, setting up such an array can be an expensive proposition involving special high-speed, high-capacity audio equipment. This patent proposes an alternative based on a layout of laptop computers, each having an audio system with sound card, microphone, and loudspeaker. It is assumed that at least some of the required processing to coordinate such a layout could be performed by the individual laptops. An additional coordinating computer would almost certainly also be required. During a calibration phase, each machine could, in turn, emit a beep, which would be picked up by all other systems, allowing the computation of the location of each machine based on sound travel times. In fact, the patent describes little else other than this calibration scheme. Many other obvious requirements are glossed over or omitted entirely.—DLR

6,954,025

43.38.Hz RESONANT ENERGY MEMS ARRAY AND SYSTEM INCLUDING DYNAMICALLY MODIFIABLE POWER PROCESSOR

Toshikazu Nishida *et al.*, assignors to University of Florida Research Foundation, Incorporated
11 October 2005 (Class 310/339); filed 13 May 2003

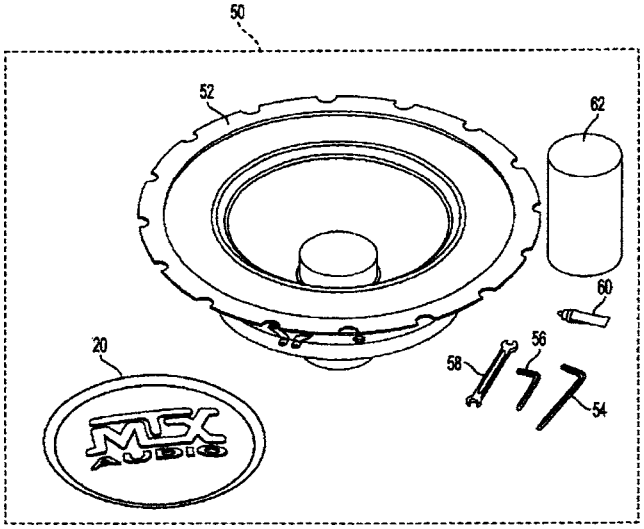
This patent describes a variant on the piezoelectric energy harvesting idea in which the piezo devices are split up and distributed across a plane substrate. This substrate also includes a controller IC that selects among generators to hook up to the output at any moment. One would guess that this division of mass makes the generators less efficient for most everyday uses, but that question is not addressed here, nor is the question of the energy consumption of the control electronics.—JAH

6,938,726

43.38.Ja FIELD REBUILDABLE LOW FREQUENCY LOUDSPEAKER DRIVER HAVING A REMOVABLE DIAPHRAGM ASSEMBLY, PARTS KIT AND METHOD FOR REBUILDING A LOUDSPEAKER DRIVER IN THE FIELD

Richard S. Roark and Daniel F. Roemer, assignors to Mitek Corporation
6 September 2005 (Class 181/171); filed 16 October 2002

It is an unavoidable fact of life that electroacoustic transducers can fail. The question then is how to repair them in an efficient, repeatable, and cost-effective manner. Peavey introduced their Black Widow line of low-frequency loudspeakers a while ago, in which the whole bracket assembly (frame, spider, surround, coil, leads, contact terminals) is removable, thereby making field replacement of a damaged unit possible and fairly easy (if you have the replacement part on hand) and where no application of adhesive is necessary. Just unbolt, remove, replace, rebolt, and go on. The present patent describes a version of the field replacement of the expended part that requires a bit more work, some skill, and more time, but probably with less cost for the replacement part. In the present patent, the replacement kit **50** consists of cone/surround/coil/spider/lead assembly **52**, a guide **62**, dust cap **20**, tools **52**, **56**, and **58**, and a tube of adhesive **60**. The repair procedure is presented in detail. Sounds easy.—NAS

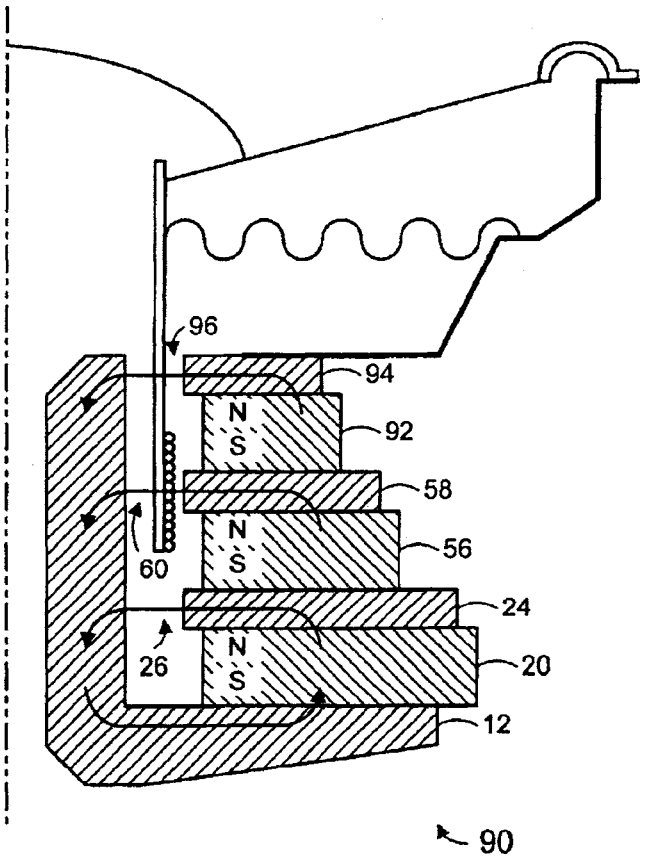


6,940,992

43.38.Ja PUSH-PUSH MULTIPLE MAGNETIC AIR GAP TRANSDUCER

Enrique M. Stiles, assignor to Step Technologies Incorporated
6 September 2005 (Class 381/412); filed 5 November 2002

Various types of multiple-gap, multiple-magnet, and multiple-coil electrodynamic motor geometries have been developed and patented. One well-known implementation is Button's dual-gap, dual-coil topology. The author for this patent is known for his "push-pull magnetic air gap topologies." The present invention describes several two-magnet, two-gap, one-coil topologies, various implementations of three-magnet, three-gap, one-coil topologies, and a version of a two-magnet, single-gap, single-coil topology.—NAS

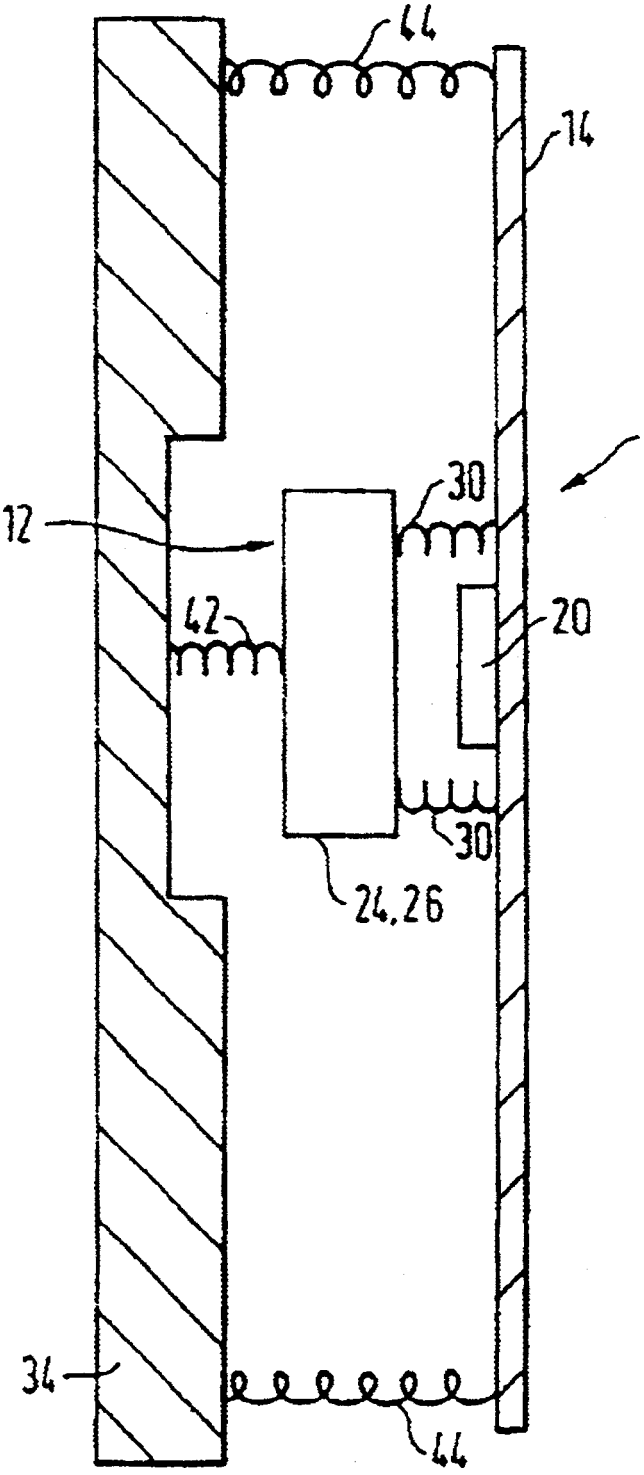


6,956,957

43.38.Ja LOUDSPEAKERS

Henry Azima *et al.*, assignors to New Transducers Limited
18 October 2005 (Class 381/431); filed in the United Kingdom 9 January 1997

A panel-type loudspeaker driven by a freely suspended inertia transducer requires substantial mass and is consequently prone to shipping damage. If the driving transducer is rigidly attached to an overall mounting



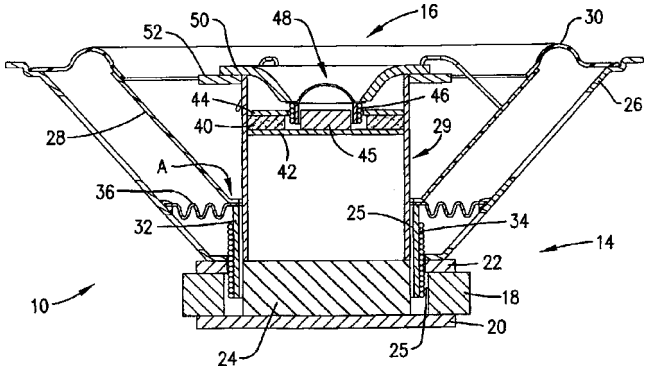
frame, then vibrations are directly transmitted to the frame and performance may suffer. This patent describes a fairly rugged resilient mounting which can also serve as an element to ''tune'' the response of the system.—GLA

6,963,650

43.38.Ja COAXIAL SPEAKER WITH STEP-DOWN LEDGE TO ELIMINATE SOUND WAVE DISTORTIONS AND TIME DELAY

Christopher Combest, assignor to Multi Service Corporation
8 November 2005 (Class 381/182); filed 9 September 2002

Exactly how does the coaxial loudspeaker shown differ from countless examples of prior art? The answer is step-down ledge **52**, ''for reflecting



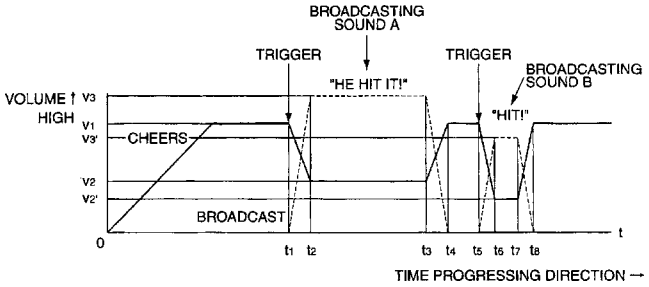
sound waves diffracting around the edge of the baffle.'' Loudspeaker designers may find this a questionable step forward in the quest for audio perfection.—GLA

6,945,871

43.38.Lc RECORDING MEDIUM STORING VOLUME CONTROL PROGRAM, VOLUME CONTROL METHOD, VIDEO GAME MACHINE, AND VOLUME CONTROL PROGRAM

Takahito Uenishi and Yasuhiro Kawabata, assignors to Konami Computer Entertainment Osaka, Incorporated
20 September 2005 (Class 463/35); filed in Japan 16 July 2001

A video game typically has several things going on at once and each activity often generates its own audio signal. An obvious control arrangement is to provide a volume control stage for each sound source at a point before it is summed into the game audio output signal. This patent presents



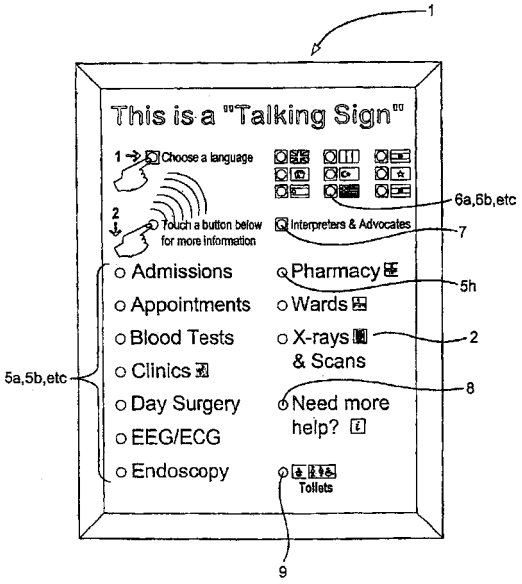
a strategy for setting such volume controls under various game states. The example given is a sports event with a crowd's cheering noise as one audio source and a play-by-play announcer as a second source. The strategy is to lower the cheering noise when the announcer speaks, much as an engineer at a mixer panel might do.—DLR

6,956,492

43.38.Lc DISPLAY SIGNS COMPRISING A FLAT PANEL LOUDSPEAKER

Michael Andrew Beadman *et al.*, assignors to Talksign Limited
18 October 2005 (Class 340/686.1); filed in the United Kingdom 18 November 2000

The inventors envision an interactive display panel that contains a number of proximity switches. The display also functions as a panel-type loudspeaker, providing audible response to switch actuation. Provisions for



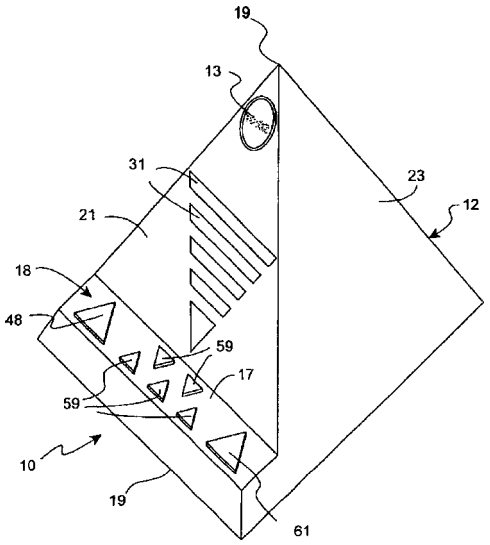
memory storage and reprogramming are described in the patent but nothing is said about the possibility of using a conventional loudspeaker or of providing visual (touchscreen) response.—GLA

6,947,352

43.38.Md AUDIO RECORDING PLAYBACK DEVICE

Aaron J. Goolkasian, Charlotte, North Carolina
20 September 2005 (Class 369/27.01); filed 15 July 2002

Pyramid power is back! "The pyramid shape can be said to have a culturally connotative significance." No doubt—and recording playback will



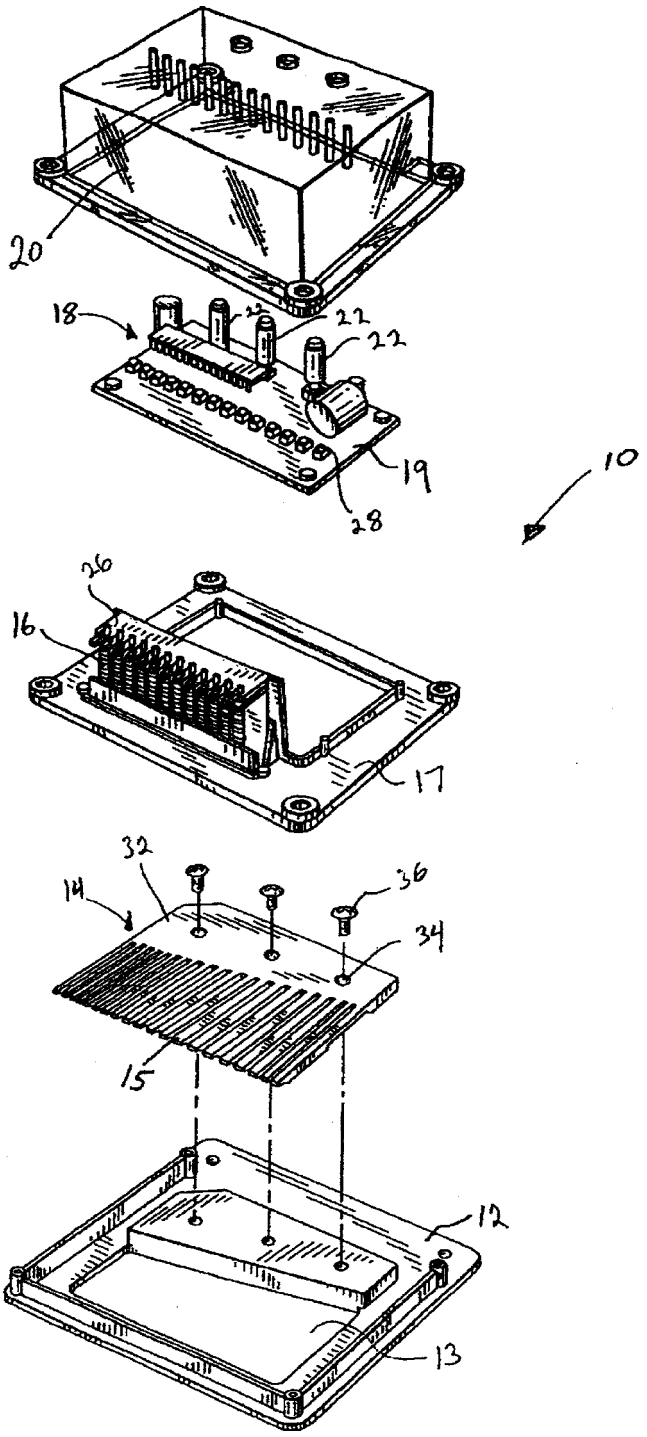
sound better, too. (Why is this a utility patent instead of a design patent?)—MK

6,960,710

43.38.Ne ELECTRONIC ACOUSTIC MUSIC ENGINE

Marc Howard Segan *et al.*, assignors to M.H. Segan Limited Partnership
1 November 2005 (Class 84/94.1); filed 2 April 2004

A conventional music box includes a rotating cylinder with spikes and a set of pretuned tines. One major problem is the limited playing time (limited by the circumference of the cylinder). The inventors solve this



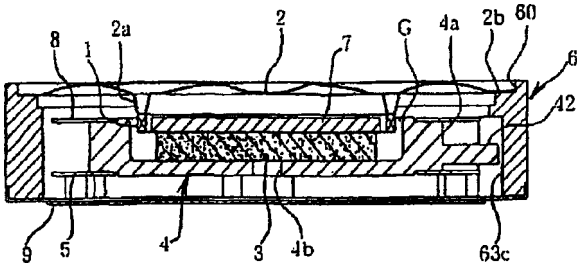
problem by replacing the spikes with a set of computer-controlled (solenoid) actuators. The magnetic force pulls the tine and then releases it, resulting in a brief vibration. Detailed schematics, but not the firmware, are available in the patent.—MK

6,954,016

43.38.Si VIBRATING ACTUATOR AND A POWER SUPPLY MECHANISM THEREOF

Minoru Ueda *et al.*, assignors to Namiki Seimitsu Hoseki Kabushiki Kaisha
11 October 2005 (Class 310/81); filed in Japan 16 April 1999

According to this patent, most cellular phones employ moving coil transducers as ringers. These produce tactile vibrations when driven at low frequencies and audible tones when driven at higher frequencies. We are told that such transducers can be damaged by impact and the flexible wires



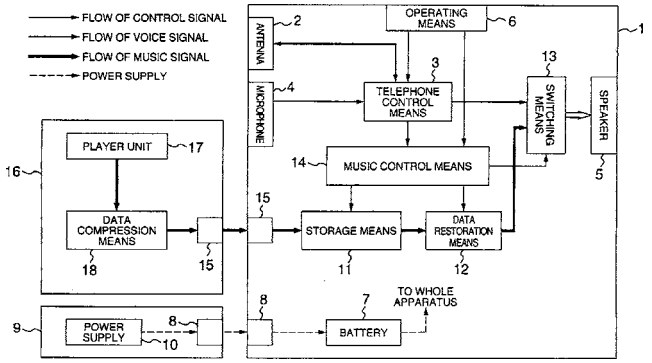
connecting the coil to the circuit board can be broken. An improved design described at great length in the patent actually follows the same basic geometry as prior art. The sole novel feature appears to be the use of conductive leaf springs to make signal connections to the voice coil.—GLA

6,954,652

43.38.Si PORTABLE TELEPHONE APPARATUS AND AUDIO APPARATUS

Kenji Sakanashi, assignor to Matsushita Electric Industrial Company, Limited
11 October 2005 (Class 455/550.1); filed in Japan 13 April 1999

The current hot item in personal electronics is a combination cell phone and music player. To design such a device one might begin by listing the components required to make a cell phone, followed by a list of additional components required to make a personal music player. Preparing such a list would be wasted effort, however, because Matsushita has already patented it. In all fairness, the patent also discusses a number of operational refinements and describes seven possible embodiments.—GLA

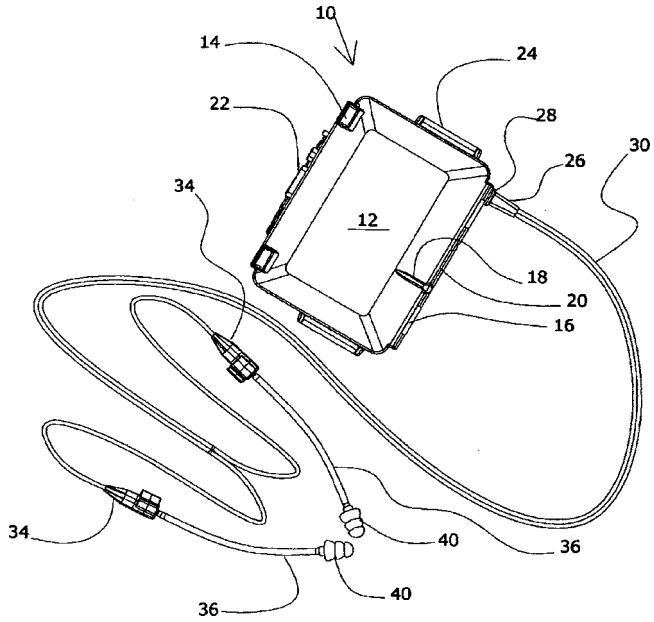


6,961,434

43.38.Si SUBMERSIBLE HEADPHONES

Martin S. Silverman, assignor to DeNovo Research, LLC
1 November 2005 (Class 381/74); filed 26 June 2002

Assuming that a significant number of people want to enjoy music while scuba diving, the inventor has already designed and patented waterproof cassette players and headphones. Even so, if the headphones are dislodged at a depth of several feet, water can enter and corrode the transducer diaphragm. In this improved design, small transducers **34** transmit sound



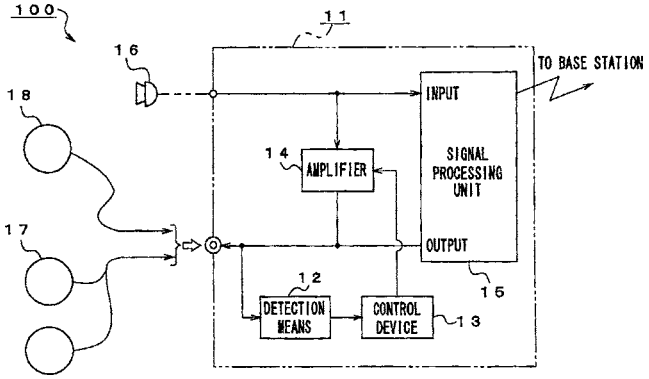
waves to ear plugs **40** through small-bore tubes **36**. The coupling chamber between the transducer diaphragm and the tube is made as small as possible, and the tube is made long enough to contain a volume of air greater than that of the coupling chamber. Although water pressure may be enough to substantially compress the air in the tube, the water itself will not reach the transducer diaphragm.—GLA

6,961,591

43.38.Si HAND HELD TELEPHONE SET AND AUDIO PROCESSING METHOD

Keiji Osano, assignor to Sony Corporation
1 November 2005 (Class 455/569.1); filed in Japan 22 September 2000

A conventional cellular telephone can sense the connection of an external earphone/microphone and disable the corresponding internal transducers. But suppose the unit includes music playback and the user has plugged in a stereo headset instead of a monophonic earphone/microphone. What



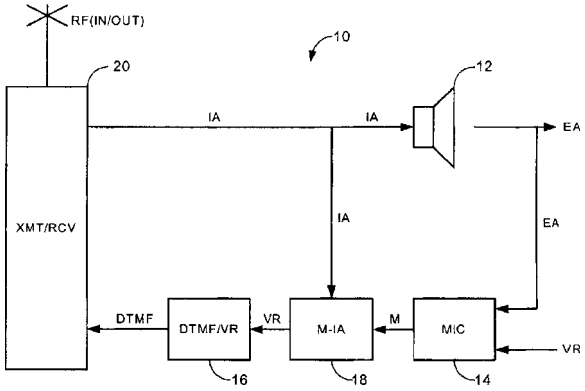
happens when a phone call is received? This patent argues that the level of ambient sound picked up by the microphone and fed back to the headphones should be automatically increased when stereo headphones are in use.—GLA

6,963,760

43.38.Si METHOD AND APPARATUS FOR GENERATING DTMF TONES USING VOICE-RECOGNITION COMMANDS DURING HANDS-FREE COMMUNICATION IN A VEHICLE

James J. Piwowarski, assignor to General Motors Corporation
8 November 2005 (Class 455/563); filed 1 October 2001

While driving a vehicle, it would be desirable to control a wireless communication device entirely through voice commands. In this situation, electronic voice recognition is hampered by the presence of background noise as well as incoming audio signals. What is needed is a ''means for enunciating an audio signal to the cabin of the vehicle, and for simultaneously receiving the enunciated audio signal and a voice-command from the user speaking in the cabin of the vehicle; and means for generating a DTMF signal in response to the voice-command without muting or otherwise interrupting enunciation of the audio signal in the vehicle cabin.'' The preceding quotation is taken directly from the first patent claim. As such, it does not describe the goal of the invention, but rather what has actually been patented. To a layman, this is like patenting the idea of a self-guided automobile or a better tasting peanut butter sandwich.—GLA

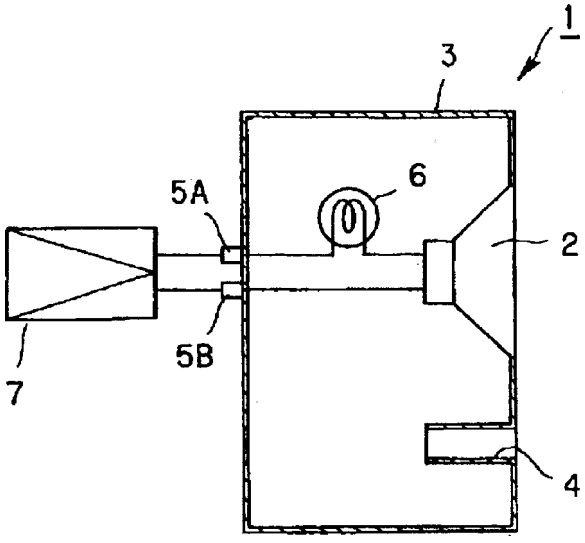


6,928,177

43.38.Tj SPEAKER-USE PROTECTION ELEMENT AND SPEAKER DEVICE

Ikuo Chatani and Yukio Yasuda, assignors to Sony Corporation
9 August 2005 (Class 381/189); filed in Japan 21 February 2001

The patent appears to involve a sheath around lamp **6** that prevents the light emitted from the lamp, when the lamp is in protect mode and the current exceeds a certain limit, from being seen through any openings, such as ports, in an enclosure in which the protection device is used. Several embodiments of the sheath and of the lamp leads are described.—NAS

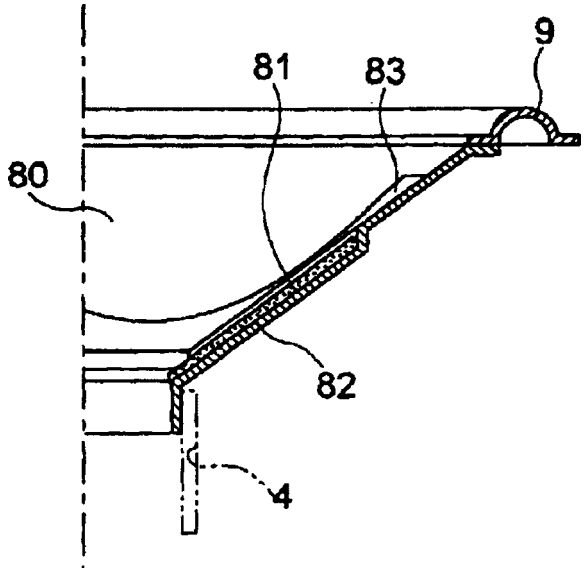


6,929,092

43.38.Tj SPEAKER DIAPHRAGM

Yasuhisa Abe *et al.*, assignors to Pioneer Corporation
16 August 2005 (Class 181/168); filed in Japan 23 October 2000

Speaker diaphragm **80** is manufactured using an injection molding process to include a metallic plate **81** that can be made from aluminum or an aluminum alloy. The plate, about 1 mm in thickness, improves the thermal ''radiation'' efficiency of a subbass electrodynamic speaker assembly for use in automobiles.—NAS

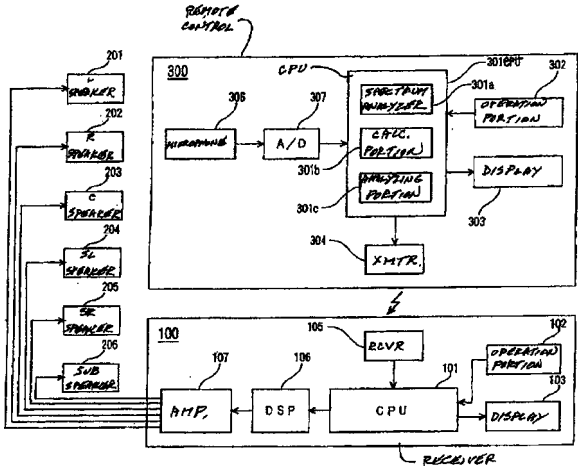


6,954,538

43.38.Vk REMOTE CONTROL APPARATUS AND A RECEIVER AND AN AUDIO SYSTEM

Tadashi Shiraishi, assignor to Koninklijke Philips Electronics N.V.
11 October 2005 (Class 381/105); filed in Japan 8 June 2000

Automatic equalization of audio systems is not a new idea, however the complicated adjustments required to set up a 5.1 home theater make such an approach especially attractive. The system described in this patent includes one or more sensing microphones, plus analysis and control circuitry, all packaged in the remote control unit. During the setup procedure, the



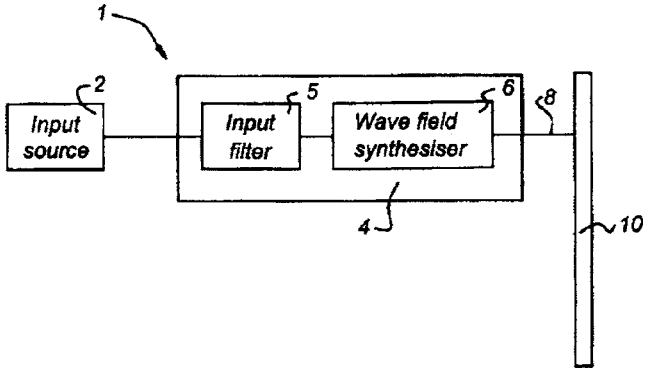
remote control decides what corrections are required and transmits appropriate control signals to the main receiver. One might speculate that the remote control could thus function as a self-contained audio analyzer, but the patent makes no such suggestion, nor does it give any other reason for the arrangement.—GLA

6,959,096

43.38.Vk SOUND REPRODUCTION SYSTEM

Marinus Marias Boone *et al.*, assignors to Technische Universiteit Delft
25 October 2005 (Class 381/152); filed in the European Patent Office 22 November 2000

It is theoretically possible to accurately reproduce a three-dimensional sound field within a confined space. However, practical implementations of this wave-synthesis approach usually ignore vertical directivity and surround the listening area with horizontal arrays of tightly packed identical



loudspeakers. This densely written (almost unreadable) patent suggests that a number of practical advantages can be realized by replacing such arrays with multiple-driver, panel-type loudspeakers. Prefiltering is used to compensate for (at least some of) the drawbacks of this arrangement.—GLA

6,961,439

43.38.Vk METHOD AND APPARATUS FOR PRODUCING SPATIALIZED AUDIO SIGNALS

James A. Ballas, assignor to The United States of America as represented by the Secretary of the Navy
1 November 2005 (Class 381/309); filed 26 September 2001

Static head-related transfer functions (HRTFs) are of limited utility during realistic spatialization because the listener changes their head position, thereby changing the HRTF. This patent proposes using the orientation of the listener (as detected by a helmet) to control the HRTF of the listener. This concept has been discussed for at least 15 years prior to the filing in 2001. In this reviewer's opinion, this patent is DOA.—MK

6,943,969

43.40.Kd METHOD OF DETECTING PROTRUSION ON RECORDING MEDIUM AND DETECTING APPARATUS THEREFOR

Toru Yokohata, assignor to Fujitsu Limited
13 September 2005 (Class 360/25); filed in Japan 20 September 2002

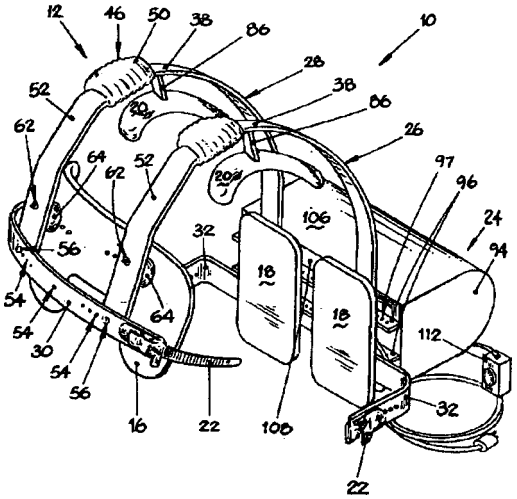
By adding a piezoelectric transducer onto a disk-drive head, the sound output can be used to sense collisions of the head with the media. Regular bumps could be used to provide index marks to denote segment boundaries or impacts.—MK

6,958,047

43.40.Ng CHEST VIBRATING DEVICE

Marten Jon DeVlieger, Taber, Alberta, Canada
25 October 2005 (Class 601/71); filed 2 October 2002

The purpose of this chest-vibrating device is to aid people afflicted by cystic fibrosis or other lung conditions by clearing obstructed airways and lungs. The device includes a frame, shoulder pads, chest pad, and back pad.



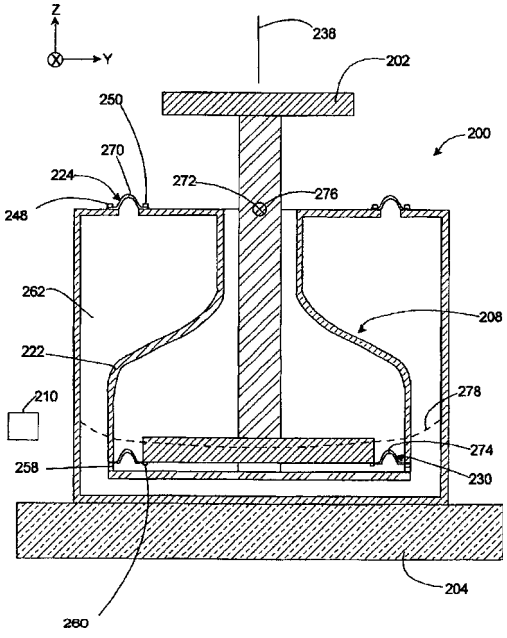
The chest pad is configured to fit around the upper body of the user. A vibrating unit attached to the frame produces a vibration that travels through the frame onto the chest pad and shoulder pad.—DRR

6,953,109

43.40.Tm VIBRATION ISOLATOR WITH LOW LATERAL STIFFNESS

Douglas C. Watson and Alton H. Phillips, assignors to Nikon Corporation
11 October 2005 (Class 188/378); filed 8 October 2002

This isolator, intended for protecting sensitive equipment such as is used in microelectronics manufacture, incorporates a pressurized air volume **262** contained between an essentially cylindrical housing and a sleeve **222**. A resilient seal **270** permits the sleeve to move relative to the housing. Within the sleeve there is mounted a pendulum arrangement whose



disc-shaped bottom is supported from the sleeve wall via resilient elements **274**. Appropriate selection of the geometry results in rotation of the pendulum arrangement about axes **272**. Vertical isolation is due primarily to the air-spring effect between the sleeve and the housing, whereas lateral isolation is due primarily to rocking of the pendulum.—EEU

6,953,424

43.40.Tm ROTOR DRIVING APPARATUS WITH TEMPERATURE ADJUSTMENT OF ELASTIC SUPPORTING PORTION

Shoji Kusumoto and Hiroyuki Takahashi, assignors to Hitachi Koki Company, Limited
11 October 2005 (Class 494/10); filed in Japan 31 July 2002

The motor of a centrifugal separator is attached to its base via a rubber isolator. A Peltier element that heats or cools the isolator in response to a signal from a controller, activated by a temperature sensor, keeps the isolator temperature in the range within which the isolator performs well.—EEU

6,954,686

43.40.Tm METHOD OF TUNING A VIBRATION ABSORBER

Pierre-Antoine Aubourg and Gianni Naccarato, assignors to Eurocopter
11 October 2005 (Class 701/3); filed in France 17 November 2003

A vibration absorber, intended for use on a helicopter, consists of a mass mounted near the tip of a cantilever leaf spring. The arrangement is configured so that the mass can be moved along the length of the leaf spring, thus changing the spring's effective stiffness and thereby the absorber's natural frequency. An accelerometer is provided at the base of the spring and a second one on the mass. Flight test measurements are used to compare the absorber's measured transfer function to its theoretical transfer function. Minimization of the difference between these two functions is employed as the basis for adjusting the absorber without the need for extensive testing and repeated adjustments.—EEU

6,955,250

43.40.Tm APPARATUS FOR DAMPING VIBRATION USING MACRO PARTICULATES

Thomas R. Kreider and Steve L. Hadden, assignors to Honeywell International Incorporated
18 October 2005 (Class 188/268); filed 3 December 2003

A damping arrangement suitable for use at extreme temperatures consists of a container filled with macroparticles, such as graphite, alumina, titanium dioxide, and titanium. These materials may be crushed, pulverized, sputtered, or processed in any way that yields extremely small particles. These particles preferably are of essentially spherical shapes, so that they can readily slide past each other, enabling the contents of the container to slosh back and forth in the presence of vibrations, much as a liquid would. Some larger particles are mixed in with the small ones in order to ''stir'' the small ones and keep them from agglomerating.—EEU

6,958,567

43.40.Tm ACTIVE/PASSIVE DISTRIBUTED ABSORBER FOR VIBRATION AND SOUND RADIATION CONTROL

Christopher R. Fuller and Pierre E. Cambou, assignors to Virginia Tech Intellectual Properties, Incorporated
25 October 2005 (Class 310/321); filed 5 May 2003

A nonmicromechanical method is presented that combines passive modal damping with active piezoelectric damping. This patent covers the use of hybrid passive/active damping layers on sheets and beams of material for the purposes of structural-vibration and sound-emission damping. The active layer is preferably PVDF and is corrugated to allow for mechanical amplification of the actuation. The effectiveness of such a layer in general would be compromised due to the lack of subdivision of the actuation electrodes. A performance comparison plot is presented which compares this method of controlling beam vibration with a point-feedback and drive approach.—JAH

6,959,787

43.40.Tm ELEVATOR CAR FRAME VIBRATION DAMPING DEVICE

Josef Husmann, assignor to Inventio AG
1 November 2005 (Class 187/292); filed in the European Patent Office 7 March 2002

The device described in this patent is intended to suppress vibrations of a frame that is guided on guiderails by means of guide elements and that carries an elevator car body. Vibrations perpendicular to the direction of travel are sensed and used to control actuators that act between the frame and guide elements. Shear deformations of the frame are sensed by means of one or more accelerometers and/or strain gages, laser systems, or gyros.—EEU

6,954,666

43.50.Ki METHOD FOR LOCAL REDUCTION OF THE OPERATING NOISE PRODUCED BY A MEDICAL DIAGNOSTIC OR THERAPY DEVICE, AND MEDICAL DIAGNOSTIC OR THERAPY DEVICE HAVING A DEVICE FOR IMPLEMENTING SUCH A METHOD

Mario Bechtold and Ralph Oppelt, assignors to Siemens Aktiengesellschaft
11 October 2005 (Class 600/410); filed in Germany 16 April 2002

In order to reduce noise levels generated by the operation of a medical diagnostic or therapeutic device, such as a magnetic resonance tomograph or lithotripter, an active noise-cancellation system was developed to sense the noise. A highly directional parametric loudspeaker is used to generate the antinoise from an inversion modulation unit. Disturbing operating noises can be limited locally and are said to be suppressed effectively.—DRR

6,963,647

43.50.Ki CONTROLLED ACOUSTIC WAVEGUIDE FOR SOUNDPROOFING

Jan Krueger and Philip Leistner, assignors to Fraunhofer-Gesellschaft zur Foerderung der angewandten Forschung e.V.
8 November 2005 (Class 381/71.5); filed in Germany 15 December 1998

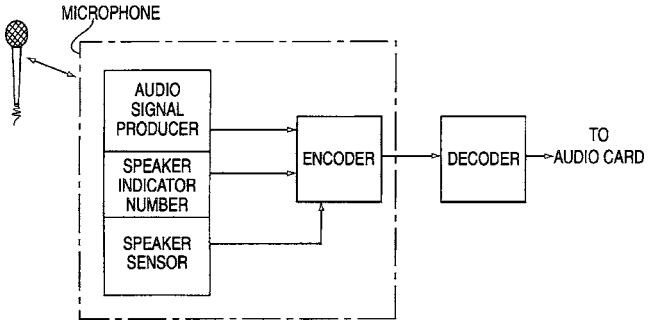
This patent covers modifications of an existing patent to enhance the active noise cancellation system in a duct by tuning the resonance of a hollow chamber aside the duct.—CJR

6,959,095

43.60.Bf METHOD AND APPARATUS FOR PROVIDING MULTIPLE OUTPUT CHANNELS IN A MICROPHONE

Raimo Bakis and Mark E. Epstein, assignors to International Business Machines Corporation
25 October 2005 (Class 381/122); filed 10 August 2001

This simple patent deals with the use of a single microphone in news gathering or other activities involving a number of different sound sources. The output of the single microphone may be assigned automatically to a



specific channel downstream by recognition of the source (speaker sensor), by audio quality (audio signal producer), or by intervention (speaker indicator number). The process is shown in the figure.—JME

6,963,649

43.60.Bf NOISE CANCELLING MICROPHONE

Michael A. Vaudrey and William R. Saunders, assignors to Adaptive Technologies, Incorporated
8 November 2005 (Class 381/94.7); filed 3 October 2001

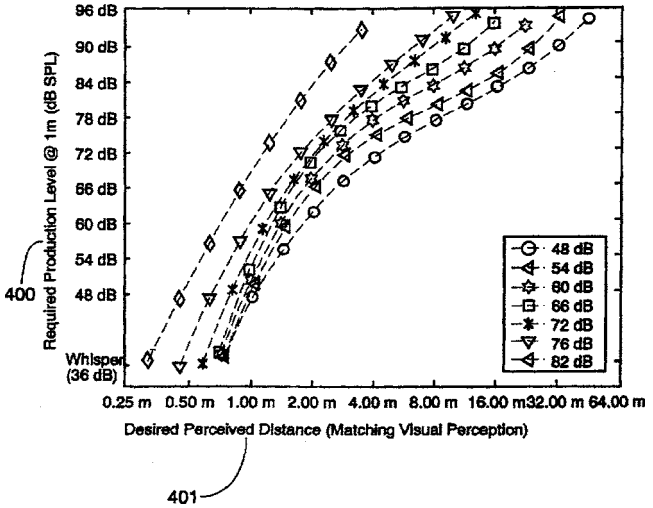
Traditionally, noise cancelling microphones have relied on the differences between far and near sources and the pressure gradients that each source produces at the transducing element. This patent describes an alternate method in which adaptive filtering is used to identify, and thus cancel, external noises. A similar technique has been used previously in effectively altering microphone pickup patterns. Properly implemented, it should work very well.—JME

6,956,955

43.60.Jn SPEECH-BASED AUDITORY DISTANCE DISPLAY

Douglas S. Brungart, assignor to The United States of America as represented by the Secretary of the Air Force
18 October 2005 (Class 381/310); filed 6 August 2001

In many military cockpit applications, HRTFs are used to define the angular direction of a virtual sound source. The virtual distance of the source, and the requirement that it be in agreement with certain visual cues, is another matter and is the subject of this patent. Using a prerecorded



library of speech signals, both amplitude and vocal production cues can be used if both are carefully calibrated and presented to the subject. The figure shows the ranges in level and distance that may be used for target speech signals referred to a distance of 1 m.—JME

6,958,950

43.60.Sx METHOD OF VISUALIZING SOUND FIELDS OF INDIVIDUAL SOUND SOURCES USING ACOUSTIC HOLOGRAPHY

Yang Hann Kim and Kyoung Uk Nam, assignors to Korea Advanced Institute of Science and Technology
25 October 2005 (Class 367/8); filed in the Republic of Korea 14 January 2003

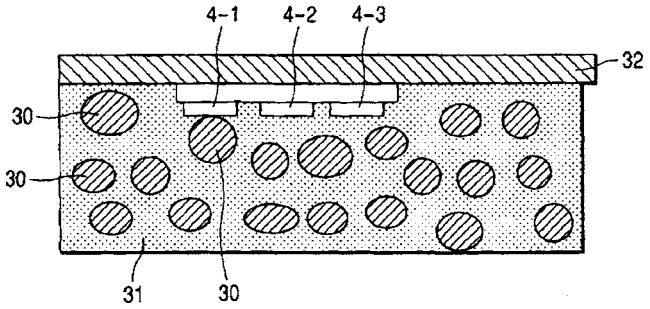
This patent deals with acoustic holography and methods of refining source acutance via a recursive program described in the abstract as follows: ''A single sound source have a position of a maximum sound pressure on the sound source plane is extracted, and a value of the sound pressure at the position is evaluated as a signal coherent to the sound source. A sound field of the extracted sound source is obtained using the coherent signal. The sound field is eliminated from sound fields of all sound sources, and it is determined whether any remaining sound field exists. The process returns to the sound field obtaining operation if any remaining sound field exists.'' Got that?—JME

6,931,929

43.60.Vx FILLER DETECTION METHOD AND FILLER DETECTION DEVICE

Takashi Sakai and Minoru Kaneko, assignors to Akebono Brake Industry Company, Limited
23 August 2005 (Class 73/579); filed in Japan 10 April 2002

By using swept sine waves, a piezoelectric transducer can be used to determine when a concrete form **32** is filled. The patent states that this can be done because the frequency characteristics of the transducer change when the material adjacent to the transducer is air (peak about 4.7 kHz), the



constituency of concrete or mortar **31** (peak at about 3.3 kHz), or aggregate **30** (peak well below 3.3 kHz). Multiple transducers **4-1**, **4-2**, **4-3** are used in an array to reduce false-fill detection and to determine the state of concrete fill throughout a complex form.—NAS

6,961,433

43.66.Qp STEREOPHONIC SOUND FIELD REPRODUCING APPARATUS

Toru Ishii, assignor to Mitsubishi Denki Kabushiki Kaisha
1 November 2005 (Class 381/17); filed 16 April 2001

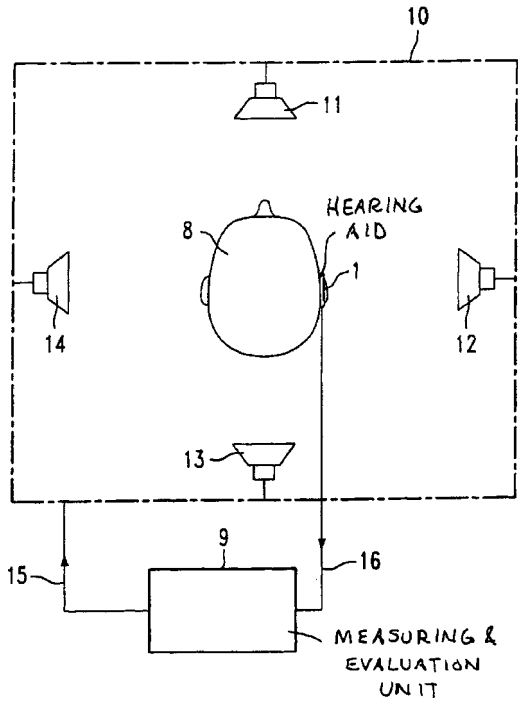
Head-tracking systems can be used to stabilize a reference sound field for normal head movements by a binaural listener. This patent describes an alternative method in which a cluster of stationary transducers at each ear is used to create a stable frame of reference for the listener.—JME

6,954,535

43.66.Ts METHOD AND ADAPTING A HEARING AID, AND HEARING AID WITH A DIRECTIONAL MICROPHONE ARRANGEMENT FOR IMPLEMENTING THE METHOD

Georg-Erwin Arndt *et al.*, assignors to Siemens Audiologische Technik GmbH
11 October 2005 (Class 381/60); filed in Germany 15 June 1999

To reduce amplitude and phase differences between two or more omnidirectional microphones used to form a directional hearing aid system,



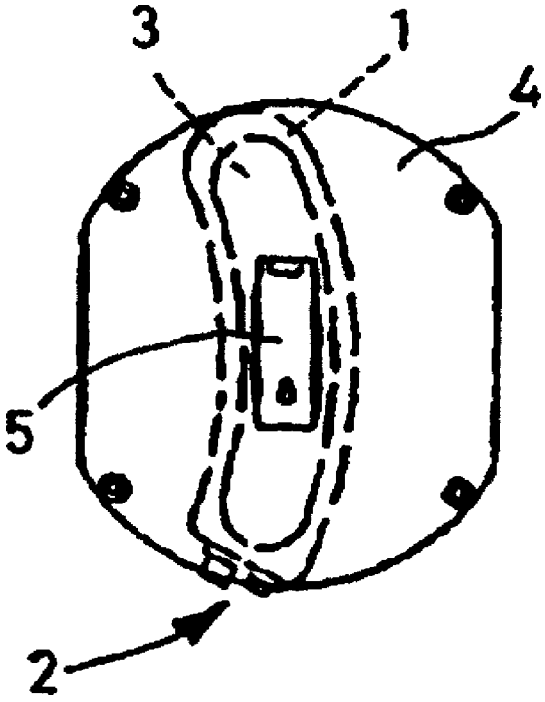
microphone matching is performed with acoustic sources from several directions *in situ* on the hearing aid wearer, rather than at the factory, based on diffraction obtained on the KEMAR mannequin.—DAP

6,959,097

43.66.Ts BEHIND THE EAR HEARING AID WITH FRONT PLATE

Hans-Dieter Borowsky, assignor to Auric Hörsysteme GmbH & Company
25 October 2005 (Class 381/322); filed in Germany 7 July 1999

A plastic faceplate that is normally used to make custom in-the-ear hearing aids is contoured to the shape of a behind-the-ear (BTE) hearing aid and attached to a BTE housing shell. On the faceplate is mounted a battery



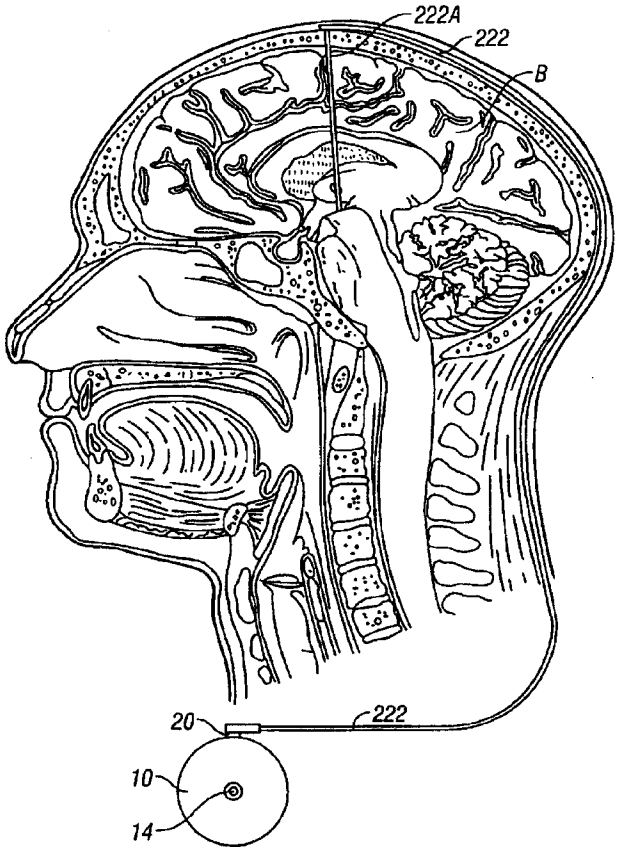
compartment and electronic components for the hearing aid at locations within the BTE housing shape. Extra plastic material around the periphery of the faceplate extending beyond the desired BTE shape is ground off.—DAP

6,944,497

43.70.Dn SYSTEM AND METHOD OF TREATING STUTTERING BY NEUROMODULATION

Paul H. Stypulkowski, assignor to Medtronic, Incorporated
13 September 2005 (Class 607/2); filed 31 October 2001

This patent describes possible treatments for stuttering, including either drug injection and/or electrical stimulation directly to one or more specific brain areas. A voicing detector would be placed (externally) near the glottis to detect the onset of voicing attempts. A controller would then adjust



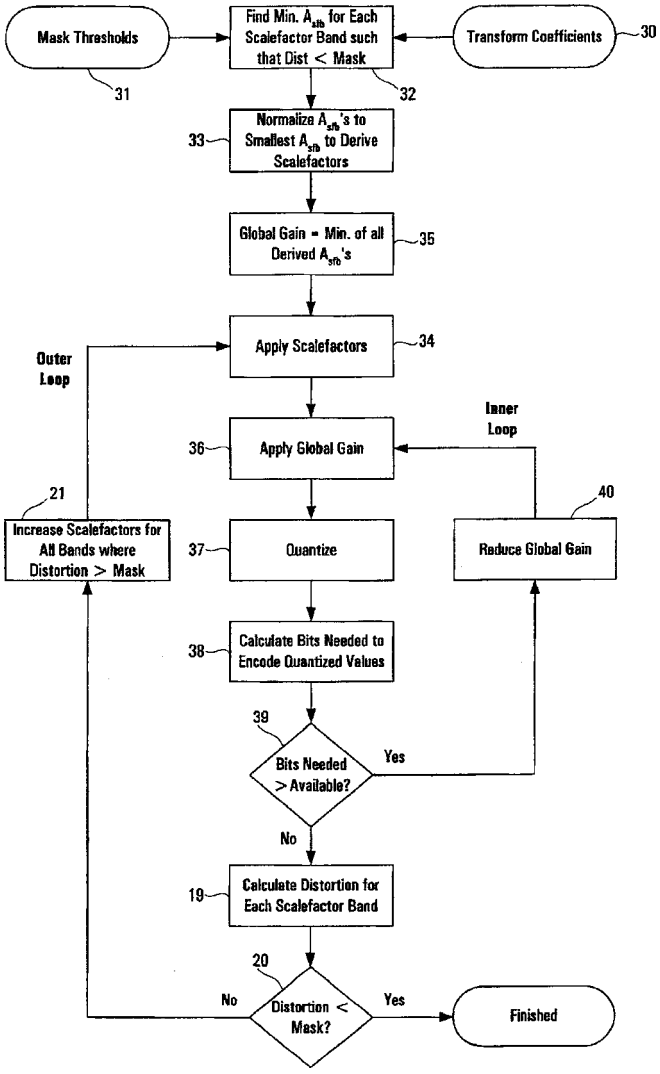
the therapeutic action according to measured timing and other patterns detected at the glottis as a result of stimulation trials. A long list of potentially relevant brain locations is given, but there is no discussion of how the specific ''preselected'' brain area is to be determined, other than some discussion of current theories of the brain control of speech.—DLR

6,950,794

43.72.Gy FEEDFORWARD PREDICTION OF SCALEFACTORS BASED ON ALLOWABLE DISTORTION FOR NOISE SHAPING IN PSYCHOACOUSTIC-BASED COMPRESSION

Girish P. Subramaniam and Raghunath K. Rao, assignors to Cirrus Logic, Incorporated
27 September 2005 (Class 704/200.1); filed 20 November 2001

To reduce the number of iterations and processing power required for quantizer bit/noise allocation during data compression, several distortion thresholds are associated with each frequency subband of the signal,



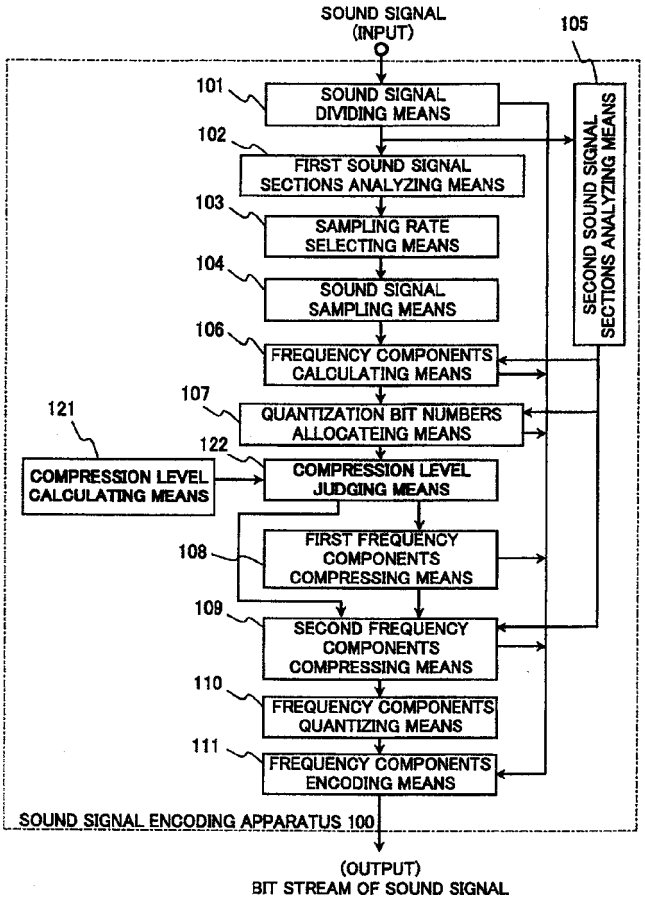
resulting in one transform coefficient and one total scaling value for each frequency subband. The product of the transform coefficient and total scaling value must be less than one of the corresponding distortion thresholds.—DAP

6,963,646

43.72.Gy SOUND SIGNAL ENCODING APPARATUS AND METHOD

Yoshiaki Takagi and Yasuhito Watanabe, assignors to Matsushita Electric Industrial Company, Limited
8 November 2005 (Class 381/23); filed in Japan 24 November 2000

To improve sound quality by preventing excessive compression of sound-signal frequency components, a method is proposed for encoding a



two-channel input signal. The decision of whether or not to perform compression is based on determining if the compression level for each input-signal segment exceeds a predetermined compression threshold.—DAP

6,934,684

43.72.Ne VOICE-INTERACTIVE MARKETPLACE PROVIDING PROMOTION AND PROMOTION TRACKING, LOYALTY REWARD AND REDEMPTION, AND OTHER FEATURES

Ahmet Alpdemir and Arthur James, assignors to Dialsurf, Incorporated
23 August 2005 (Class 704/265); filed 17 January 2003

This device is a fully computerized, voice-response system which provides descriptive or sales information on company goods or services. Customer product-assistance functions may also be provided. The system would be set up to respond via telephone, providing voice response, or as an Internet server, providing the requested information in the form of web pages. This sort of company information would be available for multiple businesses in an area in the form of subscriptions to a database service operated according to methods described in the patent.—DLR

6,937,983

43.72.Ne METHOD AND SYSTEM FOR SEMANTIC SPEECH RECOGNITION

Juan Rojas Romero, assignor to International Business Machines Corporation
30 August 2005 (Class 704/257); filed in the European Patent Office 20 December 2000

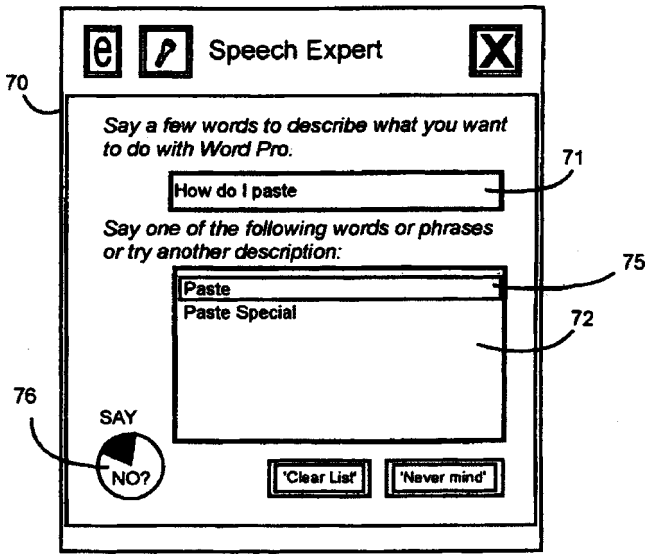
This somewhat unusual speech recognition system is described as being based on semantic principles, essentially bypassing the usual syntactic analysis. Very little description is given for the phonetic recognizer portions of the system, basically assuming that some sort of word spotting can be performed to detect any of a large vocabulary of keywords. These words are then used, possibly with n-gram statistics, or perhaps only by the combination-occurrence statistics, to retrieve an action command. One of the stated goals is that the user would not need to be familiar with a specific command vocabulary or syntax. This would seem to imply an implausibly large keyword vocabulary and raises many issues of training.—DLR

6,937,984

43.72.Ne SPEECH COMMAND INPUT RECOGNITION SYSTEM FOR INTERACTIVE COMPUTER DISPLAY WITH SPEECH CONTROLLED DISPLAY OF RECOGNIZED COMMANDS

Scott Anthony Morgan *et al.*, assignors to International Business Machines Corporation
30 August 2005 (Class 704/270); filed 17 December 1998

This patent describes a speech recognition system organized for the purpose of controlling some sort of function, either in the form of hardware or software. The present system is applicable in the case that a visual display is available during system usage, but where additional inputs requiring hand operation may not be available. Such a display would be used, among other



purposes, to assist in the resolution of recognition errors. Once a potential command has been recognized, either the command itself or some representation of the executable function would be displayed for confirmation. During a preset time interval, the user may then abort the command via further input. Upon timeout, the command would be executed as displayed. In the figure, the clock in the lower left represents a countdown timer.—DLR

6,937,986

43.72.Ne AUTOMATIC DYNAMIC SPEECH RECOGNITION VOCABULARY BASED ON EXTERNAL SOURCES OF INFORMATION

Lawrence A. Denenberg *et al.*, assignors to Comverse, Incorporated
30 August 2005 (Class 704/275); filed 28 December 2000

This is a telephone-accessible, information database intended for a wide audience of users and covering current topics, perticularly weather and news reports. Once a particular user has been identified and verified, personal information, such as address book and calendar, may also be available. In addition to the usual range of vocabulary items in the grammar related to access mechanisms and to the categories of the accessible data, additional

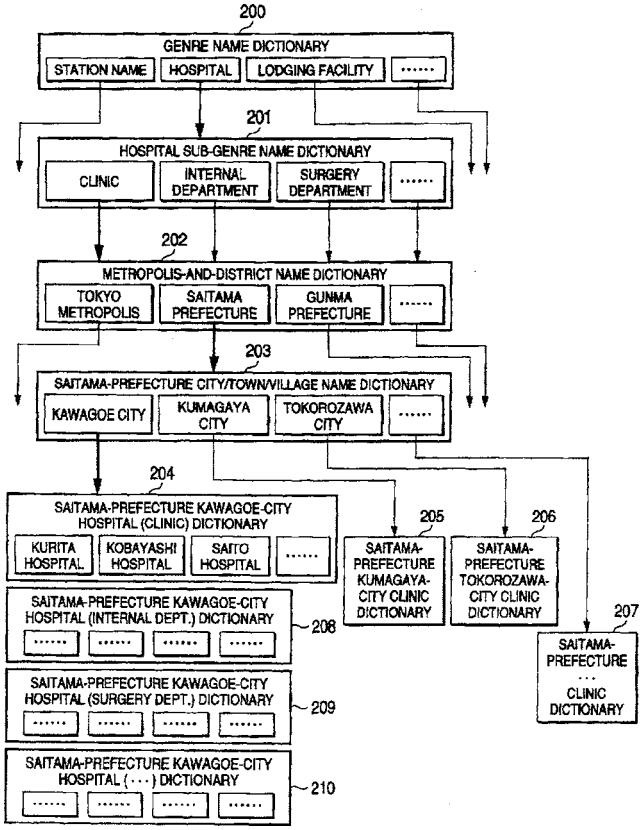
grammar items are continually extracted from current content, such as news stories. Such words, newly added to the grammar, are then keyed to the applicable stories from which they were obtained, while also considering the possibility that a given word occurred in multiple sources.—DLR

6,961,706

43.72.Ne SPEECH RECOGNITION METHOD AND APPARATUS

Hiroshi Saito, assignor to Pioneer Corporation
1 November 2005 (Class 704/275); filed in Japan 12 October 2000

Some car navigation systems use voice input and speech recognition to narrow down the possibilities for designating a target location. During the



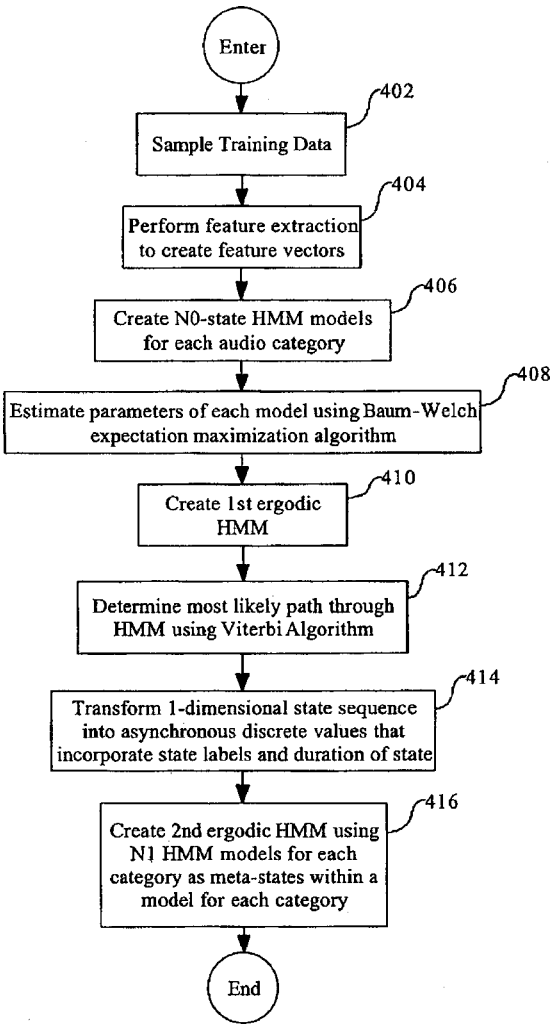
iterative process, if the user cannot respond to a prompt, the location designation is aborted. To prevent this problem, a system of hierarchical speech-recognition dictionaries with skipping capability is proposed.—DAP

6,963,835

43.72.Ne CASCADED HIDDEN MARKOV MODEL FOR META-STATE ESTIMATION

Steven F. Kimball and Joanne Como, assignors to BAE Systems Information and Electronic Systems Integration Incorporated
8 November 2005 (Class 704/235); filed 31 March 2003

Speech is distinguished from silence and sounds such as tones or music by cascading two hidden Markov models (HMMs). This facilitates modeling processes with segments of varying duration and complex temporal



characteristics. Training data are used to create models of the input signal types. The first HMM stage segments short data samples into multilevel states which are used for inputs to the second HMM stage, which, in turn, models the event sequence and duration.—DAP

6,963,633

43.72.Ne VOICE DIALING USING TEXT NAMES

William F. Diede *et al.*, assignors to Verizon Services Corporation
8 November 2005 (Class 379/88.03); filed 30 June 2000

A voice-activated dialing feature is implemented for an advanced intelligent (telephone) network (AIN) using Internet-based management techniques. A voice-dialing record is maintained for each subscriber which can also specify, for example, more than one phone number and location for the individuals who will be called. Subscribers can provide speech samples for hard-to-pronounce words.—DAP

6,949,701

43.75.Hi DRUMHEAD

Yukimasa Okumura, assignor to Yamaha Corporation
27 September 2005 (Class 84/411 R); filed in Japan 18 January 2002

Historically, drum heads were made of animal skins. This patent details an artificial drum head manufacturted by laminating two or more films together. However, as amateur laminators know, air bubbles will be formed. This inventor details all variety of air hole shapes to remove bubbles during pressing.—MK

6,946,594

43.75.Pq METHOD FOR REPRODUCING THE SOUND OF AN ACCORDION ELECTRONICALLY

Luigi Bruti *et al.*, assignors to Roland Europe S.p.A.
20 September 2005 (Class 84/604); filed in Italy 27 April 2001

Imagine taking the reeds out of a keyboard or button accordian. Then, make each key or button act as a sensor for a microprocessor. Additionally, a critical pressure sensor is added to measure the pressure created by the bellows. Now, when the bellows pressure is greater than the ''on'' pressure, the tone generator for that note can be turned on. In addition, a reed-closing click can be simulated by storing the sound in the generator.—MK

6,960,715

43.75.St MUSIC INSTRUMENT SYSTEM AND METHODS

Jerry Riopelle, assignor to HumanBeams, Incorporated
1 November 2005 (Class 84/725); filed 16 August 2002

Imagine using interrupted light beams instead of keys as a musical input device. By using multiple beams, multiple controls are possible. However, breaking a beam is essentially a binary decision and therefore this necessitates the use of a sequencer to control the note sequence from the synthesizer. Fine control really isn't possible.—MK

6,961,288

43.75.Tv SOUND-SCALE GENERATION DEVICE AND TIME-ANNOUNCING CLOCK

Takashi Iijima, Aizuwakamatsu-shi, Fukushima, Japan
1 November 2005 (Class 368/272); filed in Japan 25 October 2002

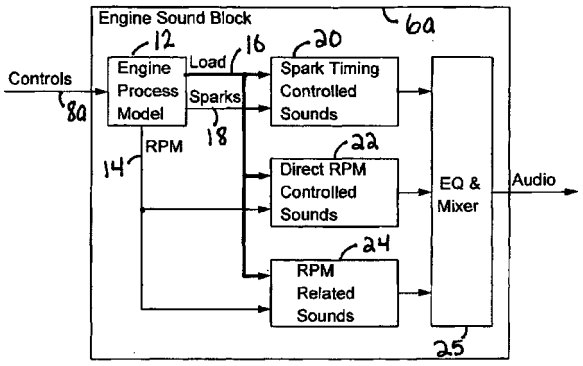
The inventor only wants to hear a ''pleasant sounding'' (i.e., consonant) scale in an audio clock, where each hour is an interval using middle C as the reference. The inventor argues that ''due to its cheefulness, it is possible to reduce the dissonant reverberation to about one-quarter psychologically.''—MK

6,959,094

43.75.Wx APPARATUS AND METHODS FOR SYNTHESIS OF INTERNAL COMBUSTION ENGINE VEHICLE SOUNDS

Kim Cascone *et al.*, assignors to Analog Devices, Incorporated
25 October 2005 (Class 381/86); filed 20 April 2001

Physical models for musical instruments are all the rage. Why not apply the same principles to the synthesis of internal-combustion engine



sounds for use in computer gaming? Although short on details, this disclosure lays out the basic approach.—MK

6,952,912

43.80.Qf METHOD AND DEVICE FOR ANALYZING RESPIRATORY SOUNDS IN HORSES

David H. Lambert, assignor to Airway Dynamics, LLC
11 October 2005 (Class 54/6.1); filed 19 February 2003

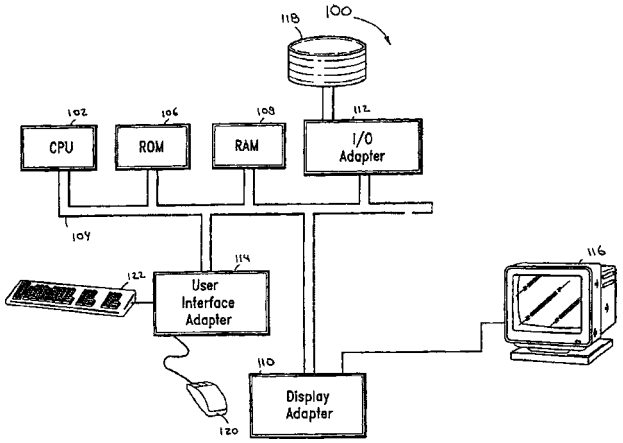
The patent covers a device and method for recording sounds during the course of exercising horses. This is said to enable the horse trainer to identify upper airway abnormalities and to assess athletic potential by measuring expiratory and inspiratory times under either exercising or actual racing conditions. The respiratory sounds are picked up by a transducer placed in direct contact with the skin overlaying the skull of the animal, namely the frontal sinus or nasal turbinates. The device incorporates an acoustic monitoring system that produces an output signal not attenuated or corrupted by extraneous noises such as wind, hoof beats, etc.—DRR

6,953,436

43.80.Qf MULTI-MODAL CARDIAC DIAGNOSTIC DECISION SUPPORT SYSTEM AND METHOD

Raymond Watrous and Nathaniel Reichek, assignors to Zargis Medical Corporation
11 October 2005 (Class 600/528); filed 12 September 2002

This patent is a continuation of United States Patent 6,572,560 that is directed to a multimodal, cardiac-diagnostic, decision-support system and method. This is a medical tool that can detect valvular heart diseases, discriminate pathological from innocent heart murmurs, and render a decision



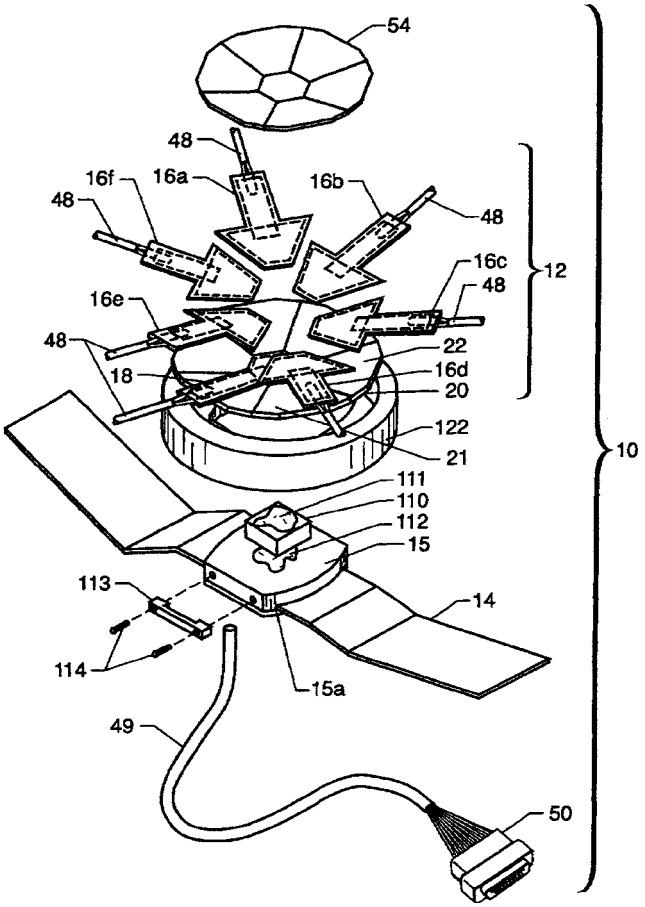
about referring a patient for an echocardiography study, while rationalizing and documenting the basis for the referral decision. The cardiac signal is analyzed through the use of a neural network analysis of the cardiac acoustic signal, using a wavelet decomposition to extract time-frequency information.—DRR

6,954,971

43.80.Qf METHOD FOR SIMULTANEOUSLY MAKING A PLURALITY OF ACOUSTIC SIGNAL SENSOR ELEMENTS

Timothy D. Bryant *et al.*, assignors to The United States of America as represented by the Administrator of the National Aeronautics and Space Administration
18 October 2005 (Class 29/25.35); filed 30 September 2002

This is a fetal heart monitoring system consisting of a backing plate that may have a generally concave front surface, a generally convex back surface, and at least one sensor element attached to the concave front surface for acquiring acoustic fetal heart signals produced by a fetus *in utero*. The



sensor element is shaped to conform to the generally concave surface of the backing plate. In one embodiment at least one sensor element comprises an inner sensor and multiple surrounding outer sensors. The fetal heart monitoring system can also consist of a web belt and a web belt guide movably attached to the web belt.—DRR

6,962,566

43.80.Qf MEDICAL DIAGNOSTIC ULTRASOUND INSTRUMENT WITH ECG MODULE, AUTHORIZATION MECHANISM AND METHODS OF USE

Jens U. Quistgaard *et al.*, assignors to Sonosite, Incorporated
8 November 2005 (Class 600/437); filed 1 February 2002

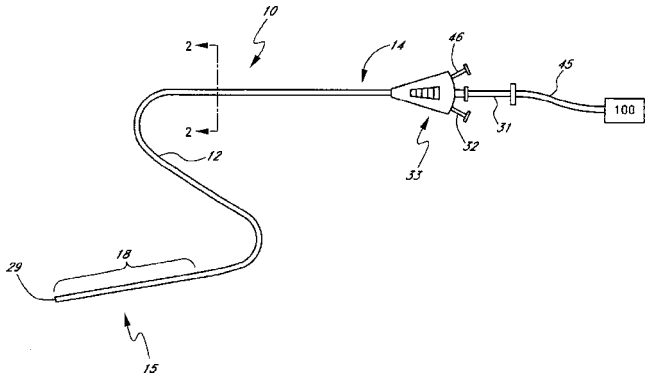
A handheld ultrasound-diagnostic instrument is disclosed having multiple diagnostic modes, including pulse/continuous wave Doppler, time-motion analysis, spectral analysis, and tissue harmonic imaging. An external electrocardiograph (ECG) recording unit is also provided. The ECG unit is said to be adaptable for use with the handheld ultrasound instrument to provide ECG monitoring, while the latter performs an ultrasound scan in B-mode, Doppler, color Doppler, M-mode, and CW/PW mode. The instrument also includes a security mechanism, allowing any combination of modes to be enabled by the manufacturer. A method is also disclosed for a manufacturer to maintain a database of handheld-ultrasound-instrument capabilities after these instruments enter the market.—DRR

6,958,040

43.80.Sh MULTI-RESONANT ULTRASONIC CATHETER

Leonard R. Oliver and Richard R. Wilson, assignors to Ekos Corporation
25 October 2005 (Class 600/439); filed 27 December 2002

This catheter system is designed to deliver ultrasonic energy and a therapeutic compound to a treatment site within a patient's vasculature. The tubular body of the catheter has an energy delivery section positioned between the proximal and distal ends. The system also has an inner core of an electrically conductive material that is configured for insertion into the



tubular body. In addition, there are multiple ultrasound radiating members positioned on the outer surface of the inner core which are electrically connected to the core material. A control circuit delivers an input signal to the ultrasound radiating members. The idea behind the use of this catheter is to provide ultrasonic energy at multiple frequencies rather than being limited to the resonant property of a single radiation member.—DRR

6,953,434

43.80.Vj METHOD AND APPARATUS TO ENHANCE ULTRASOUND CONTRAST IMAGING USING STEPPED-CHIRP WAVEFORMS

Xiaohui Hao *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
11 October 2005 (Class 600/458); filed 24 September 2002

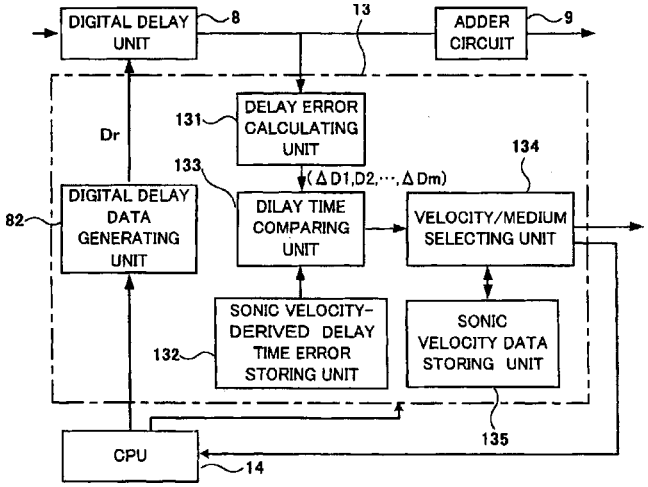
Contrast-to-tissue and signal-to-noise ratios in bubble-contrast imaging are enhanced using stepped-chirp waveforms. In this method, the first component of a transmit waveform has a frequency optimized to initiate bubble vibration and the second component has a frequency optimized to produce an enhanced, nonlinear bubble response. The waveform components are transmitted as a single-stepped chirp signal. The relative phase, switching time, and time delay between transmit waveform components are also adjusted for maximal enhancement of the bubble nonlinear response.—RCW

6,958,041

43.80.Vj ULTRASONIC IMAGING DEVICE

Hirotaka Baba *et al.*, assignors to Hitachi Medical Corporation
25 October 2005 (Class 600/443); filed in Japan 10 March 2000

The time delay used by this imaging system to form beams is adjusted



using a time delay error found by comparing time delays from various sound velocities that are stored.—RCW

6,958,042

43.80.Vj ULTRASONIC TRANSMISSION/ RECEPTION METHOD, ULTRASONIC TRANSMISSION/RECEPTION APPARATUS, ULTRASONIC IMAGING METHOD AND ULTRASONIC IMAGING APPARATUS

Masayoshi Honda, assignor to GE Medical Systems Global Technology Company, LLC
25 October 2005 (Class 600/458); filed in Japan 11 April 2001

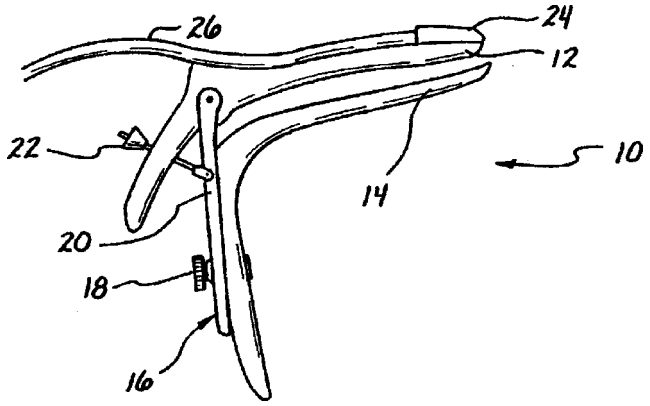
A low sound pressure that does not destroy contrast bubbles is used during intermittent scanning with transmit pulse modulation. Pulse compression is performed on the echo signals. The pulse-compressed signals are used to form an image.—RCW

6,960,166

43.80.Vj SPECULUM HAVING ULTRASOUND PROBE

Irwin Lane Wong, Tustin, California and Harry Hatasaka, Park City, Utah
1 November 2005 (Class 600/221); filed 5 November 2002

An ultrasound probe included in this speculum permits visualization



during procedures such as transcervical embryo transfer during *in vitro* fertilization.—RCW

6,960,169

43.80.Vj SPREAD SPECTRUM CODING FOR ULTRASOUND CONTRAST AGENT IMAGING

Zuhua Mao *et al.*, assignors to Siemens Medical Solutions USA, Incorporated
1 November 2005 (Class 600/458); filed 19 May 2003

The contrast-to-tissue and signal-to-noise ratios during contrast agent imaging are improved by transmitting a waveform that contains pulses at two frequencies. Signals received at an intermodulation frequency of the transmitted frequencies are isolated and used for imaging. Differences in contrast-agent and tissue responses are used to identify signals from the contrast agent with minimal influence from tissue signals.—RCW

LETTERS TO THE EDITOR

Dual-probe laser interferometer for structural health monitoring (L)

Stefan Hurlebaus[a)]
Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas 77843-3136

Laurence J. Jacobs[b)]
School of Civil and Environmental Engineering, G.W.W. School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0355

(Received 12 December 2005; revised 9 January 2006; accepted 10 January 2006)

In this note we present the development of a dual-probe laser interferometer that uses the filtering properties of a polarized beamsplitter to enable two independent (uncoupled) detection probes. The robustness of this system is demonstrated by making broadband, noncontact, high fidelity measurements of Lamb waves in an aluminum plate. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2170442]

PACS number(s): 43.35.Cg, 43.38.Ne, 43.35.Sx [YHB] Pages: 1923–1925

I. INTRODUCTION

Recent developments in quantitative structural health monitoring have presented the need for the measurement of the frequency-dependent attenuation and velocity of guided ultrasonic Lamb waves that propagate in plate-like components. These Lamb waves are dispersive and multimode, so it is advantageous to utilize the point-like, high fidelity, and noncontact nature of laser interferometry in their detection. Of particular importance is the ability to simultaneously measure the same Lamb wave at two different spatial locations. This ability to detect the same (dispersive) Lamb wave at two different propagation distances is critical for attenuation and scattering applications, which require an absolute comparison of frequency-dependent amplitudes, as a function of the spatial location. For example, Benz *et al.*[1] developed a correlation procedure to locate and size a notch in a plate by comparing the reflected and transmitted Lamb wave fields.

Existing dual-probe interferometer designs either use the reference beam and the object beam as the two individual probes to make a differential measurement,[2–5] or simply combine the components of two (independent) single-probe interferometers.[6] The multimode and dispersive nature of Lamb waves complicates their interrogation with a differential measurement system; since a time-domain Lamb wave signal is relatively long, the (optical) signals from the (coupled) reference and object beams can interfere with each other, causing spurious results in the Lamb wave signal. In contrast, in this note we present a dual-probe laser interferometer that provides the following advantages: *two independent (uncoupled) simultaneous measurements*; *and a reduced number of optical components*.

II. DEVELOPMENT OF THE DUAL-PROBE INTERFEROMETER

First consider the single-probe heterodyne interferometer shown in Fig. 1 that was developed in Bruttomesso *et al.*;[7] note that readers interested in more details on laser interferometry should see Scruby and Drain.[8] An argon-ion (continuous wave) laser generates a single beam of vertically polarized laser light (wavelength of 514.5 nm). The beam passes through an acousto-optic modulator (AOM) that splits this original beam into an infinite number of separate beams. Each beam has its frequency shifted by a specific frequency (40 MHz in this case), known as the beat frequency. The zeroth-order (unshifted) beam and the first-order (shifted) beam contain approximately 95% of the power from the original incident beam, and these are the only two beams used in this interferometer. The first-order beam eventually reflects off the specimen (object) surface, and is referred to as the object beam. The zeroth order beam (which passes to the photodiode without ever reflecting off the specimen surface) is referred to as the reference beam.

After leaving the AOM, the vertically polarized object beam passes through a polarized beamsplitter (PBS). The PBS causes vertically polarized light to be reflected at 90°,

[a)]Electronic mail: shurlebaus@civil.tamu.edu; URL: http://www.civil.tamu.edu
[b)]Electronic mail: laurence.jacobs@ce.gatech.edu; URL: http://www.ce.gatech.edu

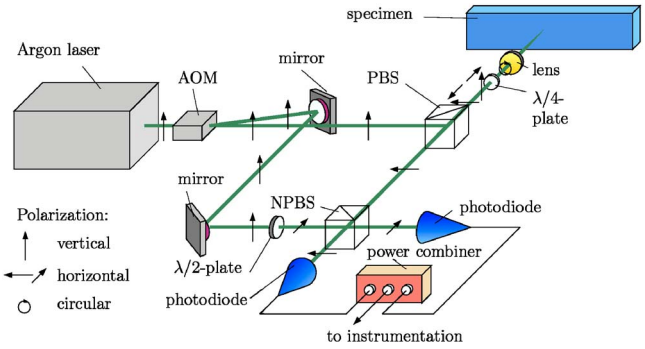


FIG. 1. (Color online) Schematic of the single-probe interferometer.

but allows horizontally polarized light to pass straight through. Consequently, the object beam is reflected 90°. The object beam then passes through a quarter-wave plate (λ/4 plate), at which point the beam becomes circularly polarized. The object beam then passes through a lens (microscope objective) that focuses the beam onto the surface of the object (specimen). The object beam then returns through the lens and the λ/4 plate, which causes the beam to become horizontally polarized. This horizontally polarized object beam then passes back through the PBS (this time unaffected) and continues into a nonpolarized beamsplitter (NPBS).

In contrast, the vertically polarized reference beam that leaves the AOM is reflected off two mirrors and passes through a half-wave plate (λ/2 plate), which changes the polarization from vertical to horizontal. This reference beam then passes through the NPBS, where it is recombined with the reflected object beam and focused onto two photodiodes. The signals of both photodiodes contain the same information and are combined with a power combiner in order to increase the carrier signal, and therefore the signal-to-noise ratio (SNR).

Now consider a dual-probe interferometer where the incident beam is separated into two different polarization directions, and each of these beams is used to build an individual probe. This design makes it possible to have the information of each of the individual probes contained in a single object beam without corrupting or interfering with each other. This dual-probe heterodyne interferometer (shown in Fig. 2) is similar to the single-probe version shown in Fig. 1, except that a λ/2 plate is now placed in front of the AOM, and this λ/2 plate is rotated such that the incident vertically polarized light is changed to 45° polarized

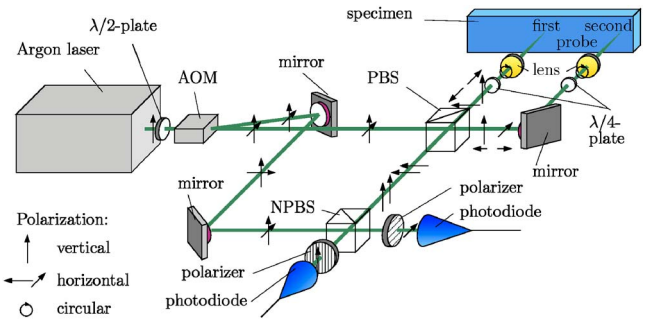


FIG. 2. (Color online) Schematic of the dual-probe interferometer.

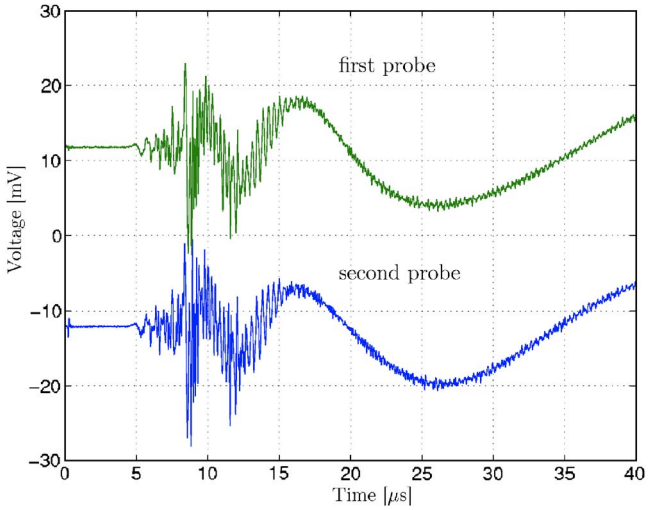


FIG. 3. (Color online) Lamb waves (independent and simultaneous) measured with the dual-probe interferometer.

light—a superposition of horizontally and vertically polarized light. By using the PBS as a filter, the vertically polarized beam is used as the object beam of the first probe, while the horizontally polarized beam (which passes through the PBS unchanged) becomes the object beam of the second probe. The reference beam of this system is also 45° polarized, and is recombined with the two (separately polarized) object beams (probe one horizontal, and probe two vertical) at the NPBS. The information of each object beam is combined with its (polarization appropriate) reference beam at the two photodiodes. Each photodiode in the dual-probe interferometer contains independent information from each of the single probes.

Note that the (electrical) signal out of each photodiode is first bandpass filtered (40 MHz), and then passes through a 40 MHz frequency modulation (FM) discriminator. It is important to note that the modulation of the combined (object/reference) beam is equal to the frequency of the object beam, plus any Doppler shift due to the (out-of-plane) velocity of the point on the specimen's (object's) surface. The output signal from the FM discriminator is then lowpass filtered (10 MHz) and captured on a digital oscilloscope.

III. MEASUREMENT OF LAMB WAVES AND DISCUSSION

The robustness and accuracy of this dual-probe interferometer is demonstrated by simultaneously measuring the same Lamb wave at two different spatial locations, but with the same propagation distance. Broad bandwidth Lamb waves are generated with an Nd:YAG laser (4–6 ns pulse) (see Scruby and Drain[8] for details on the laser generation of Lamb waves). The laser detection of these Lamb waves is accomplished with the proposed dual-probe interferometer, measuring out-of-plane surface velocity (particle velocity) at two points on the specimen's surface. The specific specimen in this demonstration is a 1 mm thick, 3003 aluminum plate (300×300 mm), with the source-to-receiver distance the same for both probes, 26 mm (in opposite directions). Figure 3 shows the (transient) time-domain signals measured using

the first and second probes, respectively. Note that the Nd:YAG laser fires at $t=0$ and generates the Lamb wave at the source location (the spot where the Nd:YAG hits the plate) and each Lamb wave signal represents an average of 30 Nd:YAG shots to increase SNR.

It is clear from Fig. 3 that this proposed dual-probe interferometer makes high-fidelity, point-like, independent (there is no bleed through from the first probe to the second probe, and vice-versa) measurements with a broad bandwidth (100 kHz to 10 MHz in this example). The independent and simultaneous measurements of this dual-probe interferometer are enabled by using the filtering properties of the PBS to separate 45° polarized light into two orthogonal polarization directions. The second advantage of this proposed interferometer is that because a common reference beam can be used, it only requires a few more optical components than a "standard" single-probe interferometer. Both of these attributes are advantageous for structural health monitoring applications, where the efficient, reliable, and accurate measurement of Lamb waves are of critical importance.

[1]R. Benz, M. Niethammer, S. Hurlebaus, and L. J. Jacobs, "Localization of notches with Lamb waves," J. Acoust. Soc. Am. **1114**, 677–685 (2003).
[2]D. Gardner, T. Hofler, S. Baker, R. Yarber, and S. Garrett, "A fiber optic interferometric seismometer," J. Lightwave Technol. **5**, 953–960 (1987).
[3]A. Lewin, "Optischer Aufbau von Polytec Vibrometern," in *4. Laser Vibrometer Seminar*, Polytec, Waldbronn, 1997, pp. 1–7.
[4]J. Huang and J. D. Achenbach, "Dual-probe laser interferometer," J. Acoust. Soc. Am. **90**, 1269–1274 (1991).
[5]A. D.W. McKie, J. W. Wagner, J. B. Spicer, and J. B. Deaton, Jr., "Dual-beam interferometer for the accurate determination of surface-wave velocity," Appl. Opt. **30**, 4034–4039 (1991).
[6]J. W. Littles, L. J. Jacobs, and A. H. Zurieck, "Single-sided ultrasonic technique to characterize thick FRP composites," J. Nondestruct. Eval. **17**, 223–230 (1998).
[7]D. Bruttomesso, L. J. Jacobs, and R. D. Costley, "Development of interferometer for acoustic emission testing," J. Eng. Mech. **119**, 2303–2316 (1993).
[8]C. B. Scruby and L. E. Drain, *Laser Ultrasonics: Techniques and Applications* (Adam Hilger, Philadelphia, 1990).

Measuring flow resistivity of porous materials at low frequencies range via acoustic transmitted waves (L)

Z. E. A. Fellah
Laboratoire de Mécanique et d'Acoustique, CNRS-UPR 7051, 31 chemin Joseph Aiguier, Marseille, 13402, France

M. Fellah
Laboratoire de Physique Théorique, Institut de Physique, USTHB, BP 32 El Alia, Bab Ezzouar 16111, Algérie

N. Sebaa and W. Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Heverlee, Belgium

C. Depollier
Laboratoire d'Acoustique de l'Université du Maine, UMR-CNRS 6613, Université du Maine, Avenue Olivier Messiaen 72085 Le Mans Cedex 09, France

(Received 19 January 2005; revised 24 January 2006; accepted 3 February 2006)

An acoustic transmissivity method is proposed for measuring flow resistivity of porous materials having rigid frame. Flow resistivity of porous material is defined as the ratio between the pressure difference across a sample and the velocity of flow of air through that sample per unit cube. The proposed method is based on a temporal model of the direct and inverse scattering problem for the diffusion of transient low-frequency waves in a homogeneous isotropic slab of porous material having a rigid frame. The transmission scattering operator for a slab of porous material is derived from the response of the medium to an incident acoustic pulse. The flow resistivity is determined from the solution of the inverse problem. The minimization between experiment and theory is made in the time domain. Tests are performed using industrial plastic foams. Experimental and numerical results, and prospects are discussed. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2179749]

PACS number(s): 43.20.Bi, 43.20.Hq [ANN] Pages: 1926–1928

I. INTRODUCTION

The acoustic characterization of porous materials saturated by air[1,2] such as plastic foams, fibrous, or granular materials is of great interest for a wide range of industrial applications. These materials are frequently used in the automotive and aeronautics industries and in the building trade. One important parameter that appears in theories of sound propagation in porous materials at a low-frequency range[3,4] is the specific flow resistivity σ.[5–11] This parameter is defined as the ratio between the pressure difference across a sample and the velocity of flow of air through that sample per unit cube; the flows being considered are steady and nonpulsating. The permeability k_0 is related to the specific flow resistivity σ by the relation $k_0 = \eta/\sigma$, where η is the fluid viscosity.

Among the various systems that have been developed for the measurement of flow resistance, a distinction can be made between direct and comparative methods. With direct methods,[6,8,9,11] the pressure drop across a sample and the rate of air flow through the porous sample are determined separately and the specific flow resistivity is computed as the ratio of the two quantities. With comparative methods,[10,12] a calibrated flow resistivity is placed in series with the porous sample. The ratio of pressure drops across each element is the same as the ratio of the values of flow resistivity, since the volumetric flow of air in the line is the constant.

In this work, we present a simple acoustical method of measuring specific flow resistivity by measuring a diffusive wave transmitted by a slab of air-saturated porous material in a guide (pipe).

II. MODEL

In the acoustics of porous materials, one distinguishes two situations according to whether the frame is moving or not. In the first case, the dynamics of the waves due to the coupling between the solid skeleton and the fluid is well described by the Biot theory.[13] In air-saturated porous media the structure is generally motionless and the waves propagate only in the fluid. This case is described by the model of equivalent fluid,[4] which is a particular case of the Biot model, in which the interactions between the fluid and the structure are taken into account in two frequency response factors: the dynamic tortuosity of the medium $\alpha(\omega)$ given by Johnson *et al.*[14] and the dynamic compressibility of the fluid included in the porous material $\beta(\omega)$ given by Allard[1] (ω is the pulsation frequency). In the frequency domain, these factors multiply the density of the fluid and its compressibility, respectively, and represent the deviation from the behavior of the fluid in free space as the frequency changes.

 0001-4966/2006/119(4)/1926/3/$22.50

Consider a homogeneous porous material that occupies the region $0 \leqslant x \leqslant L$. A sound pulse impinges normally on the medium. It generates an acoustic pressure field $p(x,t)$ and an acoustic velocity field $v(x,t)$ within the material. In the low-frequency range, the acoustic fields satisfy the Euler equation and the constitutive equation (along the x axis):

$$\sigma\phi v(x,t) = -\frac{\partial p(x,t)}{\partial x},$$

$$\frac{\gamma}{K_a}\frac{\partial p(x,t)}{\partial t} = -\frac{\partial v(x,t)}{\partial x}.$$

In these equations, ϕ is the porosity, K_a is the bulk modulus of the fluid, and γ is the adiabatic constant. The Euler equation is reduced to the Darcy's law, which expresses the balance between the driving force of the wave and the drag forces $\sigma\phi v$, due to the flow resistance of the material. The fields that are varying in time, the pressure, the acoustic velocity, etc., follow a diffusion equation,[4]

$$\frac{\partial^2 p(x,t)}{\partial x^2} - D\frac{\partial p(x,t)}{\partial t} = 0, \tag{1}$$

where $D=\sigma\phi\gamma/K_a$ is the diffusion constant. The solution of this diffusive equation gives the Green's function of the porous material:

$$G(x,t) = \frac{x\sqrt{D}}{2\sqrt{\pi}}\frac{1}{t^{3/2}}\exp\left(-\frac{x^2 D}{4t}\right). \tag{2}$$

To derive the transmission scattering operator, it is assumed that the pressure field and flow velocity are continuous at the material boundary: $p(0^+,t)=p(0^-,t)$, $p(L^-,t)=p(L^+,t)$, $v(0^-,t)=\phi v(0^+,t)$, $v(L^+,t)=\phi v(L^-,t)$, where ϕ is the porosity of the medium and the $\mp$ superscript denotes the limit from left and right, respectively. Assumed initial conditions are $p\ (x,t)|_{t=0}=0$ and $\partial p/\partial t|_{t=0}=0$, which means that the medium is idle for $t=0$. The incident $p^i(t)$ and transmitted $p^t(t)$ fields are related in the time domain by the transmission scattering operator:

$$p^t(x,t) = \int_0^t \tilde{T}(\tau)p^i\left(t-\tau-\frac{(x-L)}{c_0}\right)d\tau. \tag{3}$$

The transmission operator is independent of the incident signal and depends only on the properties of the porous material. Using the relations (1)–(3), we obtain the transmission scattering operator $\tilde{T}(t)=D(t)*G(L,t)$, where $*$ denotes the time convolution operation. The operator $D(t)$ is given by

$$D(t) = -\frac{8}{B^2}\left(1+\frac{t}{B^2}\right)\exp\left(\frac{t}{B^2}\right)\mathrm{Erfc}\left(\frac{\sqrt{t}}{B}\right) + \frac{4}{B\sqrt{\pi t}}\left(2\frac{t}{B^2}+1\right), \quad B = \frac{1}{\eta}\sqrt{\phi^3\gamma\sigma^3\rho},$$

where erfc is the error function and ρ_0 is the fluid density.

III. INVERSE PROBLEM

The inverse problem is to find value for parameter σ that minimizes the function $U(\sigma)=\int_0^t[p^t_{\exp}(x,t)-p^t(x,t)]^2\,dt$, where $p^t_{\exp}(x,t)$ is the experimentally determined transmitted signal and $p^t(x,t)$ is the transmitted wave predicted from Eq. (3). The inverse problem is not solved for the porosity ϕ because the transmission is much more sensitive to flow resistivity than to porosity; the effect of the porosity in transmitted mode is negligible, as it has been observed in the asymptotic domain[15] (high-frequency range). The analytical method of solving the inverse problem using the conventional least-square method is tedious. In our case, a numerical solution of the least-square method can be found that minimizes $U(\sigma)$ defined by $U(\sigma)=\Sigma_{i=1}^{i=N}[p^t_{\exp}(x,t_i)-p^t(x,t_i)]^2$, where $p^t_{\exp}(x,t_i)_{i=1,2,\ldots,N}$ represents the discrete set of values of the experimental transmitted signal and $p^t(x,t_i)_{i=1,2,\ldots,N}$ is the discrete set of values of the simulated transmitted signal. The inverse problem is solved numerically by the least-square method.

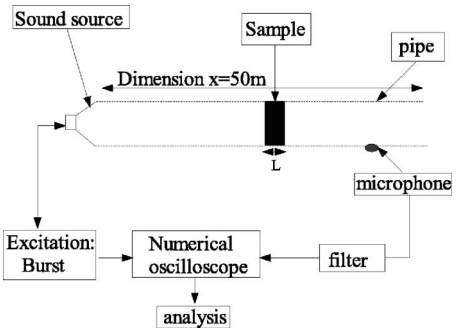


FIG. 1. Experimental setup of acoustic measurements.

Experiments are performed in a guide (pipe), having a diameter of 5 cm and of length 50 m. This length has been chosen for the propagation of transient signals at low frequency. It is not important to keep the pipe straight; it can be rolled in order to save space without perturbations on experimental signals (the cutoff frequency of the tube $f_c \sim 4$ kHz). A sound source Driver unit "Brand" constituted by loudspeaker Realistic 40-9000 is used. Bursts are provided by synthesized function generator Stanford Research Systems model DS345-30 MHz. The signals are amplified and filtered using model SR 650-Dual channel filter, Standford Research Systems. The signals (incident and transmitted) are measured using the same microphone (Bruel&Kjaer, 4190) in the same position in the tube. The incident signal is measured without a porous sample, however, the transmitted signal is measured with the porous sample. The experimental setup is shown in Fig. 1. Consider a cylindrical sample of plastic foam M of diameter 5 cm, porosity $\phi=0.9$, and thickness 2.5 cm. Sample M was characterized using classical methods[6] given $\sigma=38\,000\pm6000$ N m^{-4} s. Figure 2 shows the experimental incident signal (solid line) generated by the loudspeaker in the frequency bandwidth (85–115) Hz, and the experimental transmitted signal (dashed line). After solving the inverse problem numerically for the flow resistivity, we find the following optimized value: $\sigma=39\,500\pm2000$ N m^{-4} s. We present in Fig. 3 the variation of the minimization function U with the flow resistivity σ. In Fig. 4, we show a comparison between an experimental transmitted signal and simulated transmitted signal for the optimized value of the flow resistivity. The difference between the two curves is slight, which

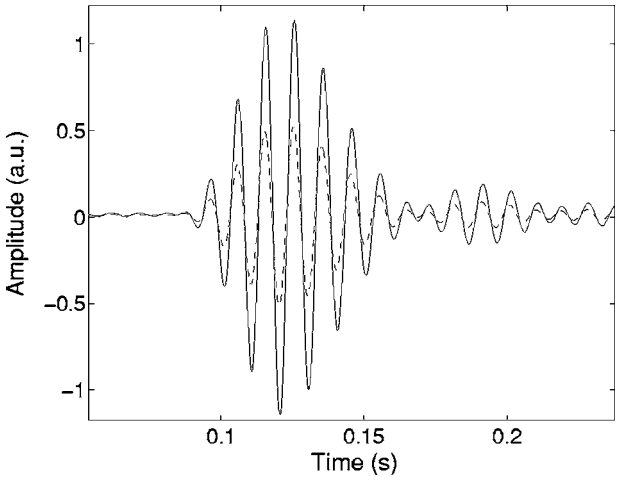


FIG. 2. Experimental incident signal (solid line) and experimental transmitted signal (dashed line).

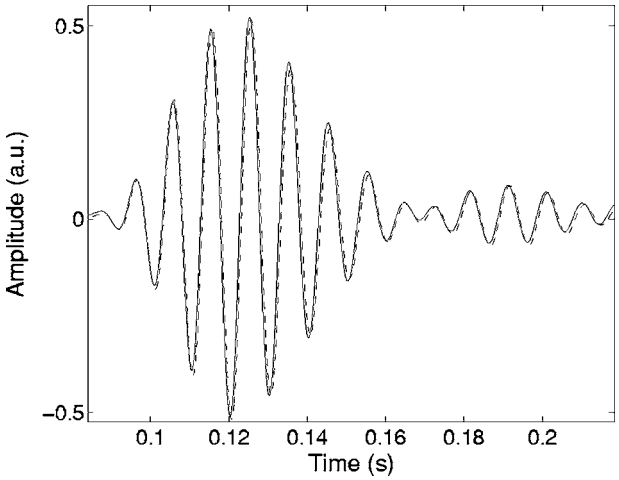


FIG. 4. A comparison between the experimental transmitted signal (dashed line) and the simulated transmitted signal (solid line) for the sample M.

leads us to conclude that the optimized value of the flow resistivity is correct. This study has been carried on, in the frequency bandwidths (25–60) Hz and (130–190) Hz and has also given good results.

The classical method[1] using continuous sound can be used only for the intermediary frequencies (it means between 1 and 3 or 4 KHz, depending of the radius of the tube). However, in this domain of frequency, we cannot use our approximation of very low frequency in which only the flow resistivity (or permeability) and porosity intervene in the propagation. The advantage of the proposed method is that the measurement can be done at very low frequency (less than 100 Hz by taking a long tube of 50 m), and thus the flow resistivity can be obtained easily. In the domain of frequency corresponding to measurement in the Kundt Tube (continuous sound), all physical parameters (porosity, tortuosity, viscous, and thermal characteristics lengths, viscous and thermal permeability, etc) intervene in the model of propagation. In this case, the inverse problem becomes very difficult to solve. In this proposed method, the use of transient signals (having a large spectrum) is possible, this is not the case for the classic Kundt tube (continuous frequency). For transient signals, the temporal approach are more appropriate, we do not need to use the Fourier transform. In our case the inverse problem is solved directly in time domain using the waveform without any transformation in Fourier domain.

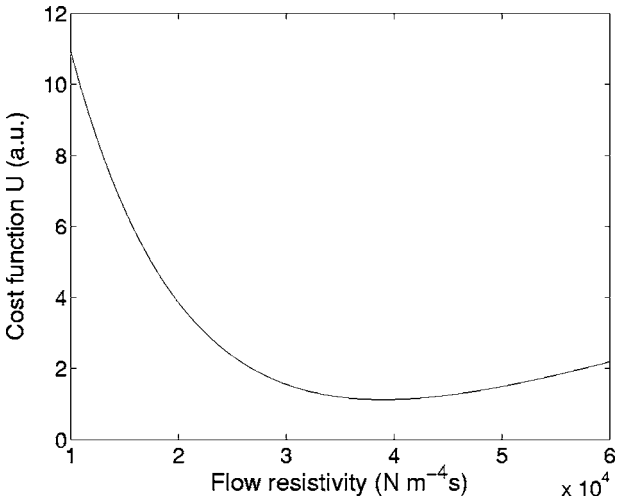


FIG. 3. Variation of the cost function U with flow resistivity σ.

[1] J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman and Hall, London, 1993).
[2] K. Attenborough, "Models for the acoustical properties of air-saturated granular media," Acta. Acust. Acust. **1**, 213 (1993).
[3] N. Chandler, "Transient streaming potential measurements on fluid-saturated porous structures: An experimental verification of Biot's slow wave in the quasi-static limit," J. Acoust. Soc. Am. **70**, 116–121 (1981).
[4] Z. E. A. Fellah and C. Depollier, "Transient acoustic wave propagation in rigid porous media: a time-domain approach," J. Acoust. Soc. Am. **107**, 683–688 (2000).
[5] M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous materials," Appl. Acoust. **3**, 105–116 (1970).
[6] D. A. Bies and C. H. Hansen, "Flow resistance information for acoustical design," Appl. Acoust. **13**, 357–391 (1980).
[7] T. F. W. Embleton, J. E. Piercy, and G. A. Daigle, "Effective flow resistivity of ground surfaces determined by acoustical measurements," J. Acoust. Soc. Am. **74**, 1239–1244 (1983).
[8] R. L. Brown and R. H. Bolt, "The measurement of flow resistance of porous acoustic materials," J. Acoust. Soc. Am. **13**, 337–344 (1942).
[9] R. W. Leonard, "Simplified flow resistance measurements," J. Acoust. Soc. Am. **17**, 240–241 (1946).
[10] M. R. Stinson and G. A. Daigle, "Electronic system for the measurement of flow resistance," J. Acoust. Soc. Am. **83**, 2422–2428 (1988).
[11] P. M. Morse, R. H. Bolt, and R. L. Brown, "Relation between acoustic impedance and low resistance of porous acoustic materials," J. Acoust. Soc. Am. **12**, 475–476 (1941) (abstract).
[12] A. Gemant, "Frictional phenomena IV," J. Appl. Phys. **12**, 725–734 (1941).
[13] M. A. Biot, "The theory of propagation of elastic waves in fluid-saturated porous solid. I. Low frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[14] 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).
[15] Z. E. A. Fellah, M. Fellah, W. Lauriks, and C. Depollier, "Direct and inverse scattering of transient acoustic waves by a slab of rigid porous material," J. Acoust. Soc. Am. **113**, 61–73 (2003).

Quantitative measurements of acoustic emissions from cavitation at the surface of a stone in response to a lithotripter shock wave (L)

Parag V. Chitnis and Robin O. Cleveland[a)]
Department of Aerospace and Mechanical Engineering, Boston University, 110 Cummington Street, Boston, Massachusetts 02215

(Received 21 September 2005; revised 20 January 2006; accepted 25 January 2006)

Measurements are presented of acoustic emissions from cavitation collapses on the surface of a synthetic kidney stone in response to shock waves (SWs) from an electrohydraulic lithotripter. A fiber optic probe hydrophone was used for pressure measurements, and passive cavitation detection was used to identify acoustic emissions from bubble collapse. At a lithotripter charging voltage of 20 kV, the focused SW incident on the stone surface resulted in a peak pressure of 43±6 MPa compared to 23±4 MPa in the free field. The focused SW incident upon the stone appeared to be enhanced due to the acoustic emissions from the forced cavitation collapse of the preexisting bubbles. The peak pressure of the acoustic emission from a bubble collapse was 34±15 MPa, that is, the same magnitude as the SWs incident on the stone. These data indicate that stresses induced by focused SWs and cavitation collapses are similar in magnitude thus likely play a similar role in stone fragmentation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2177589]

PACS number(s): 43.35.Ei, 43.25.Cb, 43.80.Gx [CCC] Pages: 1929–1932

I. INTRODUCTION

Despite the clinical success of shock wave lithotripsy (SWL) in treating kidney stones since its introduction in 1980,[1,2] there is still debate over the process by which the SWs comminute kidney stones. There are two main categories of mechanisms deemed plausible for stone breakage: direct stress effects, such as spallation,[1,3] squeezing,[4] and shearing,[5] and cavitation effects, such as micro-jets[6,7] and acoustic emissions from collapsing bubbles.[8–10]

Measurements of the stress induced in a stone by cavitation have mostly been qualitative in nature. Delacretaz *et al.*[9] performed experiments with schist stones and monitored the stress on the stone surface using a fiber-optic-based gauge. They concluded that the collapse of cavitation on the stone surface induced larger stresses than the incident shock wave although they did not report quantitative measurements of the stresses. Xi and Zhong[10] used photoelastic and shadowgraph imaging techniques to visualize the propagation of SWs and the resulting transient stress fields in the stones. They observed stresses induced in the stone due to cavitation bubble collapse but did not quantify the stresses. Estimates of lithotripsy-induced stresses in stones were also obtained by Gracewski *et al.*[5] using embedded strain gages but they did not measure stresses induced by cavitation collapse. Vogel and Lauterborn[8] studied pressures emitted by the collapse of laser-induced bubbles in the free field and near a solid boundary and measured a peak pressure of 2.5 MPa at a distance of 10 mm from the bubble.

This letter addresses the need for obtaining quantitative measurements of both the focused SWs and the cavitation-induced acoustic emissions on the stone surface. A fiber optic probe hydrophone with high spatial and temporal resolution was used for acoustic measurements. Dual passive cavitation detection was used simultaneously to identify acoustic transients resulting from cavitation collapses. Employing both measurement techniques was necessary to provide quantitative estimates of the acoustic emissions from cavitation.

II. MATERIALS AND METHODS

This study used a research lithotripter designed to mimic the behavior of the Dornier HM3 lithotripter.[11] The GALCIT lithotripter referred to in Ref. 11 is now located at Boston University and was used for these studies. Pressure measurements were carried out using a fiber optic probe hydrophone (FOPH) (Model 500, RP Acoustics, Stuttgart, Germany) as recommended by IEC standard 61846.[12] The hydrophone measures a change in refractive index of water which can be related to the pressure in the water.[13] The hydrophone has a spatial resolution of 100 μm and a bandwidth of 30 MHz. Cavitation activity was monitored using a dual passive cavitation detector (DPCD) consisting of two 1-MHz focused transducers (diameter 40 mm, focal length 64 mm, Model A392S, Panametrics, Waltham, MA) mounted confocally and orthogonally to each other. The effective spot size of the DPCD was 5 mm. Signatures corresponding to the arrival of SWs at the focus and those corresponding to the inertial collapse of the bubbles were detected using coincidence detection and then converted to pressure using a calibration scheme based on spherical spreading.[14] The result of the PCD calibration process[14] was that a 1-V PCD signal corresponded to $p_0 r_0 = 30$ MPa·mm where p_0 is the peak radiated pressure at a radius r_0 from the bubble. In these experiments, the signal from the second PCD transducer was amplified by a factor of 3 to compensate for signal loss due to obstruction by a bolt used to hold the stone in place.

a)Author to whom correspondence robinc@bu.edu

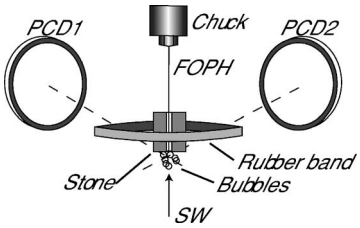


FIG. 1. Schematic of the experimental setup. The PCD transducers were confocal and orthogonal to each other. Both of the PCDs and the proximal surface of the stone were aligned at the lithotripter focus. The FOPH was inserted through the stone such that its tip was flush with the proximal surface.

The schematic of the experimental setup is shown in Fig. 1. The artificial stones were made of Ultracal 30 (U30) gypsum, were formed and stored in water, and have acoustic properties similar to natural stones.[15] The U30 stone was cylindrical in shape (6.5 mm diameter and 7.5 mm long) and a 1-mm-diam hole was drilled through the stone along its cylindrical axis. The stone was held in place by a rubber band which was attached to a frame and positioned such that one end of the stone was at the lithotripter focus. The FOPH fiber was held by a four-jaw chuck about 40 mm from the tip. The chuck was attached to a second positioning system that was used to insert the fiber through the hole of the stone until the tip was flush with the proximal surface. Prior work with a FOPH and high speed camera suggests that the FOPH moves less than 0.2 mm when a SW is incident upon it and therefore it should not interact with the sides of the hole. The DPCD was positioned so that the mutual focus coincided with the proximal face of the stone. Measurements from the FOPH and the DPCD were recorded simultaneously on a digital oscilloscope (8-bit resolution and 100-MHz sampling frequency) and downloaded to a computer for later analysis.

III. RESULTS

Figure 2 shows representative waveforms acquired simultaneously by the FOPH and DPCD at a lithotripter setting of 16 kV. The FOPH signal showed two acoustic transients of roughly equivalent amplitude at 179 and 691 μs. The DPCD signal showed the signatures associated with the

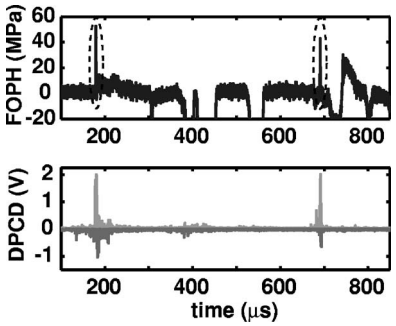


FIG. 2. Representative waveforms of the pressure measured by the FOPH (upper) and the voltages from the DPCD (lower) as a function of time. The first PCD signal is displayed as the absolute value and the second PCD signal as the negative of the absolute value which allows the signals to be distinguished clearly. The transients circled on the FOPH trace are coincident with DPCD signals confirming that the transients are associated with bubble collapses.

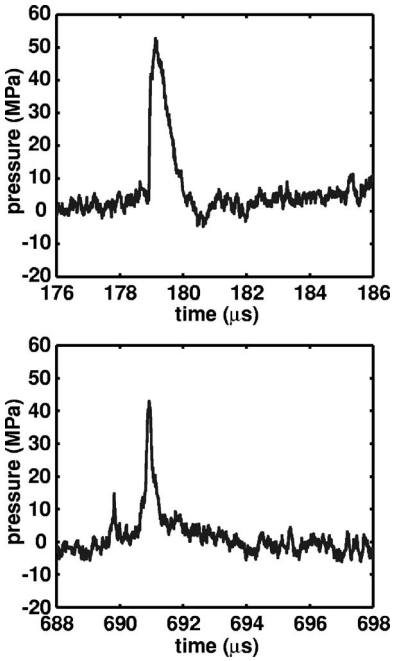


FIG. 3. FOPH pressure measurements as a function of time. Top: Focused SW; Bottom: Acoustic emission from a cavitation collapse. The cavitation-induced acoustic emission has a peak pressure and pulse width comparable to the focused SW.

arrival of the incident focused SW and the inertial collapse of the bubble. The travel time from the FOPH tip to the each PCD has been subtracted. The timing of the DPCD signal correlated well with the FOPH signal, thus confirming that the acoustic transient measured by the FOPH at 691 μs was an acoustic emission from a cavitation collapse. The FOPH signal exhibited dropout between 370 and 560 μs, which was attributed to cavitation bubbles enveloping the tip of the FOPH at which point the glass/air interface produced a large optical reflection. These observations are consistent with high speed camera images which show a bubble initially encapsulating the proximal surface of the stone and then receding before collapsing.[16] However, the signal from the FOPH returned to normal prior to the cavitation collapse which indicates that the fiber was sufficiently wetted to record pressure signals. Zoomed versions of the focused SW and the acoustic emission from the bubble collapse are shown in Fig. 3. The peak positive pressure of the acoustic emission from the bubble collapse (43 MPa) is comparable to that of the focused SW (53 MPa). The pulse width of the two waveforms are also of the same order (0.8 μs for the acoustic emission from bubble collapse and 1.2 μs for the focused SW).

The peak pressure measured with the FOPH was 43 MPa, which would be consistent with the calibration of the DPCD if a bubble collapsed at a distance r_0=0.7 mm from the FOPH. This distance is in the range of what seems reasonable. The calibration also predicts a peak pressure of 3.0 MPa at a range of 10 mm from the bubble which is consistent with the measurements of Vogel and Lauterborn.[8]

We note that simultaneous measurements using the FOPH and the DPCD were vital to accurately identify the emissions from the bubble collapse. Furthermore, since the

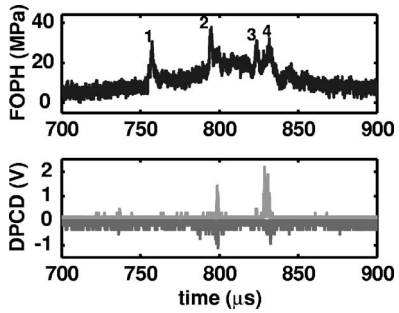


FIG. 4. Simultaneous FOPH and DPCD measurements acquired at a lithotripter driving voltage of 20 kV. Only FOPH signals 2 and 4 were coincident with DPCD signals and can be identified as signals from an inertial collapse at the stone surface.

spot size of the DPCD is about five times greater than r_0, any cavitation-related acoustic emission detected by the FOPH should also be detected by the DPCD. Figure 4 shows an example of the FOPH signal with four transients of similar amplitude, but comparison with the DPCD signal indicated that only the second and fourth emissions could have resulted from cavitation collapses in the region of interest.

Of the 20 measurements acquired at a charging voltage of 20 kV (same stone for all 20 measurements), ten definite (correlated with the DPCD) measurements of acoustic emissions from bubble collapse were obtained. Based on this data set, the focused SW incident on the stone surface had a mean peak positive pressure of 43 ± 6 MPa. The peak positive pressure for the cavitation-induced acoustic emission incident on the stone was determined to be 34 ± 15 MPa.

In the free field, the peak positive pressure of the focused SW was 23 ± 4 MPa. The plane wave reflection coefficient of the artificial stone was 0.5, which should result in an expected peak pressure at the face of 35 MPa, about 20% less than the measured pressure of 43 MPa. One possible reason for this observation is diffraction from the stone. This conjecture was examined by employing the deconvolution technique that was developed to correct for diffraction at the tip of the FOPH[17] and using it to treat the proximal surface of the stone as a finite-sized baffle. For the case of the waveform measured by the FOPH at the surface of the stone the deconvolution predicted a free field pressure of 28 MPa, which is higher than the measured free field pressure of 23 MPa. Therefore, the peak pressure of the SW incident on the stone was about 20% higher than what would be predicted from diffraction caused by the stone. The possible role of cavitation in enhancing this pressure is addressed in the discussion.

IV. DISCUSSION

Acoustic emissions due to lithotripsy-induced cavitation collapse on an artificial stone surface were identified and measured. Pressure was measured using a FOPH but only signals that a DPCD identified as acoustic emissions from cavitation collapses were considered. The peak pressure of the acoustic emission from the bubble collapse (34 ± 15 MPa) was of the same order as the focused SW incident on the stone (43 ± 6 MPa). The incident pressure on the stone was about 20% higher than predicted by reflection and diffraction at the stone surface. We speculate that the excess pressure is attributable to emissions from the forced collapse of bubbles by the incident pressure wave. This requires the presence of preexisting bubbles or nuclei at or near the stone surface. For the electrohydraulic source used here it has been shown that preexisting nuclei do exist and can be excited by the direct wave from the spark source.[18] Any preexisting bubbles could be driven to collapse by the leading compressive phase of the focused SW and the subsequent acoustic emissions could add to the incident focused SW resulting in increased peak positive pressure. The arrivals of the emissions from different bubbles will be spread out in time, thus broadening the positive peak of the focused SW. The duration of the positive phase of the focused SW in the free field was 1.5 ± 0.1 μs and that for the focused SW incident on the stone surface was 2.5 ± 1.4 μs. The longer duration of the positive phase of the SW incident on the stone is consistent with the forced collapse hypothesis. The presence of the stone also resulted in a decrease in the measured peak negative pressure: -9 ± 2 MPa in the free field and -6 ± 2 MPa on the stone surface. This too could be attributed to the acoustic emissions from collapsing bubbles located further away from the fiber than the length of the positive phase of the focused SW (approximately 2.3 mm). The transients from these bubbles would destructively interfere with the negative phase of the incident SW.

The measurements reported here are for the pressure transients generated by cavitation and do not directly include the micro-jets that could impinge on the surface.[7] However, the pressure is induced in the fluid by the water hammer effect.[19] The water hammer induces a stress in the solid and there will be a concomitant pressure in the fluid as the normal stress should be continuous across the interface. The localized pressure change will result in a change in the refractive index of the fluid which will be detected by the FOPH. We speculate that those pressure pulses detected by the FOPH which did not coincide with a PCD signature (e.g., signal 1 in Fig. 4) may be due to micro-jetting where the aspherical collapse of the bubble might produce acoustic emissions that would appear weaker to the DPCD.

In conclusion, these measurements indicate that the collapse of cavitation bubbles close to the stone surface can result in stresses in the stone similar in amplitude to the stress of the focused SWs. We infer that both acoustic emissions from cavitation and focused shock waves likely play a similar role in stone fragmentation.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health through Grant No. DK43881. We thank Dr. A. P. Evan and Dr. J. A. McAteer for providing access to the fiber optic probe hydrophone and supplying Ultracal 30 stones and Dr. R. A. Roy for helpful discussions and guidance.

[1] C. Chaussy, W. Brendel, and E. Schiemdt, "Extracorporeally induced destruction of kidney stones by shock waves," Lancet **2**, 1265–1268 (1980).
[2] M. A. Averkiou and L. A. Crum, "Cavitation: Its role in stone comminution and renal injury," in *New Developments in the Management of Urolithiasis*, edited by J. E. Lingeman and G. M. Preminger, (Igaku–Shoin, New York, 1996), pp. 21–40.

[3]D. Jocham, C. Chaussy, and E. Schmiedt, "Extracorporeal shock wave lithotripsy," Urol. Int. **41**, 357–368 (1986).
[4]W. Eisenmenger, "The mechanisms of stone fragmentation in ESWL," Ultrasound Med. Biol. **27**, 683–693 (2001).
[5]S. M. Gracewski, G. Dahake, Z. Ding, S. J. Burns, and E. C. Everbach, "Internal stress wave measurements in solids subjected to the lithotripter pulses," J. Acoust. Soc. Am. **94**, 652–661 (1993).
[6]A. J. Coleman, J. E. Saunders, L. A. Crum, and M. Dyson, "Acoustic cavitation generated by an extracorporeal shock wave lithotripter," Ultrasound Med. Biol. **13**(2), 69–76 (1990).
[7]L. A. Crum, "Cavitation micro-jets as a contributory mechanism for renal calculi disintegration in ESWL," J. Urol. (Baltimore) **140**, 1587–1590 (1988).
[8]A. Vogel and W. Lauterborn, "Acoustic transient generation by laser–produced cavitation bubbles near solid boundaries," J. Acoust. Soc. Am. **84**, 719–731 (1988).
[9]G. Delacretaz, K. Rink, G. Pittomvils, J. P. Lafaut, H. Vandeursen, and R. Boving, "Importance of the implosion of eswl-induced cavitation bubbles," Ultrasound Med. Biol. **21**(1), 97–103 (1995).
[10]X. Xi and P. Zhong, "Dynamic photoelastic study of the transient stress fields in solids during shock wave lithotripsy," J. Acoust. Soc. Am. **109**, 1226–1239 (2001).
[11]R. O. Cleveland, M. R. Bailey, N. Fineberg, B. Hartenbaum, J. A. McAteer, and B. Sturtevant, "Design and characterization of a research electrohydraulic lithotripter patterned after Dornier HM3," Rev. Sci. Instrum. **71**(6), 2514–2524 (2000).
[12]International Electrotechnical Committee, "Ultrasonics—pressure pulse lithotripters—characteristics of fields," IEC Standard 61846 (1998).
[13]J. Staudenraus and W. Eisenmenger, "Fiber-optic probe hydrophone for ultrasonic and shock–wave measurements in water," Ultrasonics **31**(4), 267–273 (1993).
[14]R. O. Cleveland, O. A. Sapozhnikov, M. R. Bailey, and L. A. Crum, "A dual passive cavitation detector for localized detection of lithotripsy-induced cavitation in vivo," J. Acoust. Soc. Am. **107**(3), 1745–1758 (2000).
[15]J. A. McAteer, J. C. Williams, Jr., R. O. Cleveland, J. Van Cauwelaert, M. R. Bailey, D. A. Lifshitz, and A. P. Evan, "Ultracal-30 gypsum artificial stones for lithotripsy research," Urol. Res. **33**, 429–434 (2005).
[16]Y. A. Pishchalnikov, O. A. Sapozhnikov, M. R. Bailey, J. C. Williams, R. O. Cleveland, T. Colonius, L. A. Crum, A. P. Evan, and J. A. McAteer, "Cavitation bubble cluster activity in he breakage of kidney stones by lithotripter shock waves," J. Endourol **17**(7), 435–446 (2003).
[17]Z. Q. Wang, P. Lauxmann, C. Wurster, M. Kohler, B. Gompf, and W. Eisenmenger, "Impulse response of a fiber optic probe hydrophone determined with shock waves in water," J. Appl. Phys. **85**(5), 2514 (1999).
[18]T. J. Matula, P. R. Hilmo, and M. R. Bailey, "A suppressor to prevent direct wave-induced cavitation in shock wave therapy devices," J. Acoust. Soc. Am. **118**, 178–185 (2005).
[19]J. E. Field, "The physics of liquid impact, shock wave interactions with cavities, and the implications to shock wave lithotripsy," Phys. Med. Biol. **36**(11), 1475–1484 (1991).

A theoretical model to predict the low-frequency sound absorption of a Helmholtz resonator array (L)

SangRyul Kim[a)]
Acoustics Laboratory, Korea Institute of Machinery and Materials, P.O.Box 101, Yuseong, Daejeon, 305-600, Korea

Yang-Hann Kim[b)]
Center for Noise and Vibration Control, Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology, Science Town, Daejeon, 305-701, Korea

Jae-Hee Jang[c)]
R&D Center, SKEC, 192-18 Kwanhun-dong, Chongro-gu, Seoul, 110-300, Korea

(Received 10 October 2005; revised 14 January 2006; accepted 20 January 2006)

A theoretical method based on mutual radiation impedance is proposed to compute the sound absorption performance of a Helmholtz resonator array in the low-frequency range. Any configuration of resonator arrangement can be allowed in the method, while all the resonators may or may not be identical. Comparisons of the theoretical predictions with those done by the past studies or experiments show that the present method can accurately predict the absorption performance in more general cases.
 [DOI: 10.1121/1.2177568]

PACS number(s): 43.55.Ev, 43.55.Dt [NX] Pages: 1933–1936

I. INTRODUCTION

An array of Helmholtz resonators is often used to reduce low-frequency noise because of its resonance characteristics.[1,2] Differently tuned resonators also have been employed to decrease broadband noise.[3] One of the popular shapes of resonator array is a plane-type array illustrated in Fig. 1(a); hereafter, we call it the Helmholtz resonator array panel (for convenience, array panel). However, it has not been well understood how and how much the panel really absorbs sound energy. For instances, Zwikker *et al.*[4] regarded the top surface of the panel as a homogeneous surface and utilized Bolt's concept[5] of spatially averaged impedance to calculate normal-incidence absorption coefficient of the surface. Allard *et al.*[6,7] presented a more generalized theory considering oblique incidence for an identical-resonator array panel, and recently Kim *et al.*[8,9] applied the mutual radiation impedance between resonator orifices to evaluating absorption coefficient of a rectangular array panel with identical resonators. But the conventional methods are valid only when the cell size is very small compared to the wavelength.

In this paper, we propose a theoretical method that considers the radiation impedance of the array panel. The panel is composed of a group of resonators which may or may not be identical. Any size of the panel or cell can be allowed in the method, and there is no restriction on the resonator disposition. For numerical examples, we consider array panels with identical and differently tuned resonators and predict their absorption coefficients, which are also compared with experimental results.

[a)]Electronic mail: srkim@kimm.re.kr
[b)]Electronic mail: yanghannkim@kaist.ac.kr
[c)]Electronic mail: jhjang@skec.co.kr

II. THEORETICAL ANALYSIS

A. Description of system

As shown in Fig. 1(b), we consider an infinite array panel that is periodically arranged with a rectangular cell containing a group of resonators, which are not necessarily identical. The size of the cell is L_x by L_y, and its top surface is assumed to be acoustically rigid, except for the circular holes that correspond to the entrances of respective resonators. We can place the resonators or orifices at arbitrary positions within the cell area. Here we assume that the radii of orifices are very small compared to the wavelength of interest and that the air layer in the opening moves with uniform velocity so that it can be regarded as a plane piston.

B. Absorption coefficient of Helmholtz resonator array panel

When the top surface of the panel is positioned in the xy plane, a plane wave p_e incident on the plane [see Fig. 1(b)] can be written as

$$P_e = P_e e^{-j(k_x x + k_y y - k_z z)}, \tag{1}$$

The time factor $e^{j\omega t}$ is suppressed for simplicity. Using the Green's function (G_N) for a semi-infinite space bounded by a rigid plane, we can describe the sound pressure p on the xy plane in terms of the Kirchhoff-Helmholtz integral equation (for example, see Ref. 10). That is,

$$p = 2p_e + jk\rho c \sum_{i=-\infty}^{\infty} u_i \int_{S_i} G_N dS_i \quad \text{with } G_N = \frac{e^{-jkR}}{2\pi R}. \tag{2}$$

The subscript i in Eq. (2) represents the index of orifice, S_i is the orifice area, u_i is the orifice velocity, counted positive into the external space (z direction), ρ is the density of air, c

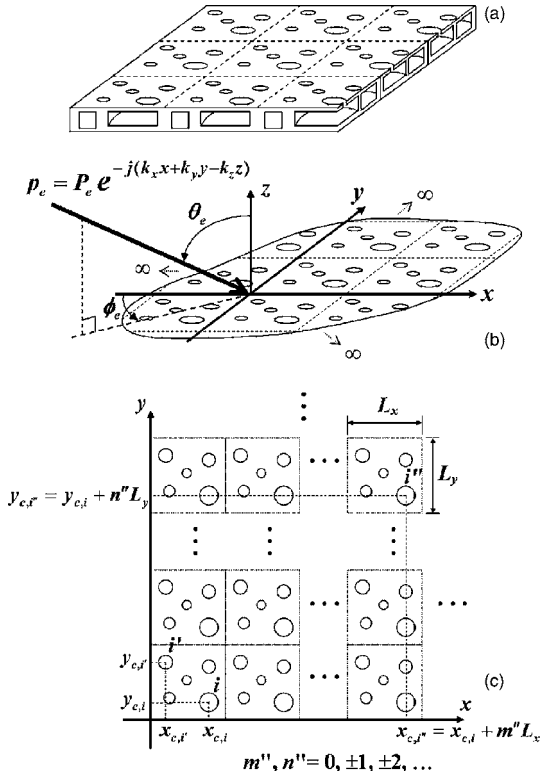


FIG. 1. Helmholtz resonator array panel and its theoretical model. (a) Helmholtz resonator array panel; (b) plane wave p_e incident on the theoretical model ($k_x = k \sin\theta_e \cos\phi_e$, $k_y = k \sin\theta_e \sin\phi_e$, $k_z = k \cos\theta_e$, and k is the wave number); and (c) xy coordinates of orifice centers. For a theoretical model, the top surface of the panel is regarded as an infinite plane periodically arranged with a rectangular cell, which corresponds to a part bordered by dotted lines in this figure.

is the speed of sound, and $R = \sqrt{(x-x_0)^2 + (y-y_0)^2}$.

If we average both sides of Eq. (2) over the area of orifice i' ($S_{i'}$), the left-hand side and the first term of the right-hand side become the pressures at the center of the orifice, denoted by $p_{i'}$ and $2p_{e,i'}$, respectively, since the radius of orifice is assumed to be very small compared with the wavelength of interest. In this case, the second term can be replaced by the mutual radiation impedance[11,12] $\rho c \zeta_{i'i}$ between orifices i' and i, that is,

$$\rho c \zeta_{i'i} = \frac{jk\rho c}{S_{i'}} \int_{S_{i'}} \int_{S_i} G_N dS_i dS_{i'}. \tag{3}$$

Implying the boundary condition on the orifice i', $p_{i'} = -\rho c \zeta_{i'} u_{i'}$, where $\rho c \zeta_{i'}$ is the surface impedance on the orifice, we can obtain an infinite set of simultaneous equations

$$\sum_{i=-\infty}^{\infty} (\zeta_{i'i} + \delta_{i'i}\zeta_i)\rho c u_i = -2p_{e,i'} \quad \text{for } i' = 0, \pm 1, \ldots, \pm\infty. \tag{4}$$

where $\delta_{i'i} = 1$ for $i' = i$ and $\delta_{i'i} = 0$ for $i' \neq i$.

Consider two identical orifices i and i'' that are located at the same position in their respective cells [see Fig. 1(c)]: not only the radii of orifices but also their surface impedances are the same. Since the surface of the panel is periodic and infinite, the acoustic behaviors in front of the orifices i and i'' must be the same except for an excitation by the incident plane wave (for example, see Ref. 13). Thus the ratio between orifice velocities can be expressed as $u_i/u_{i''} = p_{e,i}/p_{e,i''}$. Substituting the ratio into Eq. (4) and rearranging the result lead to a finite set of simultaneous equations:

$$\sum_{i=1}^{N} (\bar{\zeta}_{i'i} + \delta_{i'i}\zeta_i)\rho c u_i = -2p_{e,i'} \quad \text{for } i' = 1,2, \ldots, N, \tag{5}$$

where N is the number of orifices in a cell, and

$$\bar{\zeta}_{i'i} = \sum_{\substack{m'=-\infty \\ i''=i+m'N}}^{\infty} \zeta_{i'i''} \frac{p_{e,i''}}{p_{e,i}}. \tag{6}$$

The final expression of $\bar{\zeta}_{i'i}$ is derived in the Appendix .

The absorption coefficient of the array panel can be calculated by the ratio of the absorbed effective power to the incident sound power. Because the total power absorbed on a cell area is equal to the sum of power dissipated by individual resonators and the incident power on a cell is $(L_x L_y/2\rho c)|P_e|^2 \cos\theta_e$, the absorption coefficient α can be expressed as

$$\alpha = \sum_{i=1}^{N} \frac{\pi a_i^2}{L_x L_y} \frac{\mathrm{Re}(\zeta_i)}{\cos\theta_e} \left| \frac{\rho c u_i}{P_e} \right|^2, \tag{7}$$

where a_i is the radius of orifice. If the surface impedances $\rho c \zeta_i$ are determined, then the absorption coefficient can be obtained by solving Eq. (5) for the unknown velocity u_i and by substituting them into Eq. (7).

III. EXPERIMENTAL VERIFICATIONS

A. Experimental setup to measure the absorption coefficient

In order to examine the accuracy of the proposed method, the normal-incidence absorption coefficients of several array panels were measured by using the transfer-function method[14] with the instrumentations that are illustrated in Fig. 2(a).

Figure 2(b) shows the dimensions of two test samples made of a PVC plate. Each specimen can be regarded as a part of a corresponding array-panel with an infinite size because the four walls of the rectangular tube are rigid and a plane wave is perpendicularly incident on the surface of the specimen.

B. Theoretical model of surface impedance

From the previous study,[15,16] the surface impedance ($\rho c \zeta$) on the opening of the resonator that has a cylindrical cavity with a concentric neck of cylindrical type can be written as

$$\zeta = k\delta_\nu(2 + l/a) + j\left[k\{l + \delta_{in} + \delta_\nu(2 + l/a)\} - \frac{\pi a^2}{\pi a_c^2} \cot kl_c \right], \tag{8}$$

where $\delta_\nu = \sqrt{2\nu/\omega}$, ν is the kinematical viscosity of air ($\nu \approx 15 \times 10^{-6}$ m^2/s), ω is the angular frequency, δ_{in} is the internal end correction of resonator, a is the inner radius

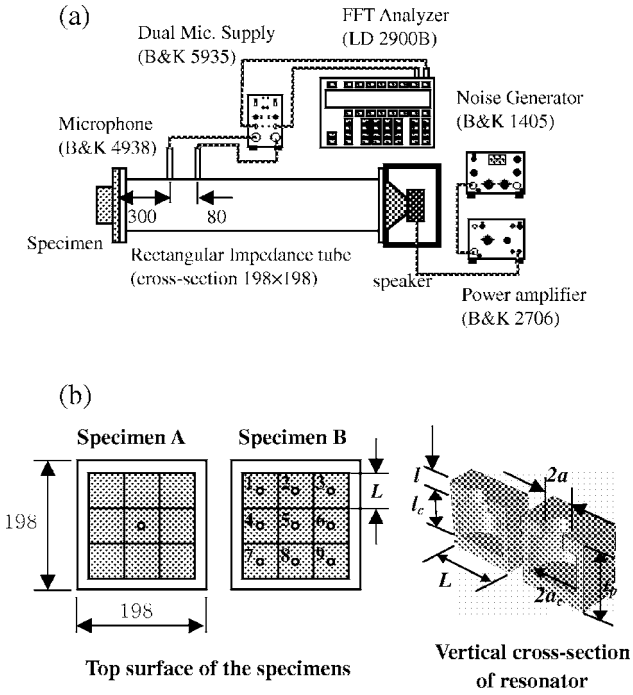


FIG. 2. (a) Experimental setup and (b) specimens to measure the normal-incidence absorption coefficient. All the dimensions of the resonators except the cavity depth l_c are the same [$a=l=5$, $a_c=25$, $t_p=110$, $L=60$ (unit: mm)]. In the case of specimen A, $l_c=64$ mm and the cavity depths of nine resonators in the specimen B are 100, 91, 82, 73, 64, 55, 46, 37, and 28 mm, respectively.

of neck, l is the neck length, and a_c and l_c represent the radius of air cavity and its depth, respectively. For the internal end correction, there have been numerous studies, but, for simplicity, we employ the approximated expression suggested by Selamet *et al.*,[17]

$$\delta_{in} = 0.82a(1 - 1.33a/a_c) \quad \text{for } a/a_c < 0.4. \tag{9}$$

C. Theoretical and experimental results

First of all, we consider an array panel that has one resonator in a cell: we will call such a panel an '$N=1$' array panel. In this case, the absorption coefficient can be obtained from Eqs. (5) and (7) as

$$\alpha = \frac{\pi a_1^2}{L_x L_y}\frac{1}{\cos\theta_e}\frac{4\,\mathrm{Re}(\zeta_1)}{|\bar{\zeta}_{11}+\zeta_1|^2}, \tag{10}$$

and $\bar{\zeta}_{11}$ represents the net radiation impedance[11,12] $\zeta_{\mathrm{rad},1}$ on the orifice:

$$\zeta_{\mathrm{rad},1} = \bar{\zeta}_{11} = \frac{\pi a_1^2}{L_x L_y}\cdot\frac{1}{\cos\theta_e} + \frac{\pi a_1^2}{L_x L_y}\sum_{\substack{m=-\infty \\ m,n\neq 0,0}}^{\infty}\sum_{n=-\infty}^{\infty}\frac{jk}{\sqrt{\beta_{mn}^2-k^2}}\left(\frac{2J_1(a_1\beta_{mn})}{a_1\beta_{mn}}\right)^2, \tag{11}$$

where $\beta_{mn}=\sqrt{(2\pi m/L_x-k_x)^2+(2\pi n/L_y-k_y)^2}$. If all β_{mn} except β_{00} are larger than k, i.e.,

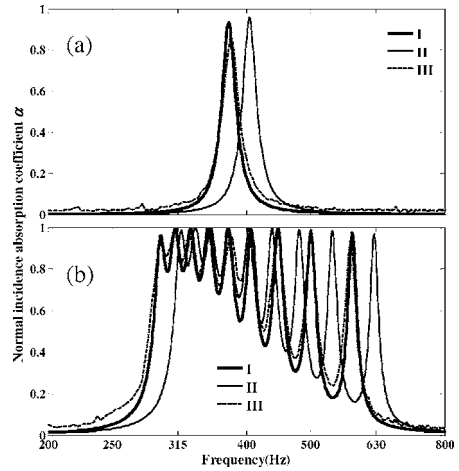


FIG. 3. Comparison of predicted and measured absorption coefficients for the specimens (a) A and (b) B shown in Fig. 2(b). Theoretical predictions are obtained by using the proposed method (graph I) and the conventional one (graph II); graph III represents the experimental result.

$$(m\lambda/L_x - \sin\theta_e\cos\phi_e)^2 + (n\lambda/L_y - \sin\theta_e\sin\phi_e)^2 > 1 \quad \text{for } (m,n)\neq(0,0), \tag{12}$$

where $\lambda=2\pi/k$ is the wavelength, the double summation term of Eq. (11) becomes purely imaginary, and so the use of external end correction $\delta_{ex,1}$, which equals the imaginary part of net radiation impedance divided by k, gives

$$\alpha = \frac{4\mathrm{Re}(\zeta_1)\cos\theta_e/\epsilon_1}{|1+(jk\delta_{ex,1}+\zeta_1)\cos\theta_e/\epsilon_1|^2} \quad \text{with } \epsilon_1 = \frac{\pi a_1^2}{L_x L_y}. \tag{13}$$

The above equation is identical to those used in the past studies[4,7,8] except that in the present method, $\delta_{ex,1}$ can be calculated for all angles of incidence. It is also found that in the case of $L_x=L_y$, the proposed formula of $\delta_{ex,1}$ becomes equal to that done by Allard[7] at the special incidence angle that he considered. It is noteworthy that the expressions such as Eq. (13) can be adopted for all angles only if $\max(L_x,L_y)/\lambda<0.5$ because the allowed range of L_x and L_y in Eq. (12) varies with the angle of incidence.

Figure 3(a) illustrates the comparison between the theoretical and measured results for an $N=1$ array panel [specimen A in Fig. 2(b)], whose cell area is equal to the cross-section area of the impedance tube ($L_x=L_y=198$ mm). The theoretical results in the figure were obtained by using respectively the proposed method and the conventional one[4] based on the spatially averaged impedance and $\delta_{ex,1}=8a/3\pi$: the value $8a/3\pi$, which is widely used in many applications, is an approximate external-end-correction of a single resonator with infinite baffle. As mentioned above, Allard's formula is equal to the proposed one in this case. Figure 3(b) shows the results for the specimen B in Fig. 2(b). The corresponding array panel has 36 resonators (nine differently tuned resonators) within a cell, whose size is $L_x=L_y=396$ mm because the period of the panel become doubled to take into account the mirror imaging effect of the side wall; Allard *et al.*[6,7] did not present any way to deal with such a panel. The comparisons shown in Figs. 3(a) and 3(b) represent that the theoretical prediction agrees better with the measurement in the case of the proposed method than the

other theory,[4] particularly in respect to the resonance frequency at which the maximum absorption occurs. These results certainly confirm that the model, which we have proposed, is capable of predicting the absorption coefficients of array panels with more various resonator arrangment.

IV. CONCLUSIONS

We have presented a method that can compute absorption coefficients of Helmholtz resonator array panels. Any size cell can be allowed in the method, and the number of resonators and their disposition in the cell can be also selected without restriction. Comparisons of the theoretical predictions with the measurements have confirmed that the present method can more accurately predict the absorption coefficient in general cases than the conventional method does.

ACKNOWLEDGMENTS

This study was partly supported by SK Engineering & Construction Co., Ltd. and the Ministry of Education and Human Resources Development of Korea (Brain Korea 21 project).

APPENDIX: DERIVATION OF RADIATION IMPEDANCE ON ORIFICE

In Eq. (3), the use of spatial Fourier transform of the Green's function G_N (for example, see Ref. 18), and Bessel's integral[19] leads to

$$\zeta_{i'i} = \frac{a_i}{a_{i'}} \int_0^\infty \frac{jk}{\sqrt{\beta^2 - k^2}} \cdot \frac{2J_1(a_{i'}\beta)J_1(a_i\beta)J_0(R_{i'i}\beta)}{\beta} d\beta, \quad \text{(A1)}$$

where $R_{i'i} = \sqrt{(x_{c,i'} - x_{c,i})^2 + (y_{c,i'} - y_{c,i})^2}$, J_n represents the Bessel function of the first kind of order n, and x_c and y_c are respectively the x and y coordinates at the center of orifice area. Substituting Eqs. (1) and (A1) and the identities of $a_{i''} = a_i$, $x_{c,i''} - x_{c,i'} = m''L_x$, and $y_{c,i''} - y_{c,i'} = n''L_y$ [see Fig. 1(c)] into Eq. (6) gives

$$\bar{\zeta}_{i'i} = \frac{a_i}{a_{i'}} \int_0^\infty d\beta \cdot \frac{jk}{\sqrt{\beta^2 - k^2}} \cdot \frac{2J_1(a_{i'}\beta)J_1(a_i\beta)}{\beta} \cdot \sum_{m''=-\infty}^{\infty} \sum_{n''=-\infty}^{\infty} J_0(R_{i'i}\beta) \cdot e^{-j(m''k_xL_x + n''k_yL_y)}. \quad \text{(A2)}$$

Here, the double summation can be rewritten from Fourier series expansion as

$$\sum_{m''=-\infty}^{\infty} \sum_{n''=-\infty}^{\infty} J_0(R_{i'i''}\beta) e^{-j(m''k_xL_x + n''k_yL_y)} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{2\pi}{L_xL_y} e^{j\{(2\pi m/L_x - k_x)(x_{c,i'} - x_{c,i}) + (2\pi n/L_y - k_y)(y_{c,i'} - y_{c,i})\}} \times \int_0^\infty J_0(R\beta)J_0(R\beta_{mn})RdR \quad \text{(A3)}$$

where $\beta_{mn} = \sqrt{(2\pi m/L_x - k_x)^2 + (2\pi n/L_y - k_y)^2}$, and so the utilization of Fourier-Bessel integral[20] produces the final expression of $\bar{\zeta}_{i'i}$

$$\bar{\zeta}_{i'i} = \frac{\pi a_i^2}{L_xL_y} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{jk}{\sqrt{\beta_{mn}^2 - k^2}} \cdot \frac{2J_1(a_{i'}\beta_{mn})}{a_{i'}\beta_{mn}} \cdot \frac{2J_1(a_i\beta_{mn})}{a_i\beta_{mn}} \cdot e^{j\{(2\pi m/L_x - k_x)(x_{c,i'} - x_{c,i}) + (2\pi n/L_y - k_y)(y_{c,i'} - y_{c,i})\}}. \quad \text{(A4)}$$

[1] J. M. Garrelick, "The transmission loss of a wall incorporating an array of resonators," J. Acoust. Soc. Am. **65**, S52 (1979).

[2] R. A. Prydz, L. S. Wirt, and H. L. Kuntz, "Transmission loss of a multilayer panel with internal tuned Helmholtz resonators," J. Acoust. Soc. Am. **87**, 1597–1602 (1990).

[3] S.-H. Seo and Y.-H. Kim, "Silencer design by using array resonators for low-frequency band noise reduction," J. Acoust. Soc. Am. **118**, 2332–2338 (2005).

[4] C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, New York, 1949), Chap. 7, 157–163.

[5] R. H. Bolt, "On the design of perforated facings for acoustic materials," J. Acoust. Soc. Am. **19**, 917–921 (1947).

[6] P. Guignouard, M. Meisser, J. F. Allard, P. Rebillard, and C. Depollier, "Prediction and measurement of the acoustic impedance and absorption coefficient at oblique incidence of porous layers with perforated facings," Noise Control Eng. J. **36**, 129–135 (1991).

[7] J. F. Allard, *Propagation of Sound in Porous Media* (Elsevier, London, 1993), Chap. 10.

[8] S. R. Kim, M.-S. Kim, Y.-H. Kim, and Y.-W. Kim, "*Absorptive characteristics of resonator panel for low frequency noise control, and its application*," on the CD-ROM Seogwipo, Korea, August 25–28, 32nd International Congress and Exposition on Noise Control Engineering (ISBN 89-952189-1-6 98060, Proceedings of Inter-Noise 2003), paper N932.

[9] M.-S. Kim, "Absorption characteristics of resonator array panel or low frequency applications," M.S. dissertation, Mechanical Engineering Department, Korea Advanced Institute of Science and Technology, 2003.

[10] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chap. 7, pp. 804–806.

[11] R. L. Pritchard, "Mutual Acoustic impedance between radiators in an infinite rigid plane," J. Acoust. Soc. Am. **32**, 730–737 (1960).

[12] E. Skudrzyk, *The Foundations of Acoustics: Basic Mathematics and Basic Acoustics* (Springer-Verlag, Wien, 1971), Chap. XXVIII, pp. 663–676.

[13] Y. W. Lam, "A boundary integral formulation for the prediction of acoustic scattering from periodic structures," J. Acoust. Soc. Am. **105**, 762–769 (1999).

[14] ISO 10534-2: 1998 Acoustics - Determination of Sound Absorption Coefficient and Impedance in Impedance Tubes - Part 2: Transfer-Function Method First Edition.

[15] U. Ingard, "On the theory and design of acoustic resonators," J. Acoust. Soc. Am. **25**, 1037–1061 (1953).

[16] L. L. Beranek and I. L. Ver, *Noise and Vibration Control Engineering, Principle and Application* (Wiley, New York, 1992), Chap. 8, pp. 232–234.

[17] A. Selamet and Z. L. Ji, "Circular asymmetric Helmholtz resonators," J. Acoust. Soc. Am. **107**, 2360–2369 (2000).

[18] E. Skudrzyk, *The Foundations of Acoustics: Basic Mathematics and Basic Acoustics* (Springer-Verlag, Wien, 1971), Chap. XXVII, pp. 659–660.

[19] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge U.P., London, 1966), Chap. II, pp. 20–22.

[20] G. N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge U.P., London, 1966), Chap. XIV, pp. 453–453.

Automated pure-tone threshold estimations from extrapolated distortion product otoacoustic emission (DPOAE) input/output functions (L)[a]

Nicolas Schmuziger,[b),c)] Jochen Patscheke,[b)] and Rudolf Probst
Department of Otorhinolaryngology, University Hospital, CH-4031 Basel, Switzerland

(Received 27 November 2005; revised 4 February 2006; accepted 6 February 2006)

A promising approach to the prediction of pure-tone thresholds through the estimation of DPOAE thresholds by input/output functions was recently published by Boege and Janssen [J. Acoust. Soc. Am. **111**, 1810–1818 (2002)]. On the basis of their results, a device that enables automated measurements of these thresholds was recently developed. The purpose of the current study was to evaluate the reliability of this instrument for the objective assessment of hearing loss in 101 ears with either normal hearing or with cochlear hearing loss of up to 50 dB HL. The median difference between pure-tone hearing and DPOAE thresholds was approximately 2 dB. For individual subjects, however, DPOAE thresholds differed from pure-tone thresholds by up to 40 dB. We find, therefore, that the clinical benefits of this method are probably limited. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2180531]

PACS number(s): 43.64.Jb [BLM] Pages: 1937–1939

I. INTRODUCTION

Measurements of DPOAEs have been used to make dichotomous decisions, in which an ear is classified as having either normal to nearly normal hearing or as showing evidence of hearing loss. Several studies have attempted to move beyond this two-state classification scheme and actually attempted to predict pure-tone thresholds from DPOAE measurements (Gorga *et al.*, 1996; Martin *et al.*, 1990). One of the most promising approaches in this regard was recently developed by Janssen and co-workers using DPOAE-threshold estimation by extrapolating DPOAE input/output (I/O) functions with sophisticated parameter settings (Boege and Janssen, 2002). In their study, Boege and Janssen found that the majority of measurements showed a good relationship between DPOAE and pure-tone hearing thresholds in both ears with normal hearing and with mild to moderate cochlear hearing loss. The authors concluded that DPOAE measurements represent a reliable estimation of cochlear hearing thresholds up to a moderate cochlear hearing loss of 50 dB HL. However, for a significant minority of their measurements, pure-tone thresholds were poorly predicted by DPOAE input/output functions.

On the basis of the findings of Boege and Janssen, a device for measuring automated pure-tone threshold estimation by means of such DPOAE I/O functions, the Cochlea-scan, was recently developed by Fischer-Zoth, Germany. According to the manufacturer, this fast, easy to handle, and portable instrument provides "extended hearing screening" with both frequency-specific and quantitative information on hearing loss. It has also been proclaimed to be suitable for the fast evaluation of hearing function in any age group and monitoring of noise-exposed subjects. The aim of our present report was to evaluate the suitability of this device as a clinical tool for the objective assessment of hearing loss.

II. METHODS

A. Subjects

Subjects were classified into two groups. The first of these included 101 ears from 53 subjects (43 males; mean age 34 years, range 21–69 years) with either normal hearing or mild to moderate cochlear hearing loss (pure tone thresholds $\leqslant$50 dB HL from 1.5 to 4 kHz). Forty-two of these subjects were nonprofessional pop/rock band members (i.e., the main income of each subject had to be earned from "nonmusical" activities) who had been active in music for more than 5 years and had weekly exposures to intense sound levels by electro-amplification for at least 2 h. Our exclusion criteria were determined using the *Questionnaire for Hearing Tests* (ISO/TC43/WG1, 1996): (a) the occurrence of acoustic trauma, (b) excessive noise exposure during occupational activities, (c) a history of recurrent otitis media, (d) ear surgery, (e) fractures of the cranium, (f) ingestion of potentially ototoxic drugs, and (g) reported hearing difficulties in other family members.

The second group (control group) included 10 ears from 5 subjects (4 males; mean age 32 years, range 1–57 years) with documented profound hearing loss or deafness.

B. Pure-tone threshold measurements

Pure-tone air conduction thresholds at all standard frequencies from 0.25 to 8 kHz, including interoctave frequencies of 1.5, 3, and 6 kHz, were measured with a digital,

[a)]Neither the authors of this study, nor the departments of otorhinolaryngology from the University Hospital of Basel and Cantonal Hospital of Aarau, received any financial support from industry or any other external source related to the material discussed herein.
[b)]Current affiliation: Department of Otorhinolaryngology, Cantonal Hospital, CH-5000 Aarau, Switzerland.
[c)]Elecronic mail: nicolas.schmuziger@ksa.ch

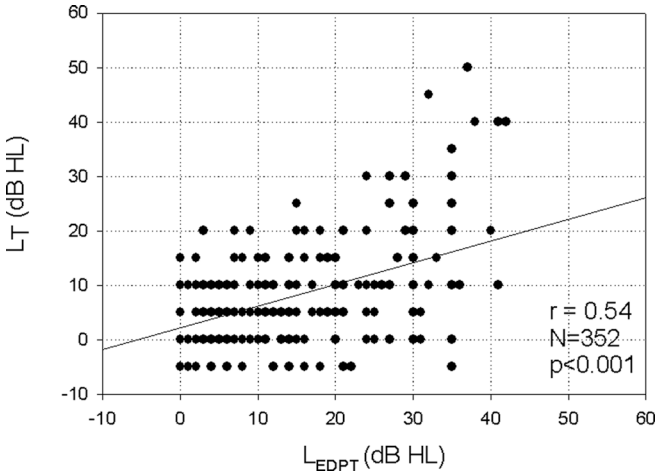


FIG. 1. Pure-tone thresholds (L_T) plotted against the estimated DPOAE thresholds (L_{EDPT}) for all frequencies.

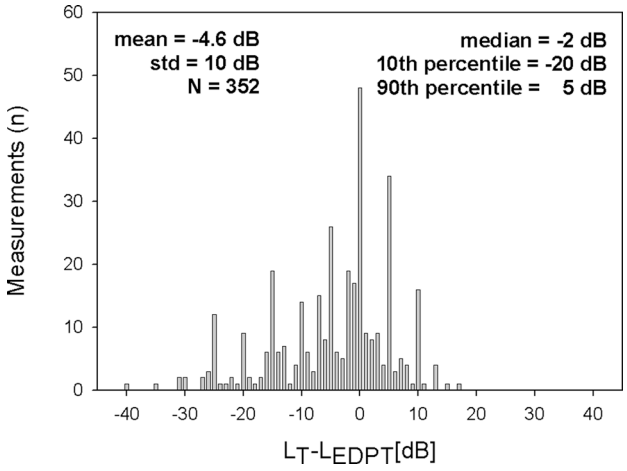


FIG. 2. Differences between pure-tone thresholds and estimated DPOAE thresholds ($L_T - L_{EDPT}$) for all frequencies.

PC-controlled audiometer (Insider of Audiocare, Switzerland) equipped with circumaural Sennheiser HDA 200 earphones. These measurements were obtained using a modified Hughson-Westlake procedure and were performed in a sound-treated booth, in which the ambient noise level was less than that recommended by ISO 8253-1 (1989). Conductive hearing loss was excluded by screening immittance measurements and otoscopy. The audiometer and earphones were calibrated according to the manufacturer's instructions.

C. DPOAE measurements

Automated pure-tone threshold estimations, by means of DPOAE I/O functions, were performed using the Cochlea-scan device under study according to the manufacturer's instructions. DPOAE I/O functions were measured at 1.5, 2, 3, 4, and 5 kHz for f_2. Primary stimulus levels for L_2 were 65 dB SPL and were reduced in 5-dB steps down to 15 dB SPL. The primary tone level for L_1 was $0.4^{*}L_2+39$ dB (Boege and Janssen, 2002). The estimated thresholds were displayed by the Cochlea-scan device.

D. Data analysis

Estimated thresholds from extrapolated DPOAE I/O functions were compared with the corresponding "behavioral" pure-tone thresholds for single standard frequencies at 1.5, 2, 3, and 4 kHz. Pure-tone thresholds at 5 kHz were not measured in this study because we used a commercially available audiometer with standard parameter settings.

III. RESULTS

A. Ears with normal hearing or mild cochlear hearing loss

From the 101 ears with normal hearing or mild cochlear hearing loss, a total of 402 DPOAE I/O-function measurements were performed. Three hundred fifty-two of these (88%) could be processed by the Cochlea-scan system for the estimation of pure-tone thresholds. Figure 1 shows the pure-tone thresholds (L_T) plotted against the estimated DPOAE thresholds (L_{EDPT}) for all frequencies in the 101 ears. A moderate but significant relationship between the DPOAE thresholds and pure-tone thresholds is demonstrated by this linear regression, with a calculated slope of almost 0.5 and a correlation coefficient (r) equal to 0.54. A good relationship was also shown for measurements with pure-tone thresholds $\geqslant$30 dB HL in comparison to the estimated DPOAE thresholds, whereas the relationships between measurements with pure-tone thresholds of $\leqslant$20 dB HL and the corresponding estimated DPOAE thresholds were clearly poorer. For example, the estimated DPOAE thresholds ranged from 0 to 42 dB HL for a given pure-tone threshold of 10 dB HL. The distribution of the differences between the pure-tone thresholds and estimated DPOAE thresholds are shown in Fig. 2. The median difference was −2 dB, the range between 10th and 90th percentile 25 dB. To allow comparison with the results in the literature, the mean value and standard deviation are also indicated in spite of a sleeved distribution of these differences. The distribution of the differences between the pure-tone and the estimated DPOAE thresholds is shown in Table I for the tested frequencies. The mean values and standard deviation were the lowest at 4 kHz in comparison to lower frequencies, consistent with previous reports (Gorga *et al.*, 2003).

Threshold estimations by means of DPOAE I/O functions could not be accomplished for approximately 12% of the tests. The mean measuring time for one ear was found to be 760 s (s.d. 214, range 137–999).

TABLE I. Differences at single frequencies between the pure-tone thresholds and estimated DPOAE thresholds ($L_T - L_{EDPT}$) for 101 ears with either normal hearing or mild to moderate cochlear hearing loss.

	Frequency (kHz)				
	1.5	2	3	4	1.5–4
Median (dB)	−7.5	−5	−1	0	−2
Mean (dB)	−9.5	−5.2	−3.9	0.5	−4.6
SD (dB)	9.9	10.1	10.4	6.5	10
L_T (n)	100	101	100	101	402
L_{EDPT} (n)	92	92	84	84	352

B. Ears with deafness or profound hearing loss (control group)

DPOAE thresholds could not be determined for any of these ears ($n=10$) with severe hearing loss. The mean measuring time for each ear was 215 s (s.d. 98, range 114–420).

IV. DISCUSSION

Our results showed median differences of 2 dB between the pure-tone and DPOAE thresholds for ears with either normal hearing or moderate cochlear hearing loss, which is excellent for clinical use. A number of our findings, however, must be considered as limitations to the clinical use of the Cochlea-scan. First, threshold differences of up to 40 dB were detected in individual cases, as shown in Fig. 1, and our analysis demonstrated a 10-dB standard deviation for these calculations. These results were consistent with the results of Gorga *et al.* (2003). Second, pure-tone estimation by means of DPOAE I/O functions could not be accomplished in 50 of the 402 measurements (12%), which is consistent with previous reports (Boege and Janssen, 2002; Gorga *et al.*, 2003). Third, the measuring times for DPOAE I/O functions is significantly longer in comparison to the measuring times for routine DPOAE measurements (Meier *et al.*, 2004).

V. CONCLUSION

Prediction of pure-tone thresholds through estimation of DPOAE thresholds by input/output function using the Cochlea-scan device was excellent for group means, but poor for a significant portion of the individual subjects. This is consistent with the results of both Boege and Janssen (2002) and Gorga *et al.* (2003). The clinical benefit of threshold estimation by means of extrapolated DPOAE I/O functions is therefore probably limited.

Boege, P., and Janssen, T. (**2002**). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears," J. Acoust. Soc. Am. **111**, 1810–1818.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (**2003**). "Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions," J. Acoust. Soc. Am. **113**, 3275–3284.

Gorga, M. P., Stover, L., Neely, S. T., and Montoya, D. (**1996**). "The use of cumulative distributions to determine critical values and levels of confidence for clinical distortion product otoacoustic emission measurements," J. Acoust. Soc. Am. **100**, 968–977.

International Organization for Standardization. (1989). "Acoustics. Audiometric test methods—Part 1: Basic pure tone air and bone conduction threshold audiometry," ISO 8253-1, **1989** (ISO, Geneva).

ISO/TC43/WG1. (**1996**). "Threshold of hearing:Preferred test conditions for determining hearing thresholds for standardization." Scand. Audiol. **25**, 45–52.

Martin, G. K., Ohlms, L. A., Franklin, D. J., Harris, F. P., and Lonsbury-Martin, B. L. (**1990**). "Distortion product emissions in humans. III. Influence of sensorineural hearing loss," Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 30–42.

Meier, S., Narabayashi, O., Probst, R., and Schmuziger, N. (**2004**). "Comparison of currently available devices designed for newborn hearing screening using automated auditory brainstem and/or otoacoustic emission measurements," Int. J. Pediatr. Otorhinolaryngol. **68**, 927–934.

Cochlear compression wave: An implication of the Allen-Fahey experiment (L)

Tianying Ren[a] and Alfred L. Nuttall
Oregon Hearing Research Center, Department of Otolaryngology and Head & Neck Surgery, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, NRC04, Portland, Oregon 97239-3098

(Received 19 October 2005; revised 23 January 2006; accepted 24 January 2006)

In order to measure the gain of the cochlear amplifier, de Boer and co-workers recently extended the Allen-Fahey experiment by measuring otoacoustic emissions and basilar membrane vibration [J. Acoust. Soc. Am. **117**, 1260–1266 (2005)]. Although this new experiment overcame the limitation of the original Allen-Fahey experiment for using a low-frequency ratio, it confirmed the previous finding that there is no detectable cochlear amplification. This result was attributed to destructive interference of the otoacoustic emission over its generation site. The present letter provides an alternative interpretation of the results of the Allen-Fahey experiment based on the cochlear fluid compression-wave theory. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2177586]

PACS number(s): 43.64.Jb, 43.64.Kc, 43.64.Ri, 43.64.Tk [BLM] Pages: 1940–1942

For processing a variety of environmental sounds, the mammalian cochlea has remarkable sensitivity with exquisite frequency selectivity and time resolution (Robles and Ruggero, 2001). It is believed that an active amplification mechanism, called "the cochlear amplifier," uses metabolic energy to boost cochlear partition vibration in response to soft sound (Davis, 1983). However, despite more than 20 years of intensive study, the cochlear amplifier theory remains to be tested.

An ingenious experiment was designed and conducted by Allen and Fahey (1992) to estimate the cochlear amplifier gain by comparing the otoacoustic emission (Kemp, 1986) to the auditory nerve fiber response. Allen and Fahey assumed that the distortion product (DP) otoacoustic emission propagates from its generation site in either direction along the cochlear partition as traveling waves. The forward traveling wave reaches its characteristic frequency (CF) location and the backward traveling wave reaches the stapes. Either the forward or the backward traveling wave is amplified when passing through the amplifier location. When the DP response at the CF site is kept constant, the emission in the ear canal changes as the square of the cochlear amplifier gain, while the DP generation site is moved from the basal side to the apical side of the amplification location (Fig. 4, Allen and Fahey, 1992).

Allen and Fahey conducted their experiment in anesthetized cats. A pair of acoustic tones at frequencies f1 and f2 ($f2 > f1$) was presented to the ear canal to evoke the cubic DP at frequency 2f1-f2. To monitor the DP response at its CF site, 2f1-f2 was chosen equal to the CF of the recorded auditory nerve fiber. The DP generation site was moved by varying f1 and f2 frequencies while the DP frequency was kept constant. The intensities of tones f1 and f2 were varied to achieve a constant DP response at its CF site, while the DP emission at 2f1-f2 was measured in the ear canal using a sensitive microphone.

Allen and Fahey found that the cochlear amplifier gain was less than 10 dB, and they believed that their results indicate that there is no cochlear amplifier. This result, however, has been questioned because of the limited f2/f1 ratio they could use (de Boer *et al.*, 2005). When the ratio of f2/f1 approaches 1, f1, f2, and 2f1-f2 frequencies are close together, and their traveling waves are largely overlapped on the basilar membrane (BM). Under this condition, the f1 rather than the 2f1-f2 traveling wave excites the recorded neuron because f1 has a much larger amplitude than 2f1-f2. Therefore, the lack of the detectable cochlear amplifier gain was attributed to measurement limitations.

To overcome the above constraint, Shera and Guinan (1997) performed an experiment similar to that of Allen and Fahey with an entirely different technique. The strength of the DP at its CF location was monitored via a secondary DP, which was evoked by adding a third tone. The secondary DP was monitored in the external ear canal while the ratio of f2/f1 was varied. The power gain of the cochlear amplifier was measured by the relationship of the first and second DP. Shera and Guinan reported evidence of significant amplification in a broad region basal to the peak of the traveling wave.

Recently, de Boer and co-workers (2005) made another attempt to explore the puzzle of the Allen-Fahey experiment by directly measuring the DP response at the CF location, using a laser interferometer. The intensity of the DP emission at 2f1-f2 was measured as a function of f2/f1 while the amplitude of the DP response at the CF location was kept constant. The ratio of f2/f1 was varied from 1.01 to 1.28 in six steps. It was found that as f2/f1 approached 1, the DP intensity in the ear canal decreased rather than increased as expected. In spite of not having the frequency-ratio limitation, de Boer and associates confirmed the results of Allen and Fahey.

De Boer and co-workers (2005) employed and extended computation results (Shera, 2003) to interpret their finding. Shera's computation showed that the DPs are generated over

[a]Also at Department of Physiology, School of Medicine, Xi'an Jiaotong University, 76 Yanta Xilu, Xi'an, Shaanxi 710061, P. R. China. Electronic mail: rent@ohsu.edu

 0001-4966/2006/119(4)/1940/3/$22.50

a certain distance on the BM, resulting in a considerable phase variation. When the ratio of f2 to f1 is close to 1, significant destructive interference occurs in the backward traveling wave, while no such interference occurs in the forward traveling wave. Consequently, the magnitudes of the backward traveling wave and the DP emission in the ear canal are decreased; the cochlear amplifier gain, therefore, cannot be measured in the Allen-Fahey experiment. Based on the measured frequency responses of the BM, de Boer and co-workers (2005) obtained a spatial pattern of the BM response based on a cochlear model and the cochlear frequency-location map. They also showed that the large phase variation over the DP generation region could cause a significant cancellation of the backward traveling wave.

A completely different view on the emission backward propagation is the cochlear compression-wave theory. According to this theory, the cochlea-generated emission reaches the stapes through the cochlear fluid as fast-propagating compression waves. For interpreting how the otoacoustic emission is generated in the cochlea, Wilson (1980) proposed a sensory outer-hair-cell swelling model, in which hair cell volume changes displace the stapes footplate and result in the emission in the external ear canal. Although the hair-cell swelling mechanism is questionable, this theory implies a pressure wave in the cochlear fluids, which results in otoacoustic emission. The cochlear compression theory was further advanced by other researchers (Avan *et al.*, 1998; Ruggero, 2004; Siegel *et al.*, 2005) by measuring BM vibration, pressures in the cochlear fluids, and the emissions in the ear canal. Using a scanning laser interferometer, Ren (2004) measured the longitudinal pattern of BM vibration at the emission frequency. The wave propagation direction was measured by phase longitudinal-location data: i.e., that phase decrease with distance from the cochlear base indicates a forward traveling wave. He found a normal forward traveling wave but no indication of a backward traveling wave at the emission frequency. A comprehensive measurement of the intracochlear pressure by Dong and Olson (2005) confirmed the feasibility of the cochlear compression wave.

The compression wave theory can satisfactorily interpret the well-documented results of the Allen-Fahey experiment, and these results in turn support the cochlear compression wave theory. De Boer and co-workers (2005) believe that the most perplexing property of the results of the Allen-Fahey experiment is that the nearer the frequency ratio of f2/f1 is to 1, the smaller the DP amplitude is in the ear canal. In contrast, the BM vibration data show that the closer the frequency ratio of f2/f1 is to 1, the greater the magnitude of the DP response at the CF place on the BM (Cooper and Rhode, 1997; Robles *et al.*, 1997). This inconsistency between the emission and BM vibration is interpreted using the cochlear compression wave theory in Fig. 1. The intensity of the emission in the ear canal is proportional to the magnitude of the stapes vibration, which results from the pressure difference between the scala vestibuli and scala tympani at the emission frequency [P_{DP} in Fig. 1(a)]. This pressure is determined by the transverse vibration velocity and the acoustic impedance of the cochlear partition at the DP generation site. When the ratio of f2/f1 is significantly larger than 1, the f1, f2, and

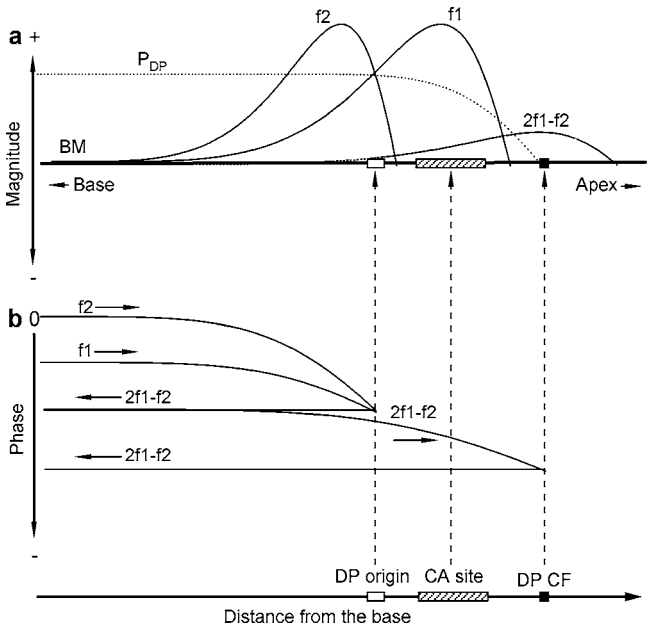


FIG. 1. Graphical interpretation of the Allen-Fahey experiment using the cochlear compression-wave mechanism. (a) Magnitudes of the f1, f2, and 2f1-f2 traveling waves (solid lines) and acoustic pressure across the basilar membrane at the DP frequency (P_{DP}) (dotted line). (P_{DP}) is independent on the longitudinal location between the cochlear base and DP generation site (DP origin), and decreases significantly near the DP CF location. Cochlear amplifier location for 2f1-f2 is indicated by "CA site." (b) Phase curves of f1, f2, and 2f1-f2 traveling waves as well as the fluid compression waves. The acoustic energy generated near the DP origin propagates through the cochlear fluids to the cochlear base at the speed of sound in water. It vibrates the stapes, resulting in emissions in the ear canal, and launches a forward traveling wave at 2f1-f2. Thus, in (b), the backward propagation of the emission shows no significant phase delay.

2f1-f2 traveling waves are spread out along the longitudinal direction. The acoustic impedance at the DP generation site near the f2 place (DP origin in Fig. 1) is significantly higher for 2f1-f2 DP according to a cochlear mathematic model (Neely, 1981). This impedance mismatch helps to maintain the pressure difference across the BM (P_{DP}). This pressure reaches the stapes instantly and drops near the DP resonant site, due to the low acoustic impedance at this site [dotted line in Fig. 1(a)]. However, when f2/f1 is close to 1, the f1, f2, and 2f1-f2 traveling waves are largely overlapped due to small frequency differences among f1, f2, and 2f1-f2. Under this condition, the magnitude of BM vibration at 2f1-f2 increases due to the low impedance, and the large vibration at f1 and f2. However, because P_{DP} is determined by the product of the velocity and the impedance, the small acoustic impedance near the DP CF may compensate for the increased BM vibration, and result in a decrease in P_{DP} and emission. The imaginary part of the impedance should be negative because the BM stiffness dominates the impedance at the region basal to the DP CF site (Neely, 1981). The f2/f1 ratio-dependent P_{DP} is shown by the dotted line in Fig. 1(a), can interpret the contradictory results of the Allen-Fahey experiment that the emission decreases when f2/f1 is close to 1. This same mechanism can also interpret the absence of the expected large cochlear amplifier gain because a cochlear traveling wave is required for cochlear amplification (Robles and Ruggero, 2001), and a compression wave provides no

magnification. The above interpretation is apparently not consistent with the active cochlear models (de Boer *et al.*, 2005), in which the acoustic impedance does not approach zero near the CF site. However, the impedance as a function of the longitudinal location remains to be measured experimentally.

In summary, the cochlear compression wave theory provides a satisfactory interpretation of the results of the Allen-Fahey experiment. The particular features of the Allen-Fahey experiment results, i.e., the decreased emission in the ear canal with an increased BM response when the frequency ratio of f2/f1 is close to 1 and the lack of cochlear amplification, imply that the cochlea-generated emission reaches the stapes through the cochlear fluid as a compression wave rather than as a backward traveling wave.

ACKNOWLEDGMENTS

We would like to thank E. de Boer, E. V. Porsov, and S. Matthews for valuable discussions. This research was supported by NIH-NIDCD and the NCRAR, Portland VA Medical Center.

Allen, J. B., and Fahey, P. F. (**1992**). "Using acoustic distortion products to measure the cochlear amplifier gain on the basilar membrane," J. Acoust. Soc. Am. **92**, 178–188.

Avan, P., Magnan, P., Smurzynski, J., Probst, R., and Dancer, A. (**1998**). "Direct evidence of cubic difference tone propagation by intracochlear acoustic pressure measurements in the guinea-pig," Eur. J. Neurosci. **10**, 1764–1770.

Cooper, N. P., and Rhode, W. S. (**1997**). "Mechanical responses to two-tone distortion products in the apical and basal turns of the mammalian cochlea." J. Neurophysiol. **78** (1), 261–270.

Davis, H. (**1983**). "An active process in cochlear mechanics," Hear. Res. **9**, 79–90.

de Boer, E., Nuttall, A. L., Hu, N., Zou, Y., and Zheng, J. (**2005**). "The Allen-Fahey experiment extended," J. Acoust. Soc. Am. **117**, 1260–1266.

Dong, W., and Olson, E. S. (**2005**). "Two-tone distortion in intracochlear pressure," J. Acoust. Soc. Am. **117**, 2999–3015.

Kemp, D. T. (**1986**). "Otoacoustic emissions, travelling waves and cochlear mechanisms," Hear. Res. **22**, 95–104.

Neely, S. T. (**1981**). "Finite difference solution of a two-dimensional mathematical model of the cochlea," J. Acoust. Soc. Am. **69**, 1386–1391.

Ren, T. (**2004**). "Reverse propagation of sound in the gerbil cochlea," Nat. Neurosci. **7**, 333–334.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Robles, L., Ruggero, M. A., and Rich, N. C. (**1997**). "Two-tone distortion on the basilar membrane of the chinchilla cochlea," J. Neurophysiol. **77**, 2385–2399.

Ruggero, M. A. (**2004**). "Comparison of group delays of 2f1-f2 distortion product otoacoustic emissions and cochlear travel times," ARLO **5**, 143–147.

Shera, C. A. (**2003**). "Wave interference in the generation of reflection- and distortion-source emissions," in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, M. P. Scherer (World Scientific, Singapore), pp. 439–454.

Shera, C. A., and Guinan, J. J., Jr. (**1997**). "Measuring cochlear amplification and nonlinearity using distortion product otoacoustic emissions as a calibrated intracochlear sound source," Abstracts of the Midwinter Meeting of the Association for Research in Otolaryngology, Vol. **20**, pp. 13.

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.

Wilson, J. P. (**1980**). "Model for cochlear echoes and tinnitus based on an observed electrical correlate," Hear. Res. **2**, 527–532.

Loudness of brief tones measured by magnitude estimation and loudness matching (L)

Michael Epstein[a)]
Institute for Hearing, Speech, and Language, Communications and Digital Signal Processing Center, ECE Dept. (440 DA); and Department of Speech-Language Pathology and Audiology (133 FR), Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Mary Florentine[b)]
Institute for Hearing, Speech, and Language and Department of Speech-Language Pathology, Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

(Received 19 September 2005; revised 24 January 2006; accepted 25 January 2006)

McFadden [J. Acoust. Soc. Am. **57**, 702–704 (1975)] questioned the accuracy and reliability of magnitude estimation for measuring loudness of tones that vary both in duration and level, whereas Stevens and Hall [Percept. Psychophys. **1**, 319–327 (1966)] reported reasonable group data. To gain insight into this discrepancy, the present study compares loudness measures for 5- and 200-ms tones using magnitude estimation and equal-loudness matches from the same listeners. Results indicate that both procedures provide rapid and accurate assessments of group loudness functions for brief tones, but may not be reliable enough to reveal specific characteristics of loudness in individual listeners. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2177592]

PACS number(s): 43.66.Cb [AJO] Pages: 1943–1945

I. INTRODUCTION

Magnitude estimation has frequently been used to measure the growth of loudness (e.g., Hellman and Zwislocki, 1964; Stevens, 1975; Poulton, 1989). Only a few previous experiments have used magnitude estimation to assess temporal integration, and they have had mixed success. Stevens and Hall (1966) measured the growth of loudness for bursts of white noise with different durations for 12 listeners using magnitude estimation. They made measurements that spanned a wide range of levels from 36 to 109 dB SPL and fitted a power function to the data. They found that the average slopes of the power functions were nearly the same for noise bursts ranging in duration from 5 to 500 ms. They did not report the variability or individual measures, but the reported data appear to be internally consistent. On the other hand, McFadden (1975) measured the growth of loudness for tones of different durations for four listeners using magnitude estimation and reported magnitude estimation unsuitable for obtaining individual loudness functions for pure tones that varied both in duration and level. His data show very large variability and the results are inconsistent with other measures of loudness as a function of duration. It is not completely clear why the results of these two studies differ so dramatically or whether magnitude estimation is suitable for obtaining detailed information about individual loudness growth.

The purpose of the present study is to gain insight into the usefulness of magnitude estimation as a method for measuring loudness functions for pure tones of different durations. Temporal integration was measured by magnitude estimation, and the results were compared to previously published data from the same listeners using cross-modality matching (Epstein and Florentine, 2005).

II. METHOD

Each of nine normal-hearing listeners estimated the loudness of individual tones by typing a number whose magnitude matched the tone's loudness on a digital microterminal. No reference or range was given as a basis for this judgment. First, absolute thresholds were measured using the method described in Epstein and Florentine (2005). Then, ten magnitude estimates were obtained for each level and duration in 5-dB steps from 5 dB SL to 100 dB SPL for the 200-ms tone and to 110 dB SPL for the 5-ms tone. These ten measurements were separated into two blocks of five each with a break in between.

The stimulus in each trial was chosen at random from all combinations of level and duration with the following restrictions: The SL had to be within 30 dB of the level in the previous trial for tones of the same duration and within 25 dB for tones of the other duration, and fewer than five presentations of that duration and level had been presented in the current block. The final estimates were the geometric mean of ten estimates completed for each duration and level.

III. RESULTS AND DISCUSSION

Figure 1 shows the geometric means of the magnitude estimates for each of the nine listeners. All listeners gave larger numbers for 200-ms tones than for 5-ms tones at the same level. The data for individual listeners are generally consistent, as indicated by the small standard errors and the overall monotonicity. For most listeners, functions for both the 5- and 200-ms tones are slightly shallower at moderate levels than at low and high levels.

[a)]Electronic mail: m.epstein@neu.edu
[b)]Electronic mail: florentin@neu.edu

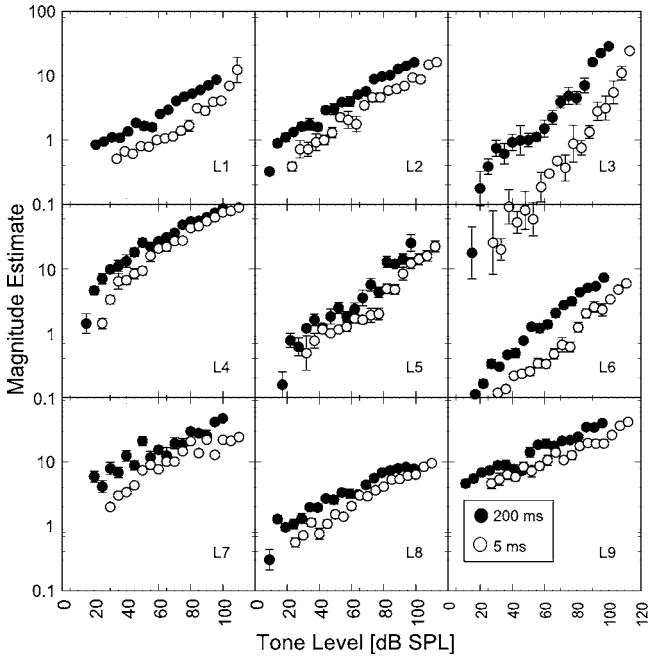


FIG. 1. Magnitude estimates as a function of tone level for nine listeners plotted on a log scale. The estimates for the 5-ms tones are shown by unfilled circles and the estimates for the 200-ms tones are shown by filled circles. The error bars show standard errors of the means.

Figure 2 shows the geometric means of the nine listeners' magnitude estimates presented in Fig. 1. The average loudness functions have a mid-to-high-level slope of about 0.18. This is lower than the frequently reported slope of 0.3 (e.g., Hellman, 1991), but it is within the range of slopes reported by Viemeister and Bacon (1988) and consistent with the observation that using a large range of stimuli reduces the measured slope (Scharf and Fishken, 1970). The mean data are much more orderly and monotonic than the individual results.

In order to compare the magnitude estimates with the loudness matches (by using both measures to determine the amount of temporal integration), functions were individually fit to the magnitude estimates for both durations for each

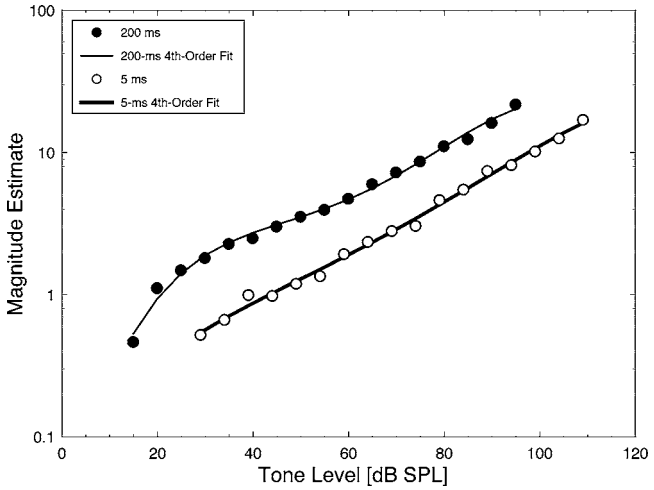


FIG. 2. Geometric means of nine listeners' magnitude estimates for 200- (filled circles) and 5-ms (unfilled circles) tones. The solid lines show fourth-order polynomial fits to the data.

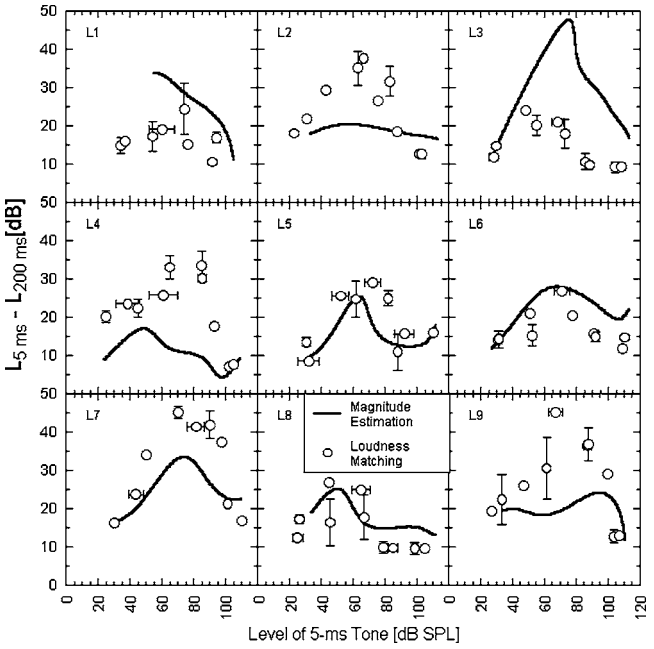


FIG. 3. Temporal integration of loudness derived from loudness matches replotted from Epstein and Florentine (2005) (unfilled circles) and magnitude estimation (solid lines). The horizontal error bars show the standard errors of the data for which the 5-ms tones were varied and the vertical error bars show the standard errors of the data for which the 200-ms tones were varied.

listener and for the group means. The polynomial fits were used to estimate the amount of temporal integration by determining the difference in level for 5- and 200-ms tones that produced the same magnitude estimates (i.e., the horizontal distance in dB between the two loudness functions at a particular loudness). Only one loudness-function fit had an R-squared value below 0.95 (L7 at 200 ms; R-squared $=0.89$).

Figures 3 and 4 show the amount of temporal integration derived from magnitude estimation and loudness matches [replotted from Epstein and Florentine (2005)] for individual listeners (Fig. 3) and the group mean (Fig. 4). The amount of temporal integration for individual listeners was derived from magnitude estimates by determining the horizontal distance (in dB) between the 200- and 5-ms magnitude-

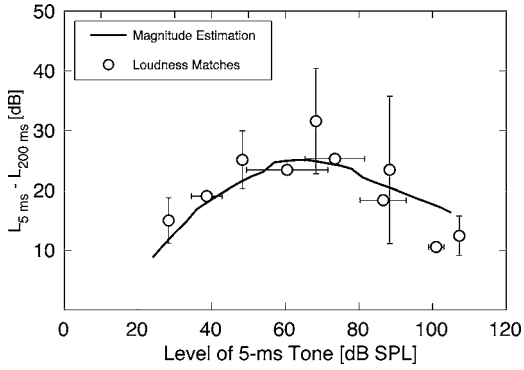


FIG. 4. Temporal integration of loudness derived from loudness matching replotted from Epstein and Florentine (2005) (unfilled circles) and magnitude estimation (solid line). Error bars indicate the standard deviation of the means of nine listeners.

estimation-derived loudness functions as a function of the level of the 5-ms tone. The mean amount of temporal integration was derived by averaging level differences between points yielding equal magnitude estimates for each listener.

Most listeners show a clear mid-level maximum in the amount of temporal integration in agreement with previous studies (e.g., Florentine *et al.*, 1996, 1998; Buus, 1999). The magnitude of the mid-level maximum varies greatly across individuals and covers a somewhat larger range than previous studies (cf., Florentine *et al.*, 1996). The amount of temporal integration derived from magnitude estimation very poorly matches that derived from loudness matching for several listeners, particularly L1, L2, L4, and L9. Listeners L2, L4, and L9 show much smaller amounts of temporal integration based on magnitude estimation than those based on loudness matches.

The mean temporal-integration functions obtained from loudness matching and magnitude estimation (Fig. 4) agree with one another, indicating that both measures yield reasonable group estimates of loudness. The measures are also consistent with similar studies (Florentine *et al.*, 1996, 1998; Buus *et al.*, 1999) also showing that the amount of temporal integration varies nonmonotonically with level and is largest at moderate levels.

In contrast to the mean data, individual results vary significantly. The general form of the loudness functions for all individual listeners was clear, but detailed information was obscured by variability in most cases. The loudness matches for individuals contained some highly inconsistent data points (i.e., points at nearly the same level for the 5-ms tone are matched with two 200-ms tones with significantly different levels).

IV. SUMMARY AND CONCLUSIONS

In general, measuring the loudness of brief tones for individuals is challenging. The temporal-integration data, from both magnitude estimation and loudness matching, indicate that listeners have a difficult time with such tasks. Magnitude estimation of loudness for tones with different durations shows high variability in some individual listeners. Loudness matches between brief and long tones results in somewhat internally consistent data for some listeners, but not for others. Therefore, experimenters should use caution when using magnitude estimation to determine individual loudness functions for pure tones of various durations or when obtaining loudness matches between brief and long tones. Mean data appear to be useful in assessing the shape of the loudness functions and making loudness matches for tones of different durations for a uniform-population group. However, mean data may not apply to any one individual. Large individual differences in reliability help explain why McFadden's (1975) data are highly variable and why average data obtained by Stevens and Hall (1966) showed little variability. If variability is a result of random effects, then larger listener pool means will be very effective at eliminating the variability.

ACKNOWLEDGMENTS

We would like to thank Editor Andrew Oxenham and one anonymous reviewer for helpful suggestions. Bert Scharf also gave helpful comments on an earlier version of this manuscript. This research was supported by NIH/NIDCD Grant No. R01DC02241. A portion of this work was presented to the International Society for Psychophysics in 2001.

Buus, S. (**1999**). "Loudness functions derived from measurements of temporal and spectral integration of loudness," in *Auditory Models and Nonlinear Hearing Instruments* edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen (GN ReSound, Taastrup, Denmark).

Buus, S., Florentine, M., and Poulsen, T. (**1999**). "Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin," J. Acoust. Soc. Am. **105**, 3464–3480.

Epstein, M., and Florentine, M. (**2005**). "A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions," J. Acoust. Soc. Am. **118**, 907–913.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.

Florentine, M., Buus, S., and Robinson, M. (**1998**). "Temporal integration of loudness under partial masking," J. Acoust. Soc. Am. **104**, 999–1007.

Hellman, R. P. (**1991**). "Loudness scaling by magnitude scaling: Implications for intensity coding," in *Ratio Scaling of Psychological Magnitude: In Honor of the Memory of S.S. Stevens* edited by G. A. Gescheider and S. J. Bolanowski (Erlbaum, Hillsdale, NJ).

Hellman, R. P., and Zwislocki, J. J. (**1964**). "Loudness function of a 1000-cps tone in the presence of a masking noise," J. Acoust. Soc. Am. **36**, 1618–1627.

McFadden, D. (**1975**). "Duration-intensity reciprocity for equal loudness," J. Acoust. Soc. Am. **57**, 702–704.

Poulton, E. C. (**1989**). *Bias in Quantifying Judgments* (Erlbaum, Hillsdale, NJ).

Scharf, B., and Fishken, D. (**1970**). "Binaural summation of loudness: reconsidered," J. Exp. Psychol. **86**, 374–379.

Stevens, J. C., and Hall, J. W. (**1966**). "Brightness and loudness as a function of stimulus duration," Percept. Psychophys. **1**, 319–327.

Stevens, S. S. (**1975**). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects* (Wiley, New York).

Viemeister, N. F., and Bacon, S. P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," J. Acoust. Soc. Am. **84**, 172–178.

Auditory temporal resolution in normal-hearing preschool children revealed by word recognition in continuous and interrupted noise (L)[a)]

Andrew Stuart,[b)] Gregg D. Givens, Letitia J. Walker,[c)] and Saravanan Elangovan[d)]
Department of Communication Sciences and Disorders, East Carolina University, Greenville, North Carolina 27858-4353

(Received 1 October 2005; revised 18 January 2006; accepted 27 January 2006)

The purpose of this study was to examine temporal resolution in normal-hearing preschool children. Word recognition was evaluated in quiet and in spectrally identical continuous and interrupted noise at signal-to-noise ratios (S/Ns) of 10, 0, and −10 dB. Sixteen children 4 to 5 years of age and eight adults participated. Performance decreased with decreasing S/N. At poorer S/Ns, participants demonstrated superior performance or a release from masking in the interrupted noise. Adults performed better than children, yet the release from masking was equivalent. Collectively these findings are consistent with the notion that preschool children suffer from poorer processing efficiency rather than temporal resolution *per se*. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2178700]

PACS number(s): 43.66.Mk, 43.66.Dc [JHG] Pages: 1946–1949

I. INTRODUCTION

The normal development of auditory temporal processing (i.e., resolution/acuity and integration/summation) in children has been of interest to psychoacousticians and clinicians. Temporal resolution refers to the ability of a listener's auditory system to resolve/separate auditory events or perceive changes in auditory stimuli over time. Temporal integration refers to the ability of a listener's auditory system to sum acoustic information over time to improve detection, recognition, or discrimination of stimuli.[1] Those examining normal auditory temporal processing development with various test paradigms have found that the performance of normal-hearing infants and children is inferior to adults.[2]

For clinicians, understanding normal development is essential for determining if the perceptual capacity of a child is abnormal. Identification of impaired auditory temporal processing is a necessary precedent for rehabilitative measures for those that implicate an underlying temporal auditory processing deficit for some communicative impairments[3–5] or for those evaluating temporal processing as part of an auditory processing test battery.[6] The early identification of a temporal processing deficit could result in the ability to begin remediation programs for such impairments, perhaps even before a child reaches school age.

Stuart and colleagues[7–15] have utilized word recognition in spectrally identical continuous and interrupted broadband noise as a function of signal-to-noise ratio (S/N) to examine temporal resolution abilities of normal-hearing and impaired listeners. Listeners experience a perceptual advantage or "release from masking" in interrupted noise. Since the noises differ only in temporal continuity, better performance in interrupted noise has been attributed to the ability of listeners to get glimpses or looks of each word between silent gaps and patch the information together in order to identify the specific word.[16–18] Any release from masking observed with listeners in the interrupted noise compared to the continuous noise, at equivalent S/Ns, is evidence of auditory temporal resolution. Assessing the auditory temporal resolution capacity between groups of listeners can be done by comparing overall performance in the interrupted noise and also by examining the amount of release from masking in the interrupted noise relative to the continuous noise.

Stuart[10] recently reported the development of word recognition in continuous and interrupted noise in 80 normal-hearing children aged 6 to 15 years. Word recognition performance was evaluated in quiet and in continuous and interrupted noise at S/Ns of 10, 0, −10, and −20 dB. Children displayed better performance in the interrupted noise compared to the continuous noise at poorer S/Ns (i.e., <10 dB) and performance increased with improving S/N. Performance also improved with increasing age. Younger children were more vulnerable to noise in that they required more favorable S/Ns to perform the same as older children and adults. Children's performance in noise equated adults after 11 years of age.

The purpose of this study was to examine word recognition performance of normal-hearing preschool-aged children in continuous and interrupted noise relative to adult listeners. It was of interest to see whether children of this age demonstrate a temporal perceptual advantage in the interrupted noise condition. Ultimately, it was of interest to generate a normative base for word recognition performance in continuous and interrupted noise for children aged 4 to 5 years of age. As such, these data could be used as a

[a)]This work was presented in part at the 2004 American Speech-Language-Hearing Association Annual Convention, Philadelphia, PA, 20 November 2004.
[b)]Electronic mail: stuart@ecu.edu
[c)]Currently affiliated with the Department of Communication Sciences and Disorders, Missouri State University, Springfield, MO.
[d)]Currently affiliated with the Department of Communicative Disorders, East Tennessee State University, Johnson City, TN.

 0001-4966/2006/119(4)/1946/4/$22.50

clinical tool to assess auditory temporal processing ability of young preschool children. It was hypothesized that performance would improve with increasing S/N, performance in the interrupted noise would be better than in the continuous noise, and children would perform poorer than adults.

II. METHODS

A. Participants

Sixteen preschool children (five males and eleven females) aged 4 to 5 years ($M=4.8$, SD=0.6) and eight young adults (five males and three females; $M=23.9$, SD=2.4) participated. Children were solicited through their parents, whom were faculty, staff, or students at East Carolina University, Greenville, NC. All participants presented with normal-hearing sensitivity as defined by pure-tone thresholds at octave frequencies from 250 to 8000 Hz and spondee recognition thresholds of ≤20 dB HL and normal middle-ear function.[19] Participants were native speakers of English and had a negative history of speech, language, cognitive, learning, and vision disorders. The children presented with an age-equivalent receptive vocabulary score as assessed by the *Peabody Picture Vocabulary Test-Revised*—3rd Edition.[20]

B. Apparatus and procedure

Northwestern University—Children's Perception of Speech[21] (NU-CHIPS, Auditec of St. Louis) monosyllabic words and custom competing continuous and interrupted noises served as test stimuli. The noises are described in detail elsewhere.[9,12] The interrupted noise was constructed with rectangular gated noise bursts and silent periods, both with durations varying randomly from 5 to 95 ms. The noise duty cycle for the interrupted noise was 0.50. All speech and noise files were normalized to have equal power. The long-term average spectra of both noises were the same.

The compact disc recordings of the stimuli were delivered through a dual-disc compact disc player (Phillips model CDR 765 K02) or two compact disc players (Sony model CDP-CE415) to a clinical audiometer (Grason Stadler GSI 61 model 1761-9780XXE). Stimuli were presented monaurally, in a double-wall sound-treated audiometric suite, to each participant's right ear through a supraaural earphone (Telephonics model TDH-50P).

The NU-CHIPS speech stimuli were presented at 50 dB HL to the right ear of participants. Average presentation levels were 36.2 dB (SD=3.4) and 44.0 (SD=3.2) above the spondee recognition threshold for the children and adults, respectively. In no case was the presentation level less than 30 dB above the listener's spondee recognition threshold where age effects on performance are evident in children less than 10 years of age.[21] Eight half-lists (i.e., 25 monosyllabic words) of the four NU-CHIPS lists were employed. The speech stimuli were presented in quiet and in both noises at S/Ns of −10, 0, and 10 dB. List presentation order was counterbalanced while noise and S/N conditions were randomized across participants. Participants were instructed to point to the picture from a set of four alternatives (i.e., one stimulus and three foils) of the word that they heard.[21]

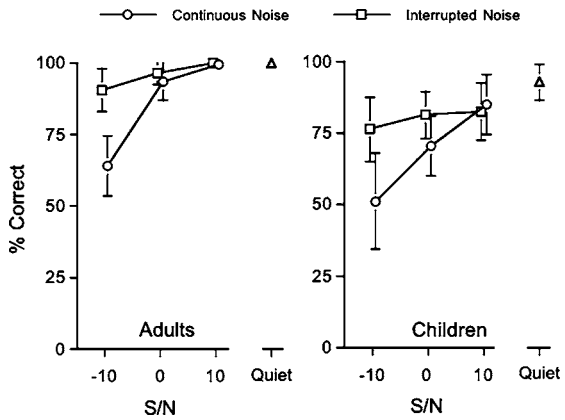


FIG. 1. Mean percent-correct word recognition scores in quiet and noise as a function of group, noise type, and S/N. Error bars represent plus/minus one standard deviation of the mean.

III. RESULTS

Participants' responses were scored as total whole word percent correct. Figure 1 illustrates the mean group word recognition performance in quiet and in both noises as a function of S/N and group. These proportional scores were transformed to rationalized arcsine units prior to inferential statistical analyses.[23] Violations of the analysis of variance (ANOVA) assumptions were examined before investigating differences in word recognition performance. Levene's test of equality of error variance was significant ($p<0.05$) for S/Ns of +10 for both noises and quiet. Consequently, scores for S/Ns at +10 for both noises were excluded from the omnibus analyses.

A three-factor mixed ANOVA was performed to investigate mean word recognition performance differences as a function of group, S/N, and noise condition. The results of

TABLE I. Summary table of a three-factor mixed ANOVA investigating differences in word recognition performance as a function of group, noise, and S/N.

Source	*df*	*F*	*p*	η^2	ϕ
Group	1	31.22	<0.0001[a]	0.59	1.0
Noise	1	130.18	<0.0001[a]	0.86	1.0
S/N	1	65.19	<0.0001[a]	0.75	1.0
Noise × group	1	0.001	0.98	0.00	0.050
S/N × group	1	7.49	0.012[a]	0.25	0.74
Noise × S/N	1	19.05	<0.0001[a]	0.46	0.99
Noise × S/N × group	1	1.66	0.21	0.07	0.24

[a]Significant at $p<0.05$.

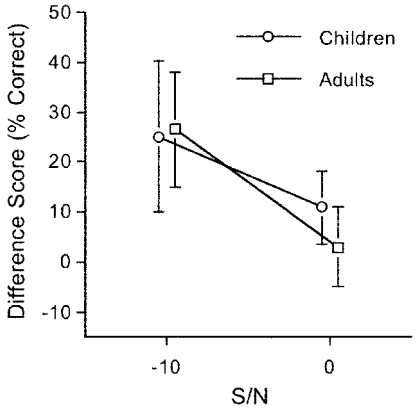


FIG. 2. Mean percent-correct word recognition difference score (i.e., interrupted noise minus continuous noise score) as a function of group and S/N.

that ANOVA are displayed in Table I. As expected, significant main effects of group, noise, and S/N were found, indicating better performance by adults, better performance in the interrupted noise, and improvement in performance with increasing S/N. The significant noise by S/N interaction reflects the release from masking phenomenon. That is, as S/N deteriorates performance worsens more rapidly in the continuous noise versus the interrupted noise. The significant group by S/N interaction reflects the fact that adults' performance improves much more as S/N improves.

The extent of the release from masking that was experienced in the interrupted noise relative to the continuous noise was examined by computing a difference score where participants' scores in continuous noise were subtracted from their scores in interrupted noise at 0 and −10 dB S/N. All participants had better scores in the interrupted noise compared to continuous noise at −10 dB S/N, and with the exception of two listeners in each group, all scored better at 0 dB S/N. Those that scored better in the continuous noise at 0 dB S/N did so by only one or two words (i.e., 4% or 8%). These difference scores as a function of group and S/N are displayed in Fig. 2. A two-factor mixed ANOVA was performed to investigate differences in mean word recognition difference scores as a function of group and S/N. A main effect of S/N was found [$F(1,22)=19.05$, $p<0.0001$, $\eta^2=0.46$, $\phi=0.99$], while a nonsignificant main effect of group [$F(1,22)=0.001$, $p=0.98$, $\eta^2=0.00$, $\phi=0.050$] and group by S/N interaction [$F(1,22)=1.67$, $p=0.21$, $\eta^2=0.070$, $\phi=0.24$] was found.

IV. DISCUSSION AND CONCLUSIONS

As hypothesized, performance improved with increasing S/N, was superior in interrupted noise, and children performed poorer than adults. Most important was that children as young as 4 to 5 years of age demonstrated better performance in the interrupted noise relative to the continuous noise at the poorer S/Ns (i.e., <10 dB). This is consistent with previous findings where normal-hearing adult listeners experience a release from masking in interrupted noise with monosyllabic word recognition.[7–18,23,24] This is the first demonstration in a preschool-aged cohort of children of this phenomenon. Significant main effects of noise and S/N with word recognition and noise by S/N interaction in this test paradigm have been reported repeatedly with normal-hearing adult listeners by Stuart and colleagues[7–15] and others.[16,17,24]

The results from this study are consistent with previous reports of 3- and 5-year-old children with the same stimuli in quiet with a closed-set response.[21] Performance in quiet and noise was superior to that of 6- to 7-year-olds reported by Stuart.[10] This is likely due to an open-set response employed by Stuart.[10] Under similar listening conditions, NU-CHIPS performance is better in a closed-versus an open-set response mode.[24,25] In only one other study utilizing the NU-CHIPS stimuli in continuous noise, Chermak *et al.*[26] reported a mean performance of approximately 72% for children between the ages of 9 and 10 years at 0 dB S/N with a closed-set response mode. Considering that the older children in the Chermak *et al.*[26] study performed approximately the same as the younger 4- to 5-year-olds in this study, one may suggest that differences in the recorded stimuli and competing continuous noise may have contributed to the fact that age differences were not evident. Overall, preschool children performed poorer in noise compared to adults, consistent with previous reports demonstrating that young children need greater S/Ns to perform at adult levels.[10,21,25]

The basis of performance differences between younger and older listeners remains a contentious issue. Two schools of thought exist:[27–30] One embraces the notion that children have a broader temporal window and therefore have poorer temporal acuity than older listeners (i.e., the "temporal resolution hypothesis"). The other suggests that children have poor processing efficiency (i.e., the "processing efficiency hypothesis"). Processing efficiency refers to factors "aside from temporal and spectral resolution, that affect the ability to detect acoustic signals in noise… [and] is measured by the threshold signal-to-noise ratio" (p. 2962).[28] Hartley and colleagues[27–30] suggest that children have more "internal noise" than adults and thus require higher effective S/N in order to perform equivalently. This is consistent with the fact that the peripheral auditory system is adult-like by 4 to 6 years of age,[29,31,32] but the central auditory system is less proficient.

The data herein support the poorer processing efficiency hypothesis. Although overall performance was worse with the children, the amount of release from masking was the same as adults. This same pattern was seen with school-aged children reported by Stuart.[10] We computed difference scores at 0 and −10 dB S/N for the five groups of school-aged children and adults from this previous study. A two-factor mixed ANOVA was performed to investigate differences in mean word recognition difference scores as a function of group and S/N. A main effect of S/N was found [$F(1,90)=136.67$, $p<0.0001$, $\eta^2=0.60$, $\phi=1.0$], while a nonsignificant main effect of group [$F(5,90)=1.33$, $p=0.26$, $\eta^2=0.069$, $\phi=0.45$] and group by S/N interaction [$F(5,90)=0.91$, $p=0.48$, $\eta^2=0.048$, $\phi=0.31$] was found. Thus, the apparent difference between preschool and school-aged children less than 12 years of age is related to more general differences in their abilities to recognize speech in degraded listening conditions in which there are a number of contributors related to the development of central audition, language, and attention. As previously stated by Stuart,[10] it is important to note that

younger auditory systems are not impaired in any way, rather, that they are normally developing yet have poorer processing efficiency that impairs their performance in noise relative to older listeners. Further, their inferior performance compared to normal adults is not the same as the inferior performance seen in adult listeners with auditory pathologies reported by Stuart, Phillips, and colleagues. Those with noise-induced hearing loss,[33] unilateral high-frequency hearing loss,[11] retrocochlear demyelinating lesions,[34] and presbyacusis[12] display overall poorer performance *and* a smaller release from masking the interrupted noise relative to young normal-hearing adults. This is consistent with poorer temporal resolution in these pathologies.

In summary, this investigation demonstrated that the word recognition in continuous and interrupted noise is poorer in preschool children than adults. The release from masking observed with these preschool children in the interrupted noise compared to the continuous noise was, however, equivalent to that of adults. Collectively, these findings suggest a developmental difference in processing efficiency between preschool children and adults, rather than developmental differences in temporal resolution abilities. The findings do not, admittedly, address what perceptual process or processes are responsible for the inferior performance among these children. Further research is warranted to address this question in normal and particularly impaired children. That is, the mechanisms underlying communication disorders must be understood such that remediation strategies focus on improving those mechanisms. In terms of clinical implementation concerning time restrictions and the difficulty of maintaining children's attention at this age, it is recommended that one administer separate lists at the −10-dB S/N for each noise condition. This would only take approximately 10 min and would provide the most information regarding the release from masking in the interrupted noise. One caveat to this approach is that care must be taken to ensure that word lists used are equivalent. Although the NU-CHIPS lists and half-lists are equivalent in quiet[21] they are not in continuous noise.[26] List equivalency of other word recognition material has not been demonstrated[9] with the same interrupted noise, and should therefore not be expected with the NU-CHIPS stimuli until demonstrated otherwise.

[1]B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 5th ed. (Academic Press, London, 2003).

[2]L. A. Werner and G. C. Marean, *Human Auditory Development* (Brown & Benchmark, Madison, WI, 1996).

[3]M. M. Merzenich, W. M. Jenkins, P. Johnston, C. Schreiner, S. L. Miller, and P. Tallal, "Temporal processing deficits of language-learning impaired children ameliorated by training," Science **271**, 77–81 (1996).

[4]C. Sloan, *Treating Auditory Processing Difficulties in Children* (College-Hill, San Diego, 1986).

[5]P. Tallal, S. L. Miller, G. Bedi, G. Byma, X. Wang, S. S. Nagarajan, C. Schreiner, W. M. Jenkins, and M. M. Merzenich, "Language comprehension in language-learning impaired children improved with acoustically modified speech," Science **271**, 81–84 (1996).

[6]American Speech-Language-Hearing Association Task Force on Central Auditory Processing Consensus Development, "Central auditory processing: Current status of research and implications for clinical practice," Am J. Audiol. **5**, 41–54 (1996).

[7]S. Elangovan and A. Stuart, "Interactive effects of high-pass filtering and masking noise on word recognition," Ann. Otol. Rhinol. Laryngol. **114**, 867–878 (2005).

[8]T. Scott, W. B. Green, and A. Stuart, "Interactive effects of low-pass filtering and masking noise on word recognition," Acute Care **12**, 437–444 (2001).

[9]A. Stuart, "An investigation of list equivalency of the Northwestern University Auditory Test No. 6 in interrupted broadband noise," Am. J. Audiol. **13**, 23–28 (2004).

[10]A. Stuart, "Development of auditory temporal resolution in school-age children revealed by word recognition in continuous and interrupted noise," Ear Hear. **26**, 78–88 (2005).

[11]A. Stuart and M. Carpenter, "Unilateral auditory temporal resolution deficit: A case study," J. Commun. Disord. **32**, 317–325 (1999).

[12]A. Stuart and D. P. Phillips, "Word recognition in continuous and interrupted broadband noise by young normal-hearing, older normal-hearing, and presbycusic listeners," Ear Hear. **17**, 478–489 (1996).

[13]A. Stuart and D. P. Phillips, "Word recognition in continuous noise, interrupted noise, and in quiet by normal-hearing listeners at two sensation levels," Scand. Audiol. **26**, 112–116 (1997).

[14]A. Stuart and D. P. Phillips, "Deficits in auditory temporal resolution revealed by a comparison of word recognition under interrupted and continuous noise masking," Semin. Hear. **19**, 333–344 (1998).

[15]A. Stuart, D. P. Phillips, and W. B. Green, "Word recognition performance in continuous and interrupted noise by normal-hearing and simulated hearing-impaired listeners," Am. J. Otol. **16**, 658–663 (1995).

[16]D. D. Dirks, R. H. Wilson, and D. R. Bower, "Effect of pulsed masking on selected speech materials," J. Acoust. Soc. Am. **46**, 898–906 (1969).

[17]G. A. Miller, "The masking of speech," Psychol. Bull. **44**, 105–129 (1947).

[18]I. Pollack, "Masking by periodically interrupted noise," J. Acoust. Soc. Am. **27**, 353–355 (1955).

[19]American-Speech-Language-Hearing Association, *Guidelines for Audiologic Screening* (American-Speech-Language-Hearing Association, Rockville Pike, MD, 1997).

[20]L. M. Dunn and L. M. Dunn, *Peabody Picture Vocabulary Test-Revised* (American Guidance Service, Circle Pines, MN, 1981).

[21]L. L. Elliot and D. R. Katz, *Development of a New Children's Test of Speech Discrimination* (Auditec, St. Louis, MO, 1980).

[22]G. Studebaker, "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462 (1985).

[23]I. Pollack, "Masking of speech by repeated bursts of noise," J. Acoust. Soc. Am. **26**, 1053–1055 (1954).

[24]G. A. Miller and C. R. Licklider, "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173 (1950).

[25]L. L. Elliot, S. Connors, E. Kille, S. Levin, K. Ball, and D. Katz, "Children's understanding of monosyllabic nouns in quiet and in noise," J. Acoust. Soc. Am. **66**, 12–21 (1979).

[26]G. D. Chermak, C. M. Pederson, and R. B. Bendel, "Equivalent forms and split-half reliability of the NU-CHIPS administered in noise," J. Speech Hear Disord. **49**, 196–201 (1984).

[27]D. E. Hartley, P. R. Hill, and D. R. Moore, "The auditory basis of language impairments: Temporal processing versus processing efficiency hypotheses," Int. J. Pediatr. Otorhinolaryngol. **67** (Suppl. 1), S137–S142 (2003).

[28]D. E. Hartley and D. R. Moore, "Auditory processing efficiency deficits in children with developmental language impairments," J. Acoust. Soc. Am. **112**, 2962–2966 (2002).

[29]D. E. Hartley, B. A. Wright, S. C. Hogan, and D. R. Moore, "Age-related improvements in auditory backward and simultaneous masking in 6 - to 10-year-old children," J. Speech Lang. Hear. Res. **43**, 1402–1415 (2000).

[30]P. R. Hill, D. E. Hartley, B. R. Glasberg, B. C. Moore, and D. R. Moore, "Auditory processing efficiency and temporal resolution in children and adults," J. Speech Lang. Hear. Res. **47**, 1022–1029 (2004).

[31]P. Allen, F. Wightman, D. Kistler, and T. Dolan, "Frequency resolution in children," J. Speech Hear. Res. **32**, 317–322 (1989).

[32]J. W. Hall III and J. H. Grose, "Notched noise measures of frequency selectivity in adults and children using fixed-masker-level and fixed-signal-level presentation," J. Speech Hear. Res. **34**, 651–660 (1991).

[33]D. P. Phillips, J. M. Rappaport, and J. M. Gulliver, "Impaired word recognition in noise by patients with noised-induced cochlear hearing loss: Contribution of a temporal resolution defect," Am. J. Otol. **15**, 679–686 (1994).

[34]J. M. Rappaport, J. M. Gulliver, D. P. Phillips, R. A. Van Dorpe, C. E. Maxner, and V. Bhan, "Auditory temporal resolution in multiple sclerosis," J. Otolaryngol. **23**, 307–324 (1994).

Perceptual learning in speech: Stability over time (L)

Frank Eisner[a)]
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands and Institute of Cognitive Neuroscience, University College London, United Kingdom

James M. McQueen
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 23 June 2005; revised 27 January 2006; accepted 31 January 2006)

Perceptual representations of phonemes are flexible and adapt rapidly to accommodate idiosyncratic articulation in the speech of a particular talker. This letter addresses whether such adjustments remain stable over time and under exposure to other talkers. During exposure to a story, listeners learned to interpret an ambiguous sound as [f] or [s]. Perceptual adjustments measured after 12 h were as robust as those measured immediately after learning. Equivalent effects were found when listeners heard speech from other talkers in the 12 h interval, and when they had the opportunity to consolidate learning during sleep. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2178721]

PACS number(s): 43.71.Bp, 43.71.Es [ARB] Pages: 1950–1953

I. INTRODUCTION

When we listen to speech, we need to adjust our interpretation of speech cues in response to talker-specific differences in articulation (Ladefoged, 1989; Ladefoged and Broadbent, 1957). The variability in the speech signal that is introduced by talker idiosyncrasies continues to be problematic for automatic speech recognizers, but is usually handled with remarkable ease by the human perceptual system. By comparing comprehension of novel and familiar talkers under difficult listening conditions, Nygaard *et al.* (1994) and Nygaard and Pisoni (1998) have shown that being familiar with a talker's voice can even aid comprehension once an initial adjustment has been made.

There are likely to be various processes engaged in perceptual adjustments made to a talker, driven by different sources of talker variability, and operating at several levels, such as the phonemic, lexical, and prosodic levels. A recent study has shown one specific mechanism, which uses lexical knowledge to resolve ambiguities that arise in the signal at the sublexical level (Norris *et al.*, 2003). Exposure to an ambiguous sound [?], that was midway between [f] and [s], caused a shift of the [f]–[s] category boundary when [?] was placed in contexts that were lexically consistent with its interpretation as either [f] or [s]. Two groups of Dutch listeners heard this ambiguous sound while performing a lexical decision task, either in contexts favoring [f] (e.g., *olij?*, where *olijf* is a word, "olive," but *olijs* is not), or in contexts favoring [s] (e.g., *radij?*, where *radijs* is a word, "radish," but *radijf* is not). Listeners in the first group subsequently categorized more sounds on an [f]–[s] continuum as [f] than listeners in the second group.

The studies by Nygaard *et al.* and Norris *et al.* suggest that the perceptual system has access to previously acquired information about a talker. The present study asks whether this kind of perceptual learning remains stable over a 12-h period. This follows up on recent research using the Norris *et al.* exposure-test paradigm that has shown a solid, and under some conditions even increased, perceptual adjustment effect 25 min after learning (Kraljic and Samuel, 2005). A second question was whether conditions that favor consolidation of learning, such that there is little contact with other talkers, as well as the opportunity for sleep, produce a more robust effect than conditions where participants have normal day-to-day interaction with other talkers, and no sleep. A study in which participants were trained to understand synthetic speech has found that, for this type of learning, there is indeed a performance increase when the testing conditions allow sleep over conditions without sleep (Fenn *et al.*, 2003).

To address these questions, an adapted version of the Norris *et al.* (2003) paradigm was used for inducing a perceptual adjustment. Listeners were first pretested on their categorization of [f]–[s] sounds before having lexically biased exposure to an ambiguous fricative, in the context of listening to a story. They were tested again on [f]–[s] categorization immediately after exposure, and after a 12-h delay, either over the course of one day, or with an intervening night's sleep.

II. METHOD

A. Participants

Eighty-four native Dutch speakers with no self-reported hearing disorders took part in exchange for a cash payment. Twenty-four participated in pretests, and 60 participated in the main experiment.

B. Materials and stimulus construction

Speech recordings were made in a sound-damped booth (Sony ECM-MS957 microphone) in a single session and digitized for further processing (Sony SMB-1 A/D converter; 44.1 kHz sampling rate; 16-bit quantization). A female na-

[a)]Electronic mail: f.eisner@ucl.ac.uk

 0001-4966/2006/119(4)/1950/4/$22.50

tive Dutch speaker produced 20 tokens each of the syllables [ɛf], [ɛs], and [ɛx] for test stimulus construction, and read out two versions of a story (see below).

1. [ɛf]–[ɛs] continuum

One token each of [f] and [s] was selected from the recorded syllables and excised at zero crossings at the onset of frication (original durations: [s] 246 ms, [f] 234 ms; original intensities: [s] 67.7 dB SPL, [f] 61.3 dB SPL). The fricatives were cut to a duration of 231 ms, and equated in root-mean-square intensity (62.4 dB SPL). With these sounds as endpoints, an 81-step continuum was made by combining their waveforms in graded, equally spaced proportions (effectively manipulating the spectrum; see McQueen, 1991), where step 1 corresponded to a clear [f] and step 81 to a clear [s]. The resulting fricatives were spliced onto a vowel excised from one of the [ɛx] syllables (duration 111 ms; intensity 79.2 dB SPL). The velar vocalic context was used for all spliced sounds in the experiment in order to avoid transitional cues for [f] or [s].

The [ɛf]–[ɛs] continuum was pretested with 24 Dutch listeners in order to find a maximally ambiguous sound for the exposure materials, and to select stimuli for the test phases of the main experiment. First, 12 listeners categorized ten sounds from the ambiguous range of the continuum (between steps 17 and 53; presented ten times each, in pseudorandomized order). Using the same procedure, a further 12 listeners then categorized ten stimuli taken from a narrower ambiguous range as determined by the first group's responses (between steps 30 and 53). From the second group's responses, steps on the continuum corresponding to 90, 70, 50, 30, and 10 percent of [f] responses were identified or determined by interpolation. The resulting steps (25, 34, 43, 52, and 61) were used in the test phases of the main experiment. The most ambiguous sound, step 43 ([?]), was also used to create the materials for the exposure phase.

2. Story

The text of a Dutch translation of a story (de Saint-Exupéry, 2001, Chap. 2) was edited such that it contained an equal number of [f] and [s] sounds and neither of the sounds [v] or [z]. After editing there were 644 words in total, containing 78 [f] sounds and 78 [s] sounds. Two versions of the story were recorded. In one version, every instance of [f] was intentionally mispronounced as the voiceless velar fricative [x] (e.g., *alsof* "as if" →[alsɒx]). In the second version every [s] was pronounced as [x] (e.g., *alsof*→[alxɒf]). The 78 critical velar fricatives in both versions were then excised at zero crossings and replaced by a version of the ambiguous fricative [?]. Since in natural speech the duration of segments is conditioned by various contextual factors, there were three tokens of [?] (all based on step 43). These were made by modifying the amplitude envelope to create two shorter 60-ms and 100-ms sounds (linearly ramped over a 10-ms window at onset and offset), and a long 160-ms sound (ramped over 10 ms at onset and 40 ms at offset). For any given position, the most natural-sounding token out of these three was used. The final two versions of the story were 4.0 min long.

C. Design and procedure

All participants were given a pretest in which they categorized the five [ɛf]–[ɛs] steps, followed by an exposure phase where the task was simply to listen to one of the two story versions. Immediately after exposure, there was a first categorization post-test, and after a delay of 12 h, a second post-test.

For 30 participants, the pretest started at 9 am, and post-test–2 was at 9 pm on the same day ("day group"). For a further 30 subjects, the first session began at 9 pm, while post-test–2 took place at 9 am the following morning ("night group"). In each of those groups, there were 15 listeners who heard the [f]-biased version of the story during exposure (i.e., [?] replacing [f]), and 15 listeners who heard the [s]-biased version.

Pretest, post-test-1, and post-test-2 all consisted of ten randomizations of the same five [ɛf]–[ɛs] steps. Stimuli were presented at an interonset interval of 2600 ms. Listeners were tested in groups of up to four, and instructed to press a button labeled "F" when hearing an [f]-like sound, and a button labeled "S" for an [s]-like sound.

III. RESULTS

For every test phase, listeners' responses were converted to a percentage of [f] categorizations per step. Data from three participants (day group; [s]-biased exposure) were corrupted due to a technical error, and discarded. All listeners in the night groups confirmed having had at least 6 h of sleep between the post-tests.

A. Stability of learning

An initial analysis of variance (ANOVA) with test (pretest, post-test-1, or post-test-2) and step (the five [f]–[s] sounds) as within-subjects factors and lexical bias ([f]- or [s]-biased exposure) as a between-subjects factor revealed a significant interaction of test and lexical bias [$F(2,110)=3.68$, $p=0.028$]. The interaction was examined by conducting ANOVAs with step and lexical bias as factors separately for each test phase. While there was no significant difference between the two exposure groups at pretest [$F(1,55)=0.02$, $p=0.893$; see Fig. 1(a)], their respective categorization functions were significantly different from each other immediately after the exposure phase [$F(1,55)=5.76$, $p=0.020$; see Fig. 1(b)], and after a 12-h delay [$F(1,55)=4.76$, $p=0.033$; see Fig. 1(c)]. For a direct comparison of these perceptual learning effects in post-tests 1 and 2, we conducted an ANOVA with test (post-test-1 or post-test-2), lexical bias, and step as factors. Crucially, the interaction of test and lexical bias was not significant [$F(1,55)=0.13$, $p=0.726$], suggesting that perceptual learning was as robust after 12 h as it was immediately after learning.

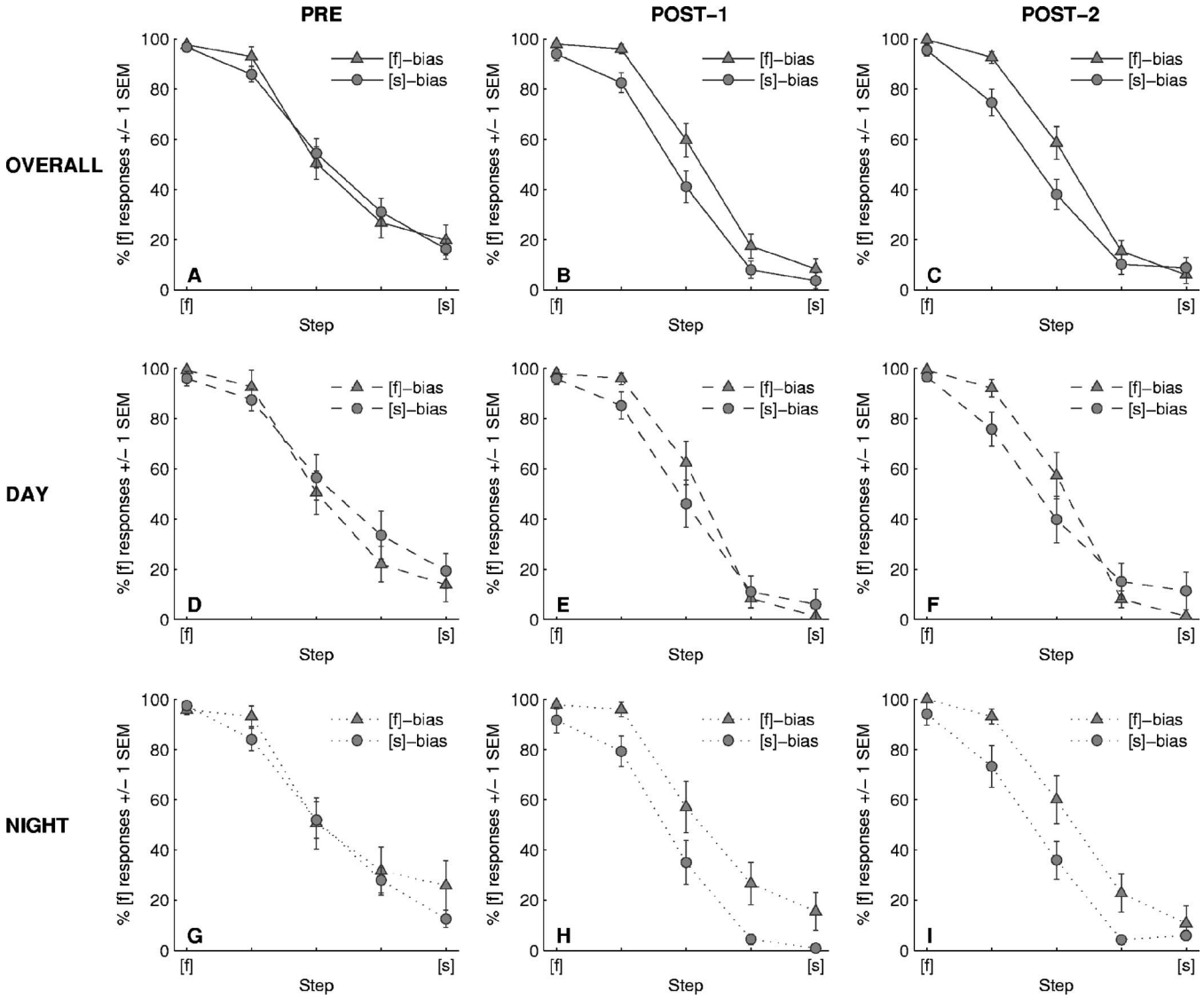


FIG. 1. Percentages of [f] responses to each of the five [f]–[s] steps for the groups with [f]-biased and [s]-biased exposure at pretest, post-test-1, and post-test-2. The top panels (A), (B), (C) show the mean performance collapsed across groups; the middle panels (D), (E), (F) show the day groups only; and the bottom panels (G), (H), (I) show the night groups.

B. Delay at night vs during the day

To test for a specific effect of sleep on post-test-2 performance, a further ANOVA comprised the factors time of exposure (9 am or 9 pm), test (pretest or post-test-2), lexical bias, and step. An effect of sleep on perceptual learning would be reflected in the interaction of time of exposure $\times$ test $\times$ lexical bias, which was not significant, $F(1,53)=0.15$, $p=0.699$. The test $\times$ lexical bias interaction was significant, $F(1,53)=4.20$, $p=0.045$. We also tested for a potential effect of time of exposure on immediate learning by conducting an equivalent ANOVA with the post-test-1 data. Again, the three-way interaction of time of exposure $\times$ test $\times$ lexical bias was not significant [$F(1,53)=0.26$, $p=0.615$], while the overall learning effect was significant [test $\times$ lexical bias, $F(1,53)=4.60$, $p=0.037$]. Thus, although the immediate learning effect was numerically stronger in the night group than in the day group (see Fig. 1), the difference between groups was not significant. The effect, once established, remained stable over the 12-h delay, both overnight and during the day, as shown by a lack of interaction of time of exposure $\times$ test $\times$ lexical bias in an equivalent ANOVA comparing post-tests 1 and 2 [$F(1,53)=0.05$, $p=0.823$].

IV. DISCUSSION

The results show an immediate perceptual learning effect after hearing an ambiguous fricative sound [?] in lexically biased contexts for a few minutes. In contrast to previous studies using a lexical decision task on a list of words and nonwords as the exposure phase (Eisner and McQueen, 2005; Kraljic and Samuel, 2005; Norris *et al.*, 2003), this lexically guided learning effect was observed here when exposure was listening to a short story and thus involved no decision task. Listeners who heard the ambiguous sound placed in words that favor its interpretation as an [f] labeled more sounds on an [f]–[s] continuum as [f] than they did before exposure to [?], while listeners who heard the same sound in [s]-biased contexts showed the reverse pattern. The effect remained robust after a 12-h interval: No change in

magnitude in either direction was observed (relative to the immediate post-test), both for the groups which had the opportunity for consolidation during sleep and received relatively little speech input from other talkers, and the groups which had no sleep and more contact with other talkers.

Fenn *et al.* (2003) showed that, for learning to understand synthetic speech, there is a decrease in performance during 12 h of waking but subsequent recovery during sleep. The lack of such a pattern in the present data suggests that the type of perceptual learning examined here is less susceptible to decay. In contrast to learning about synthetic speech, a perceptual adjustment to a talker idiosyncrasy is a very fast-occurring process in which listeners already are highly skilled, and of which they are therefore usually unaware. The perceptual system in this case is not learning a novel skill as such, but applying a subtle adjustment in the processing of a particular phoneme contrast. For this kind of learning to be helpful to the listener in benefiting subsequent recognition of the exposure talker's speech (Norris *et al.*, 2003), it ought to occur rapidly and remain stable, regardless of whether the listener is awake or asleep. Although learning to understand synthetic speech better presumably taps into existing prelexical adjustment routines, it is likely to also involve learning at other processing levels (e.g., the unusual prosody of the synthetic "talker"), all of which may be subject to unlearning during waking. This type of learning also takes time and effort (Greenspan *et al.*, 1988), and often requires explicit feedback during training. It is therefore quite possible that a more drastic distortion of the natural speech signal than the manipulation in the present experiment (e.g., affecting more than one phoneme contrast, or additional levels of processing) will also be more liable to the process of unlearning and recovery that Fenn *et al.* have demonstrated for synthetic speech.

The picture that is emerging for lexically driven perceptual adjustments in response to talker idiosyncrasies is that these remain very stable. Using a similar paradigm as the present study, Kraljic and Samuel (2005) have already shown that learning effects are reliable after a 25-min interval, unless listeners are exposed to unambiguous tokens of the critical sound that come from the voice of the exposure talker. Together with these results, the evidence at present suggests that, once the perceptual system has adjusted to a given talker, it does not return to its original state through either the effects of speech input from other talkers or the mere passage of time.

ACKNOWLEDGMENTS

We would like to thank Femke Verhees for help with constructing the stimulus materials and Dennis Pasveer for testing some of the participants. We also thank Arty Samuel and an anonymous reviewer for comments on a previous version of this manuscript.

de Saint-Exupéry, A. (**2001**). *De Kleine Prins* (Ad Donker, Rotterdam [original work published **1943**]).

Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**(2), 224–238.

Fenn, K. M., Nusbaum, H. C., and Margoliash, D. (**2003**). "Consolidation during sleep of perceptual learning of spoken language," Nature (London) **425**, 614–616.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (**1988**). "Perceptual learning of synthetic speech produced by rule," J. Exp. Psychol. Learn. Mem. Cogn. **14**(3), 421–433.

Kraljic, T., and Samuel, A. G. (**2005**). "Perceptual learning for speech: Is there a return to normal?" Cogn. Psychol. **51**, 141–178.

Ladefoged, P. (**1989**). "A note on 'Information conveyed by vowels'," J. Acoust. Soc. Am. **85**, 2223–2224.

Ladefoged, P., and Broadbent, D. E. (**1957**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.

McQueen, J. M. (**1991**). "The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity," J. Exp. Psychol. Hum. Percept. Perform. **17**, 433–443.

Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cogn. Psychol. **47**, 204–238.

Nygaard, L. C., and Pisoni, D. B. (**1998**). "Talker-specific learning in speech perception," Percept. Psychophys. **60**, 355–376.

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). "Speech perception as a talker-contingent process," Psychol. Sci. **5**(1), 42–46.

The method of waveform images for finite waveguides with resistive terminations subject to arbitrary initial conditions

Jerry H. Ginsberg
G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 7 June 2005; revised 11 October 2005; accepted 10 January 2005)

Reflection at normal incidence of a plane wave can be described by imaging the incident wave profile on the opposite side of the boundary. This concept has been introduced in a few texts, but only for nondissipative conditions. In this paper the procedure to describe a purely resistive boundary is generalized, and then the concept is extended to describe the transient response of a one-dimensional, finite length waveguide. The field generated by arbitrary initial conditions is characterized by an infinite number of images, which leads to a representation of the acoustic field as oppositely propagating waves in an unbounded waveguide. Both graphical and mathematical descriptions of these waves are derived, with the former shown to provide significant insights. Mathematical analysis of the image construction leads to identification of several fundamental acoustic phenomena, including acoustic modes and reverberation time. From an instructional viewpoint the ability to explore fundamental acoustic phenomena without recourse to solving differential equations makes the waveform image concept especially useful as an introductory tool. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2172167]

PACS number(s): 43.10.Sv, 43.20.Mv, 43.20.Ks [VWS] Pages: 1954–1960

I. INTRODUCTION

In currently popular acoustics texts, such as Blackstock (2000) or Pierce (1983), the first solution of the wave equation is the one associated with d'Alembert. This solution enables the student to be comfortable with the interchangeability of time and space for wave propagation processes without concurrently engaging in a sophisticated mathematical analysis. This solution usually is used to explain reflection from rigid and pressure-release surfaces, but it is abandoned when the topic is reflection from surfaces whose specific impedance is frequency dependent. However, the hypothetical case of a purely resistive surface having constant impedance in the frequency domain, which one could realize with a massless dashpot, can readily be described by applying the reflection coefficient to the d'Alembert solution.

Images of sources are commonly used to treat reflection processes involving planar rigid and pressure-release boundaries, and much work has been done to extend the simple concept to other boundary conditions. In the present work a different kind of image is employed, specifically a mirrored spatial profile of the incident wave. A time domain representation of transient reflections in a finite length one-dimensional waveguide will be described in terms of an infinite number of images that are scaled by the reflection coefficients. This will be seen to lead to partial wave series corresponding to waves propagating in opposite directions within a waveguide whose terminations have been removed.

One may find the concept of a waveform image in a few textbooks, such as Morse (1964) and Morse and Ingard (1968). Both use it to describe the free response of a stretched string on stationary supports, but the modification to treat acoustic waves when the boundary is rigid or pressure release is straightforward. Blackstock (2000) uses this type of image to describe a single reflection, but switches to tracking waveforms along characteristic lines when the topic is propagation in a closed waveguide. None of these texts consider dissipative boundary conditions. Much earlier, Eyring (1930) used the concept of a waveform image to explain some features of room acoustics. Miller (1951) offered modifications that led to a better match with measurements in a one-dimensional waveguide. In both cases the tendency was to use waveform images to explain, rather than derive, analytical representations of wave processes.

Many beneficial aspects may be found in using the development that follows as an instructional tool. Perhaps the most important of these is that it enables one to introduce the student to a variety of acoustical phenomena without simultaneously having to address the solution of differential equations. Following a review of the fundamental concept of a waveform image, the development addresses the transient acoustic field in a finite length waveguide generated by arbitrary initial conditions. This is done by graphically constructing spatial profiles of incident and image waves, which then are expressed as mathematical functions of two phase variables. An application of the general result to the case of nondissipative boundary conditions leads to the recognition that the field may alternatively be described as a superposition of normal mode contributions whose spatial dependence and natural frequencies are shown to be a direct consequence of the partial wave series. The corresponding application to the case of dissipative walls leads to the identification of a difference equation whose solution demonstrates that there is temporal exponential decay, but in a stepwise manner that was alluded to by Miller (1951). One byproduct of this application is a formula for the reverberation time that is very close to the Norris-Eyring formula (Pierce, 1983).

 0001-4966/2006/119(4)/1954/7/$22.50

A few aspects of the present development may be found in the referenced works, but none have disclosed the wealth of phenomena demonstrated here. This paper is limited as a research application because the notion of a frequency-independent, purely resistive impedance is quite artificial. One could introduce a generalized boundary model by closing the waveguide with a piston having a certain mass, and supported by a spring and dashpot. Because the present approach works in the time domain, a determination of the wave reflected from a single incident pulse in such a model would require the solution of a linear differential equation. This reflected wave would be the incident one for the next reflection in a closed waveguide, whose reflected wave would be obtained as another differential equation solution, *ad infintum*. Tracking the multiple reflections would likely overwhelm all but the most talented students. Of course, one could contemplate switching to the Laplace or frequency domain to evaluate the reflected wave, and then inverting back to the time domain to propagate it to the other end. Here too, one encounters a conflict with the notion of presenting a fundamental description of acoustic processes that requires only basic mathematical tools.

II. BASIC CONCEPT

The mathematical foundation for the method of waveform images is the d'Alembert plane wave solution of the wave equation. According to it, the acoustic pressure and particle velocity in an infinite waveguide are described by

$$p = p_1 + p_2, \quad v = v_1 + v_2, \quad p_1 = \rho_0 c v_1 = F(\theta_1),$$
$$p_2 = -\rho_0 c v_2 = G(\theta_2), \quad \theta_1 = x - ct, \quad \theta_2 = x + ct, \tag{1}$$

where the terms bearing subscript 1 and 2 propagate in the positive x and negative x directions, respectively. Note that positive v implies that the velocity is in the sense of increasing x.

Initial conditions that set the response for the infinite domain $-\infty < x < \infty$ are the spatial distribution of pressure and velocity,

$$p(x,0) = P(x), \quad v(x,0) = V(x). \tag{2}$$

Matching these conditions to the general solution, Eq. (1), leads to

$$F(x) = \tfrac{1}{2}[P(x) + \rho_0 c V(x)], \quad G(x) = \tfrac{1}{2}[P(x) - \rho_0 c V(x)]. \tag{3}$$

Now consider a semi-infinite waveguide, $0 < x < \infty$, with the end terminated by a frequency-independent real specific impedance Z, such as what one would obtain from a hypothetical massless dashpot. The boundary condition in this case is

$$p = -Zv \ @ \ x = 0. \tag{4}$$

The presence of a boundary does not alter the fact that the field is a superposition of two waves, so Eq. (1) still applies. Substitution of that representation into the preceding boundary condition, combined with the property that $\theta_1 = -ct$ and $\theta_2 = +ct$ at $x = 0$, leads to

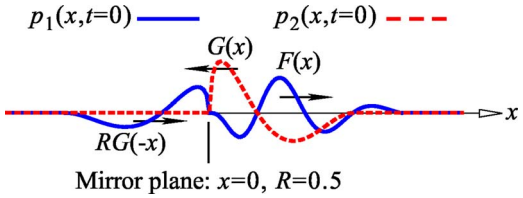


FIG. 1. (Color online) Initial pressure distributions of oppositely propagating waves in a waveguide extending over $x > 0$ when image waveforms are used to describe reflection.

$$p_1(-ct) = R p_2(ct), \quad R = \frac{Z - \rho_0 c}{Z + \rho_0 c}, \quad t > 0. \tag{5}$$

This relation is extended to any location by replacing $-ct$ with θ_1, which leads to the condition that the p_1 function for a negative value of θ_1 must be R times the p_2 function for the corresponding positive value of θ_1. Thus, the initial conditions set $p_1(\theta_1)$ and $p_2(\theta_2)$ for positive phase variables, while the reflection condition sets p_1 for negative θ_1, but not p_2. It follows that p_2 for negative θ_2 may be set arbitrarily. This leads to a global solution for each wave,

$$p_1(\theta_1) = \begin{cases} F(\theta_1), & \theta_1 > 0, \\ RG(-\theta_1), & \theta_1 < 0, \end{cases}$$
$$p_2(\theta_1) = \begin{cases} G(\theta_2), & \theta_2 > 0, \\ 0, & \theta_2 < 0. \end{cases} \tag{6}$$

The coefficient R in Eq. (5) is the familiar reflection coefficient encountered in a frequency domain analysis, except that the limitation to a real impedance means that R is a real value in the range $|R| \leqslant 1$. The description of p_2 for negative values of θ_2 is not strictly necessary, but doing so leads to an interpretation in terms of images. An alternative form of the preceding using step functions $h(\cdot)$ is

$$p_1(\theta_1) = F(x - ct)h(x - ct) + RG(-x + ct)h(-x + ct),$$
$$p_2(\theta_1) = G(x + ct)h(x + ct). \tag{7}$$

The preceding description of reflection is readily interpreted in terms of images, if one artificially extends the semi-infinite waveguide to $x < 0$. Then at $t = 0$ the spatial profile of p_1 for $x < 0$ is the incident signal p_2 plotted in the reversed direction, and scaled by the reflection coefficient. If $R = 1$ for a rigid end, then this portion of the p_1 profile is the mirror image, whereas $R = -1$ leads to a mirror image that is inverted relative to the pressure values. Initial spatial profiles of p_1 and p_2 for arbitrary initial conditions are depicted in Fig. 1.

The identification of these initial profiles facilitates an understanding of the propagation process. If one seeks to describe the spatial distribution of pressure when $t > 0$, the initial profiles of p_1 and p_2 are shifted by ct in the directions that they propagate, positive and negative x directions, respectively, and then added. This is the process depicted in Fig. 2, where the the initial p_1 in Fig. 1 is shifted to the right by ct, while the initial p_2 profiles is shifted to the left. Note that the sum of p_1 and p_2 is depicted only for $x > 0$, where the signal actually exists.

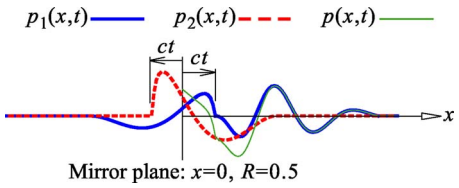


FIG. 2. (Color online) Construction of the spatial pressure profile in a waveguide extending over $x>0$ using images to describe reflection.

A similar procedure for determining the temporal waveform at a specified $x>0$ begins by evaluating the waveforms at $x=0$. Because $\theta_1=-ct$ at $x=0$ and $\theta_1=x$ at $t=0$, and $p_1=F(\theta_1)$, it follows that the temporal waveform of p_1 at $x=0$ is a function of $-ct$, whereas its initial spatial profile is the same function of x. Similar reasoning leads to the recognition that the temporal waveform of p_2 at $x=0$ is a function of ct, while its initial spatial profile is the same function of x. Thus, in Fig. 3(a) $p_1(0,t)$ at $x=0$ is described by reversing the sense in which $p_1(x,0)$ was plotted in Fig. 1, accompanied by scaling the abscissa by $1/c$, while $p_2(0,t)$ is plotted by applying the $1/c$ scaling factor to the previous plot of $p_2(x,0)$ without reversing the direction. Figure 3(b) shows that the time signature of p_1 at a specified $x>0$ is obtained by retarding $p_1(0,t)$ by x/c, while $p_2(0,t)$ is advanced by the same amount to obtain the signature of p_2 at $x>0$. The total waveform $p(x,t)$ then is obtained by adding the individual waveforms. The addition is depicted there only for $t>0$.

An important aspect of the overall development is that describing how the wave evolves spatially and temporally is fairly straightforward once the initial profiles $p_1(x,0)$ and $p_2(x,0)$ have been established for $-\infty<x<\infty$. The spatial dependencies at arbitrary t may be obtained by propagating the initial spatial profiles p_1 and p_2 by distance ct in opposite directions, then adding their values at each x. Temporal waveforms may be obtained by reversing the sense in which the initial spatial profile of p_1 is plotted, then retarding p_1 and advancing p_2 by time x/c. The development that follows will derive the same partial wave decomposition for the case of a finite length waveguide. Consequently, this notion of shifting and adding the initial spatial profile of partial waves to obtain instantaneous pressure distributions and temporal

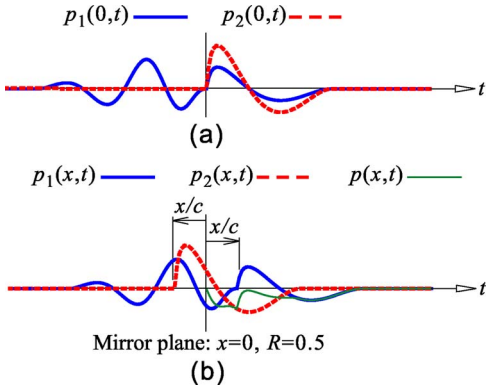


FIG. 3. (Color online) Construction of the temporal pressure waveform in a waveguide extending over $x>0$ using images to describe reflection. (a) The individual waveforms at $x=0$. (b) Retardation and advancement of the waveforms at $x>0$.

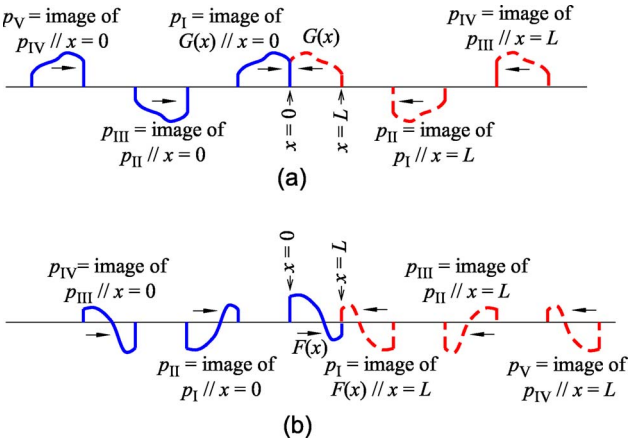


FIG. 4. (Color online) Construction of the initial spatial pressure profile in a waveguide when $R_0=1$ and $R_L=-1$. Successive waveform images are denoted as p_n. (a) Images generated by an initial disturbance that propagates in the negative x direction, (b) images generated by an initial disturbance that propagates in the positive x direction.

waveforms is equally applicable for waveguides. For this reason the emphasis henceforth is on the determination of the initial profiles.

III. FINITE WAVEGUIDE

Reflections occur at both ends when there is an initial pressure or velocity distribution in a closed waveguide of length L. Let R_0 and R_L denote the reflection coefficients at $x=0$ and $x=L$, respectively. The initial pressure distributions defined by Eq. (3) now apply solely for $0<x<L$. Correspondingly, any reference to $F(x)$ and $G(x)$ henceforth should be understood to pertain to functions that are nonzero only for $0<x<L$. It is easier to consider an initial profile in which $p_2=G(x)$ and $p_1=0$, after which the case where the initial state is one where $p_1=F(x)$ and $p_2=0$ will be addressed.

Figure 4(a) describes a situation where $x=0$ is terminated rigidly, $R=1$, and $x=L$ is a pressure-release boundary, $R=-1$. The initial wave associated with $G(x)$ propagates in the negative x direction, so it is reflected at $x=0$. This reflection is described by image p_I that propagates in the positive x direction. The consequence of the waveguide length being finite is that this image will be reflected at $x=L$. This reflection is described by image p_{II} that propagates in the negative x direction. Because the front of p_I leaves $x=0$ when $t=0$, it will arrive at $x=L$ when $t=L/c$. It is necessary that p_{II} arrive at $x=L$ synchronously with p_I, so its wave front when $t=0$ must be at $x=2L$. Another image p_{III} propagating in the positive x direction is required to satisfy the impedance boundary condition at $x=0$ when p_{II} arrives there. Thus, the front of p_{III} must be at $x=-2L/c$ when $t=0$. In turn, p_{III} will be incident at $x=L$, which requires another image p_{IV} that propagates in the negative x direction. At $t=0$, the distance from the front of p_{III} to the boundary at $x=L$ is $3L$, so the front of p_{III} when $t=0$ is at $x=4L$. This construction process leads to an infinite number of images. The initial spatial profile of the negative propagating wave p_2 consists of $G(x)$ and similarly shaped images that are shifted in the positive x direction by constant incremental distance $2L$. The p_1 profile

is composed of reflected images that are situated in $x<0$ and also are separated at a constant increment $2L$.

A similar construction describing the reflections associated with the $F(x)$ portion of the initial pressure distribution is depicted in Fig. 4(b). Because $F(x)$ propagates in the positive x direction, its reflection at $x=L$ requires an image p_I propagating in the negative x direction, with its wave front at $x=L$ when $t=0$. In turn, the incidence of p_I at $x=0$ requires image p_{II} that propagates in the positive x direction, and whose wavefront is initially situated at $x=-L$. The reflection of this image at $x=L$ requires image p_{III} that propagates in the negative x direction with its wavefront initially at $x=3L$. This construction also continues *ad infinitum*.

In the situation in Fig. 4, $R_0=1$ and $R_L=-1$. Correspondingly, images associated with reflection at $x=0$ are mirrored, while images at $x=L$ are inverted and mirrored. In general, the reflection coefficient scales the image relative to the incident wave with which it is associated. Let $p_G(x)$ and $p_F(x)$ denote the initial pressure profiles associated with $G(x)$ and $F(x)$, as described by Figs. 4(a) and 4(b), respectively. An analytical description of these profiles may be obtained by applying the appropriate reflection coefficient and two fundamental properties. A function that is defined over a finite interval may be described by a pair of step functions that turn the function on and off at the appropriate locations, that is, $f(x)[h(x-b)-h(x-a)]$ has the value of $f(x)$ for $a<x<b$, while it is zero outside that interval. Also, the profile of an image associated with reflection at $x=0$ is described by replacing the argument x of the incident wave with $-x$, which is the distance behind the reflection boundary. Thus, if $f(x)$ is the function for such a wave in $x>0$, the image will be $R_0 f(-x)$ in $x<0$. In the same manner, images associated with reflection at $x=L$ are obtained by replacing $L-x$, which is the distance in front of the boundary, where $x<L$, with $x-L$ for the distance behind the boundary, where $x>L$. Suppose $g(x)$ for $x<L$ is the function for a wave incident at $x=L$. This function may be written as $g[L-(L-x)]$. Replacing $L-x$ in this function with $x-L$, accompanied by multiplication by the reflection coefficient R_L, gives the function describing the image, specifically $R_L g(2L-x)$. These considerations lead to

$$\begin{aligned} p_G(x) = {} & G(x)[h(x)-h(x-L)] + R_0\{G(-x)[h(x+L) \\ & - h(x)]\} + R_L\{R_0 G(x-2L)[h(x-2L) \\ & - h(x-3L)]\} + R_0\{R_L R_0 G(-x-2L)[h(x+3L) \\ & - h(x+2L)]\} + R_L\{R_0^2 R_L G(x-4L)[h(x-4L) \\ & - h(x-5L)]\} + \cdots , \end{aligned} \tag{8}$$

$$\begin{aligned} p_F(x) = {} & F(x)[h(x)-h(x-L)] + R_L\{F(-x+2L)[h(x-L) \\ & - h(x-2L)]\} + R_0\{R_L F(x+2L)[h(x+2L) \\ & - h(x+L)]\} + R_L\{R_0 R_L F(-x+4L)[h(x-3L) \\ & - h(x-4L)]\} + R_0\{R_0 R_L^2 F(x+4L)[h(x+4L) \\ & - h(x+3L)]\} + \cdots . \end{aligned} \tag{9}$$

Note that all images in $p_G(x)$ situated in $x<0$ propagate in the positive x direction, and therefore are associated with p_1, while the terms in $x>0$ are associated with p_2. For $p_F(x)$ the terms that are associated with propagation in the positive x direction are situated in $x<L$, while the wave that propagates in the negative x direction is formed from the images that are situated in $x>L$. A unified picture of the pressure within a finite waveguide results from combining the descriptions of p_F and p_G in Eqs. (8) and (9). In doing so, p_1 receives those terms that are associated with positive x propagation, while the terms associated with propagation in the negative x direction constitute p_2. The result is

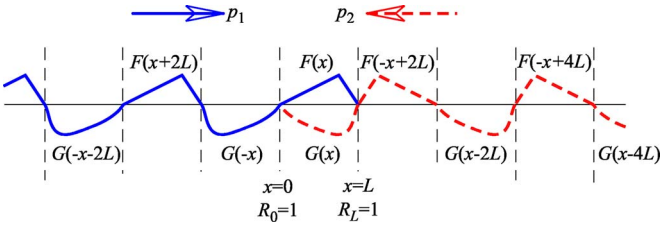


FIG. 5. (Color online) Assembly of the initial pressure disturbance and images in a closed waveguide into oppositely traveling waves. Both ends are rigid.

$$\begin{aligned} p_1(x,0) = {} & F(x)[h(x)-h(x-L)] + R_0 G(-x)[h(x+L) \\ & - h(x)] + R_0 R_L F(x+2L)[h(x+2L)-h(x+L)] \\ & + R_0^2 R_L G(-x-2L)[h(x+3L)-h(x+2L)] \\ & + R_0^2 R_L^2 F(x+4L)[h(x+4L)-h(x+3L)] + \cdots , \end{aligned} \tag{10}$$

$$\begin{aligned} p_2(x,0) = {} & G(x)[h(x)-h(x-L)] + R_L F(-x+2L) \\ & \times [h(x-L)-h(x-2L)] + R_0 R_L G(x-2L) \\ & \times [h(x-2L)-h(x-3L)] + R_0 R_L^2 F(-x+4L) \\ & \times [h(x-3L)-h(x-4L)] + R_0^2 R_L^2 G(x-4L) \\ & \times [h(x-4L)-h(x-5L)] + \cdots . \end{aligned}$$

An examination of these expressions leads to the recognition of a compact summation form. The description of the signal at an arbitrary instant is then obtained by replacing x with $x-ct$ in $p_1(x,0)$, while x is replaced by $x+ct$ in $p_2(x,0)$. This leads to

$$\begin{aligned} p_1(x,t) = {} & \sum_{n=0}^{\infty} R_0^n R_L^n \{F(x-ct+2nL)[h(x-ct+2nL) \\ & - h(x-ct+2nL-L)] + R_0 G(-x+ct-2nL) \\ & \times [h(x-ct+2nL+L)-h(x-ct+2nL)]\}, \end{aligned} \tag{11}$$

$$\begin{aligned} p_2(x,t) = {} & \sum_{n=0}^{\infty} R_0^n R_L^n \{G(x+ct-2nL)[h(x+ct-2nL)-h(x \\ & + ct-2nL-L)] + R_L F(-x-ct+2nL+2L) \\ & \times [h(x+ct-2nL-L)-h(x+ct-2nL-2L)]\}. \end{aligned}$$

Although the preceding expressions are useful for quantitative evaluations, pictures of the initial profile for a given circumstance are more conducive to gaining physical insight. The case where both walls of the waveguide are rigid, $R_0=R_L=1$, is depicted in Fig. 5. This figure shows that p_1 and

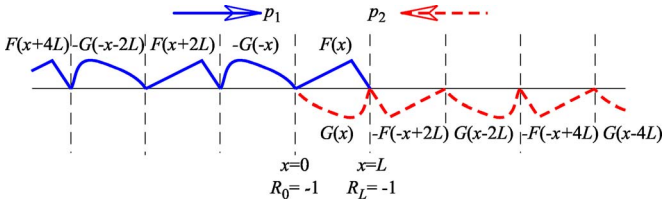


FIG. 6. (Color online) Assembly of the initial pressure disturbance and images in a closed waveguide into oppositely traveling waves. Both ends are pressure release.

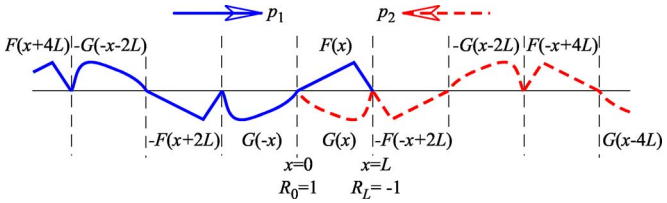


FIG. 7. (Color online) Assembly of the initial pressure disturbance and images in a closed waveguide into oppositely traveling waves. The left end is rigid and the right end is pressure release.

p_2 are periodic, and that their profiles are related, such that

$$p_1(x,0) = p_1(x-2L,0); \quad x < L,$$

$$p_2(x,0) = p_2(x+2L,0); \quad x > 0, \tag{12}$$

$$p_2(x,0) = p_1(-x,0); \quad x > 0.$$

At $t=0$, the respective phase variables are $\theta_1 = \theta_2 = x$, from which it follows that both p_1 and p_2 are representable in Fourier series form as

$$p_1(x,t) = \frac{1}{2}\sum_{n=-\infty}^{\infty} P_n \exp\left(\frac{in\pi\theta_1}{L}\right), \quad P_{(-n)} = P_n^*; \quad \theta_1 < L,$$

$$p_2(x,t) = \frac{1}{2}\sum_{n=-\infty}^{\infty} P_n \exp\left(-\frac{in\pi\theta_2}{L}\right); \quad \theta_2 > 0. \tag{13}$$

Correspondingly, the pressure signal p, which is the sum of p_1 and p_2, is representable as a Fourier series, valid only for $0<x<L$. Substitution of the definitions of θ_1 and θ_2 results in

$$p(x,t) = \frac{1}{2}\sum_{n=-\infty}^{\infty} P_n\left[\exp\left(\frac{in\pi x}{L}\right) + \exp\left(-\frac{in\pi x}{L}\right)\right] \times \exp\left(-\frac{in\pi ct}{L}\right)$$
$$\equiv \sum_{n=-\infty}^{\infty} P_n \cos\left(\frac{n\pi x}{L}\right)\exp\left(-\frac{in\pi ct}{L}\right). \tag{14}$$

Thus, the method of images leads one to recognize that the signal generated by initial conditions within a closed, rigid-rigid, waveguide is composed of an infinite series of modes whose spatial distribution is an integer multiple of half-cosines, with natural frequencies that are integer multiples of the fundamental frequency $\pi c/L$.

Similar constructions lead to modal descriptions for the other combinations of nondissipative boundary conditions. Figure 6 describes the initial profiles when both ends are pressure release. In this case the initial profiles are also periodic in an interval of $2L$, but the inversion associated with reflection at a pressure-release termination gives

$$p_1(x,0) = p_1(x-2L,0); \quad x < L,$$

$$p_2(x,0) = p_2(x+2L,0); \quad x > 0, \tag{15}$$

$$p_2(x,0) = -p_1(-x,0); \quad x > 0.$$

A Fourier series expansion of p_1+p_2 like the one in the rigid-rigid case now leads to

$$p(x,t) = \frac{1}{2}\sum_{n=-\infty}^{\infty} P_n\left[\exp\left(\frac{in\pi x}{L}\right) - \exp\left(-\frac{in\pi x}{L}\right)\right] \times \exp\left(-\frac{in\pi ct}{L}\right)$$
$$\equiv \sum_{n=-\infty}^{\infty} iP_n \sin\left(\frac{n\pi x}{L}\right)\exp\left(-\frac{in\pi ct}{L}\right); \quad 0 < x < L. \tag{16}$$

One can recognize from the preceding that the initial response within a waveguide whose both ends are pressure release is composed of an infinite number of modes whose spatial dependence is an integer number of lobes of a sine, and that the natural frequencies are the same as those when both ends are rigid.

The construction of the profiles $p_1(x,0)$ and $p_2(x,0)$ for the case where one end is rigid and the other is pressure release is described by Fig. 7. Images are inverted relative to the incident wave only at the end where $R=-1$, so an image must be reflected twice from that end before it is returned to its original appearance. Thus, each profile repeats in an interval of $4L$. Also, it is evident that $p_2(x)$ for $x>0$ equals $p_1(-x)$. Thus,

$$p_1(x,0) = -p_1(x-2L) = p_1(x-4L); \quad x < L,$$

$$p_2(x,0) = -p_2(x+2L) = p_2(x+4L); \quad x > 0, \tag{17}$$

$$p_2(x,0) = p_1(-x,0); \quad x > 0.$$

The Fourier series that result from these properties is like Eq. (14), except that the period is doubled and only odd harmonics are nonzero,

$$p(x,t) \equiv \sum_{\substack{n=-\infty \\ n \text{ odd}}}^{\infty} P_n \cos\left(\frac{n\pi x}{2L}\right)\exp\left(-\frac{in\pi ct}{2L}\right), \quad 0 < x < L. \tag{18}$$

Here, the image construction leads to the recognition that the modes consist of an odd multiple of quarter-cosines, with the natural frequencies being that multiple of the fundamental $\pi c/2L$.

When either end is dissipative, there is no spatial periodicity in the initial profiles $p_1(x,0)$ and $p_2(x,0)$, as can be seen in Fig. 8. Instead, the profile in each interval of $2L$ is attenuated by a factor of R_0R_L relative to the previous interval, such that

$$\begin{aligned} p_1(x-2L,0) &= R_0R_Lp_1(x,0); \quad x < L, \\ p_2(x+2L,0) &= R_0R_Lp_2(x,0); \quad 0 < x < 2L. \end{aligned} \tag{19}$$

These difference equations describe a decay given by

$$\begin{aligned} p_1(x-2nL,0) &= (R_0R_L)^n p_1(x,0); \quad -L < x < L, \\ p_2(x+2nL,0) &= (R_0R_L)^n p_2(x,0); \quad 0 < x < 2L. \end{aligned} \tag{20}$$

At $t=0$, both phase variables equal x, so general expressions for p_1 and p_2 are obtained by replacing x by the respective phase variable. The integer function floor(u), which is the nearest integer less than u, enables one to determine the number of periods corresponding to an arbitrary value of a phase variable relative to the reference intervals, $-L<\theta_1<L$, and $0<\theta_2<2L$. This leads to a description of the evolution of each wave according to

$$\begin{aligned} p_1(\theta_1) &= (R_0R_L)^{\text{floor}[(L-\theta_1)/2L]}p_1(\theta_1 \\ &\quad - 2L\ \text{floor}[(L-\theta_1)/2L]), \quad \text{if } \theta_1 < -L, \\ p_2(\theta_2) &= (R_0R_L)^{\text{floor}[\theta_2/2L]}p_2(\theta_2 \\ &\quad - 2L\ \text{floor}[\theta_2/2L]), \quad \text{if } \theta_2 > 2L. \end{aligned} \tag{21}$$

These relations describe how one may extend the representations of $p_1(\theta_1)$ for $-L<\theta_1<L$, and of $p_2(\theta_2)$ for $0<\theta_2<2L$, which are defined by the initial conditions and first reflections, as described by Fig. 8.

If one considers values of t that are large compared to L/c, and ignores the stepwise nature of the decay described by Eq. (21), both floor function values may be approximated as $ct/2L$. This leads to recognition of an approximate exponential decay factor, such that

$$\begin{aligned} p_1(\theta_1) &= \exp(-\beta t)p_1(\theta_1 - 2L\ \text{floor}[(L-\theta_1)/2L]), \\ &\quad \text{if } \theta_1 < -L, \\ p_2(\theta_2) &= \exp(-\beta t)p_2(\theta_2 - 2L\ \text{floor}[\theta_2/2L]), \\ &\quad \text{if } \theta_2 > 2L, \end{aligned} \tag{22}$$

where the decay constant is

$$\beta = -\frac{c}{2L}\ln(|R_0R_L|). \tag{23}$$

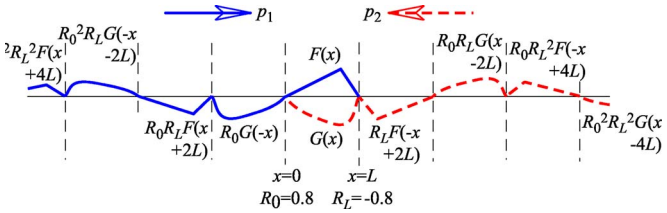


FIG. 8. (Color online) Assembly of the initial pressure disturbance and images in a closed waveguide into oppositely traveling waves. Both ends are dissipative.

In addition to providing a general picture for the exponential decay associated with dissipation, the preceding leads one to the concept of reverberation time T_{rev}, which is the time interval required for the mean-squared pressure to be attenuated by a factor of 10^{-6}. This corresponds to a decay factor of 0.001 for the instantaneous pressure amplitude. Equating the exponential factor in Eq. (22) to 0.001 leads to

$$T_{\text{rev}} = -\frac{6\ln(10)}{\ln|R_0R_L|}\frac{L}{c}. \tag{24}$$

This form is reminiscent of the Norris-Eyring reverberation time (Pierce, 1983) in the one-dimensional case.

IV. CLOSING REMARKS

The utility of using waveform images to construct the transient field in a finite length, one-dimensional waveguide is not that it provides a general analytical tool. Rather, doing so enables one to demonstrate to the beginning student many key features of waveguides and enclosures in a manner that does not require the formal mathematical solution of partial differential equations. For example, it is possible to demonstrate the periodic nature of the transient response when the walls are rigid or pressure release, with a further analysis of the imaged waveforms leading to identification of the various possible mode functions and associated natural frequencies. The application of the procedure to the case where one or both boundaries are dissipative discloses how the transient signal is attenuated, which leads to the concept of reverberation time. Another benefit of the present perspective is an appreciation for the fact that whereas a modal series obtained from a separation of variables solution of the wave equation features an infinite number of standing waves, there may be a parallel simpler representation of a transient response as a few traveling waves. At the same time, the avoidance of mathematical sophistication in the analysis limits it to cases where the ends are purely resistive. One can readily demonstrate to students that a frequency-independent impedance factor is unrealistic, which would tend to motivate the student toward the sophisticated mathematical approaches typically encountered in acoustics texts.

The key property that makes the image construction procedure accessible is the nondispersive nature of a plane wave. This observation suggests that decomposition of multidimensional waves in a rectangular waveguide into their planar components might make it possible to extend the

waveform image procedure to treat such systems. At the same time, the present approach is not advocated as a general procedure for reflection phenomena. As was noted earlier, one could contemplate employing the formulation to describe reflection when the surface pressure and normal velocity are related by a differential equation in the time domain, or a frequency-dependent impedance in the frequency domain. This would lead to a reflected image that is the particular solution of an ordinary differential equation whose inhomogeneity depends on the properties of the incident wave (Mintzer, 1950), or the need to switch back and forth between the time and Laplace or frequency domains. Having an image that depends on the functional nature of the incident wave, rather than merely scaling it, would substantially complicate the formulation. In any event such an application would be contrary to the intended purpose, which is to introduce students to some fundamental acoustic phenomena before they are confronted with sophisticated mathematical tools.

Blackstock, D. T. (**2000**). *Fundamentals of Physical Acoustics* (Wiley-Interscience, New York), pp. 4–14, 117–118.

Eyring, C. F. (**1930**). "Reverberation time in 'dead' rooms," J. Acoust. Soc. Am. **1**, 217–241.

Miller, H. B. (**1951**). "The discontinuous growth and decay of sound in a one-dimensional room," J. Acoust. Soc. Am. **23**, 517–523.

Mintzer, D. (**1950**). "Transient sounds in rooms," J. Acoust. Soc. Am. **22**, 217–241.

Morse, P. M. (**1964**). *Vibration and Sound* (McGraw-Hill, New York, reprinted by the Acoustical Society of America, Melville, NY), pp. 71–80.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York, reprinted by Princeton University Press), pp. 106–115.

Pierce, A. D. (**1983**). *Acoustics: An Introduction To Its Physical Principles and Applications* (McGraw-Hill, New York, reprinted by the Acoustical Society of America, Melville, NY), pp. 20–22, 259–265.

High order approximate low frequency theory of elastic anisotropic lining and coating

D. D. Zakharov[a)]
FESBE, London South Bank University, 103 Borough Road, London SE1 0AA, United Kingdom

(Received 30 May 2005; revised 16 December 2005; accepted 8 January 2006)

A problem of the dynamic behavior of an elastic layer coupled to one or two thick elastic solids is considered. All the materials may possess a general anisotropy and the layer is assumed to be thin enough with respect to the characteristic wavelength. Introducing the asymptotic power series with respect to the thickness-over-wavelength ratio for the main quantities and using the asymptotic integration method the displacements and stresses on the layer surfaces are related. Thus, the so-called impedance boundary conditions (IBC) are deduced for three cases—for a coated substrate with given displacements or with given stresses on the surface and for two substrates with a layer in between. In contrast to previous papers these IBC are obtained for the most general situation with the asymptotic accuracy up to the sixth order, uniform with respect to the representation of the displacements and stresses. Presented theory can be used for studying the surface and interface phenomena as well as for calculating fields and spectra of layered solids. The results are validated numerically and compared with those of other authors.
 [DOI: 10.1121/1.2169922]

PACS number(s): 43.20.Bi, 43.20.El, 43.20.Ks, 43.20.Tb [LLT] Pages: 1961–1970

I. INTRODUCTION

Many researchers in NDE, seismology, electronic devices, and other industrial applications have investigated the ultrasonic wave propagation in layered solids. The general solution to this problem is provided by Thompson[1] and Haskel,[2] whose method is based on the recurrent formulas for the propagator matrices relating the magnitudes of waves in the neighboring layers and permitting one to obtain the reflection or transmission coefficients and dispersion relations. The alternative—and numerically more stable—method of the determinant decomposition is introduced and then modified for the anelastic media by Knopoff[3,4] and Schwab.[4] From the first glance the problem addressed is solved and no longer of interest. However, for the variety of applications[5,6] which deal with the limited frequency band the consuming full analysis is not required. When the layer half-thickness h is much smaller than the minimal scale L of process (wavelength), i.e., $\varepsilon = h/L \ll 1$, the approximate layer model may accelerate the calculations with satisfactory accuracy. In particular, the layer influence can be replaced by introducing the effective boundary conditions on the interface. In what follows we use the term *impedance boundary conditions* (IBC) due to the first paper of Leontovich,[7] who introduced a similar philosophy in electromagnetics.

The equivalent boundary conditions for anisotropic layer in between two solids are constructed by Rokhlin *et al.*[8–10] Their low order analog for the wave propagation across a thin nonlinear elastic layer is developed by Sadler *et al.*[11] Bovik[12,13] suggested the asymptotic boundary conditions for the surface and interface waves guided by a thin layer. The approximate calculation of Lamb's wave spectrum and Green's function is performed by Niklasson *et al.*[14–16] They use the approximate boundary conditions for coated plate and for sandwich plate with a thin interface layer. All these papers deal with the expansion of the propagator matrix in ε-power series with the truncation error $O(\varepsilon)$ or $O(\varepsilon^2)$ as $\varepsilon \to +0$. Next iterations seem to be rather complicated. In the present paper an alternative way to deduce IBC of high order is suggested. The displacements and stresses are sought in the form of asymptotic ε-power series. Assuming then at least one of the three-dimensional (3D) boundary conditions is formulated in terms of displacements the IBC are derived by asymptotic integration[17] of a 3D dynamic boundary value problem of elasticity. This method is well developed in the theory of thin plates and shells[18–22] and provides a very efficient way to deduce the relations of any order. A simplified version of a similar method with expansion in power series of transversal coordinate is applied by Johansson *et al.*[23] to the derivation of IBC for two-dimensional (2D) problem of isotropic piezoelectric coating with a stress-free surface within the third-order accuracy (the first two terms are presented in a closed form).

Another method to approximate waves propagating as $e^{i(kX-\omega T)}$ in layered solids is developed by Wang and Rokhlin[24,25] and based on the recursive relations to deduce transfer matrices or stiffness matrices. The results presented below differ by high order terms from those in Refs. 24 and 25 obtained using the Pade approximation together with the Magnus expansion. Our internal asymptotics is uniform with respect to the layer displacements and stresses and hold the same truncation error for both of them.

The paper is organized as follows. In Sec. II the physical and mathematical statement of the problems is discussed. Section III is devoted to the asymptotic approach. In Sec. IV the impedance boundary relations are derived from 3D anisotropic elasticity. Section V presents the results of high

a)Electronic mail: dd_zakh@mail.ru

order. In Sec. VI we conclude with IBC within the sixth order for lining and coating with different boundary conditions on the surface. The numerical validation is performed in Sec. VII, discussed in Sec. VIII together with the analysis of transfer matrices and summarized in the Conclusion.

II. STATEMENT OF THE PROBLEM FOR A SINGLE LAYER

Consider a thin elastic layer occupying a region $-\infty<X_1,\ X_2<\infty,\ |Z|\leqslant h$ where $\mathbf{X}\equiv X_\alpha \mathbf{i}_\alpha$ $(\alpha=1,2)$ and $Z\equiv X_3$ are Cartesian coordinates. Denote the respective displacements of the layer particle along the axes as $\mathbf{U}\equiv U_\alpha \mathbf{i}_\alpha + W\mathbf{i}_3$ $(W\equiv U_3)$. The tensors of stresses $\underline{\underline{\sigma}}$ and strains $\underline{\underline{\varepsilon}}$ are related by Hooke's law of general anisotropy and satisfy the equations of 3D dynamic elasticity[26]

$$\begin{bmatrix}\sigma_{11}\\ \sigma_{22}\\ \sigma_{33}\\ \sigma_{23}\\ \sigma_{13}\\ \sigma_{12}\end{bmatrix}=\begin{bmatrix}c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16}\\ c_{21} & c_{22} & c_{23} & c_{24} & c_{25} & c_{26}\\ c_{31} & c_{32} & c_{33} & c_{34} & c_{35} & c_{36}\\ c_{41} & c_{42} & c_{43} & c_{44} & c_{45} & c_{46}\\ c_{51} & c_{52} & c_{53} & c_{54} & c_{55} & c_{56}\\ c_{61} & c_{62} & c_{63} & c_{64} & c_{65} & c_{66}\end{bmatrix}\begin{bmatrix}\varepsilon_{11}\\ \varepsilon_{22}\\ \varepsilon_{33}\\ \gamma_{23}\\ \gamma_{13}\\ \gamma_{12}\end{bmatrix},$$

$$\mathbf{C}\equiv\|c_{mn}\|,$$

$$\varepsilon_{pp}=\partial_p U_p,\quad \gamma_{pq}=\partial_p U_q+\partial_q U_p,\ \rho\partial_T^2 U_p=\partial_q\sigma_{pq}.$$

Here and in what follows T denotes time and ρ is the layer mass density.

Assume that at one of the surfaces $Z=\mp h$ at least one of the boundary conditions contains the boundary values of displacements U_p^- or U_p^+ in the explicit form. The problem is to relate all the boundary values of stresses $\sigma_{p3}^{\mp}$ and displacements $U_p^{\mp}$.

III. THE ASYMPTOTIC APPROACH

Introduce a small parameter $\varepsilon=h/L$, where L is a minimal scale (wavelength) in the longitudinal direction. Assuming the dynamic process to be essentially long-wave the displacements and stresses are sought in the from of asymptotic series

$$U_p=h\varepsilon^\lambda\{u_p^0+\varepsilon u_p^1+\cdots\},$$
$$\sigma_{pq}=\rho_0 c_0^2\varepsilon^\lambda\{\sigma_{pq}^0+\varepsilon\sigma_{pq}^1+\cdots\}, \tag{1}$$

where $E_0=\rho_0 c_0^2$ is a reference elastic modulus (e.g., the maximal modulus in matrix $\mathbf{G}$), ρ_0 is a reference mass density and c_0 is a reference speed of elastic wave. In what follows let us use the dimensionless variables $\mathbf{x}=\mathbf{X}/L$, $z=Z/h$, $t=T/T_0$ $(T_0=\varepsilon^{\tau-1}L/c_0)$, normalize all the elements of the stiffness matrix by E_0 and the mass density—by ρ_0. The possible values of λ and τ are considered below. For the convenience introduce additional notations for the symmetrical and antisymmetrical components of the displacements and stresses and a few useful matrices—minors of the global stiffness matrix $\mathbf{C}$

$$\mathbf{d}\equiv\begin{bmatrix}w\\ u_2\\ u_1\end{bmatrix},\quad \mathbf{t}_z\equiv\begin{bmatrix}\sigma_{33}\\ \sigma_{23}\\ \sigma_{13}\end{bmatrix},\quad \mathbf{t}_x\equiv\begin{bmatrix}\sigma_{11}\\ \sigma_{12}\\ \sigma_{22}\end{bmatrix},\quad \boldsymbol{\sigma}\equiv\begin{bmatrix}\sigma_{11}\\ \sigma_{12}\\ \sigma_{22}\\ \sigma_{23}\\ \sigma_{13}\end{bmatrix},$$

$$\mathbf{d}_\pm\equiv\frac{1}{2}(\mathbf{d}^+\pm\mathbf{d}^-),\quad \mathbf{t}_\pm\equiv\frac{1}{2}(\mathbf{t}_z^+\pm\mathbf{t}_z^-),$$

$$\mathbf{G}_0\equiv\mathbf{C}_{345}^{345},\quad \mathbf{G}_*\equiv\mathbf{C}_{345}^{162},\quad \mathbf{G}_\perp\equiv\mathbf{C}_{162}^{16245},\quad \mathbf{G}_\parallel\equiv\mathbf{C}_{345}^{16245}.$$

In these minors the upper indices correspond to rows and the lower ones denote columns whose elements are included. Similar quantities naturally arise in the dynamic elasticity of thin and thick anisotropic solids.[21,27–29] In what follows the indexed variables $\mathbf{d}^s$, $\mathbf{t}_z^s$, $\mathbf{t}_x^s$, $\boldsymbol{\sigma}^s$, $\mathbf{d}_\pm^s$, and $\mathbf{t}_\pm^s$ correspond to the respective components in asymptotic series (1). Hence, the coefficients of power ε^s in series (1) satisfy the following set of relations:

$$\partial_z\mathbf{t}_z^s=-\mathbf{D}\boldsymbol{\sigma}^{s-1}+\rho\partial_t^2\mathbf{d}^{s-4+2\tau}\quad\text{or}$$
$$\partial_z^2\mathbf{d}^s=\mathbf{G}_0^{-1}(-\mathbf{D}_1\partial_z\mathbf{d}^{s-1}+\mathbf{A}_2\mathbf{d}^{s-4+2\tau}), \tag{2}$$

$$\mathbf{t}_z^s=\mathbf{G}_0\partial_z\mathbf{d}^s+\mathbf{G}_\parallel^T\mathbf{D}^T\mathbf{d}^{s-1},\quad \mathbf{t}_x^s=\mathbf{G}_*\partial_z\mathbf{d}^s+\mathbf{G}_\perp^T\mathbf{D}^T\mathbf{d}^{s-1}, \tag{3}$$

where

$$\mathbf{D}\equiv\begin{bmatrix}0 & 0 & 0 & \partial_2 & \partial_1\\ 0 & \partial_1 & \partial_2 & 0 & 0\\ \partial_1 & \partial_2 & 0 & 0 & 0\end{bmatrix},$$
$$\mathbf{D}_1\equiv\mathbf{D}\mathbf{G}_\parallel+\mathbf{G}_\parallel^T\mathbf{D}^T,\quad \mathbf{D}_2\equiv\mathbf{D}\mathbf{C}_{16245}^{16245}\mathbf{D}^T,\quad \mathbf{A}_2\equiv\rho\partial_t^2-\mathbf{D}_2, \tag{4}$$

The maximal value of τ for which Eqs. (2) remain recurrent and consistent is $\tau=1$. Hence, the time scale T_0 corresponds to the shortest wave in the longitudinal direction.

Remark 1. When the boundary values of displacements $\mathbf{d}^{\mp}$ or stresses $\mathbf{t}_z^{\mp}$ are given and have the order $O(1)$ as $\varepsilon\to+0$, the set of relations (2) and (3) should be completed by

$$\mathbf{d}^s=\mathbf{d}^{\mp}\delta_0^{\lambda+s+1}\quad\text{or}\quad \mathbf{t}_z^s=\mathbf{t}_z^{\mp}\delta_0^{\lambda+s}, \tag{5}$$

where δ_n^k is the Kronecker delta. They can be used to determine the degree λ.

Substituting $s=0,1,2,\ldots$ into relations (2) and (3), and when necessary, into (5), one can carry out an asymptotic integration[17] seeking the respective terms in asymptotic series step by step. As one can see, the relations change when the index s steps by 1 which is in contrast to the theory of thin plates where such a step equals 2.

IV. LOW-ORDER BOUNDARY RELATIONS

To begin with, let $s=0$. Then the expressions (2) and (3) yield $\mathbf{d}^0=\mathbf{d}_0^0(\mathbf{x},t)+z\mathbf{d}_1^0(\mathbf{x},t)$, $\mathbf{t}_z^0=\mathbf{G}_0\mathbf{d}_1^0$, $\mathbf{t}_x^0=\mathbf{G}_*\mathbf{d}_1^0$, and after simple transformations we obtain

$$\mathbf{d}^0=\mathbf{d}_+^0+z\mathbf{d}_-^0,\quad \mathbf{t}_z^0=\mathbf{G}_0\mathbf{d}_-^0,\quad \mathbf{t}_x^0=\mathbf{G}_*\mathbf{d}_-^0=\mathbf{G}_*\mathbf{G}_0^{-1}\mathbf{t}_z^0,$$
$$\mathbf{t}_+^0=\mathbf{t}_z^0=\mathbf{G}_0\mathbf{d}_-^0,\quad \mathbf{t}_-^0=0. \tag{6}$$

Thus, the truncation of all terms since the order $O(\varepsilon)$ results in the linear distribution of the displacements across the thickness and in the equivalence of stresses on the surfaces. The stresses $\mathbf{t}_x^0$ can be expressed both via the boundary values of displacements or stresses. In addition, $\mathbf{t}_+^0$ is related to $\mathbf{d}_-^0$ by (6).

If the displacement $\mathbf{d}^\pm$ on the boundaries is given then, assuming $\mathbf{d}^\pm=O(1)$ as $\varepsilon\to 0$ and using (5), we obtain $\lambda+0+1=0$ and $\lambda=-1$. When the stresses $\mathbf{t}_z^{\mp}$ are given we obtain $\lambda=0$.

Proceed to the case $s=1$ now. From the relations (2) and (3) we deduce the equalities

$$\mathbf{d}^1=\mathbf{d}_0^1+z\mathbf{d}_1^1+\frac{z^2}{2!}\mathbf{d}_2^1=\mathbf{d}_0^1+z\mathbf{d}_1^1-\frac{z^2}{2!}\mathbf{G}_0^{-1}\mathbf{D}_1\mathbf{d}_1^0,$$

$$\mathbf{t}_z^1=\mathbf{t}_{0z}^1+z\mathbf{t}_{1z}^1=\mathbf{t}_{0z}^1-z\mathbf{D}\mathbf{G}_\parallel\mathbf{d}_1^0,$$

$$\mathbf{d}_+^1=\mathbf{d}_0^1+\frac{1}{2!}\mathbf{d}_2^1=\mathbf{d}_0^1-\frac{1}{2!}\mathbf{G}_0^{-1}\mathbf{D}_1\mathbf{d}_1^0,\quad \mathbf{d}_-^1=\mathbf{d}_1^1,$$

$$\mathbf{t}_+^1=\mathbf{t}_{0z}^1=\mathbf{G}_0\mathbf{d}_-^1+\mathbf{G}_\parallel^T\mathbf{D}^T\mathbf{d}_+^0,\quad \mathbf{t}_-^1=-\mathbf{D}\mathbf{G}_\parallel\mathbf{d}_-^0,$$

which provide the relations between the surface values of displacements and stresses and result in expressing all the field components through these values

$$\mathbf{d}^1=\mathbf{d}_+^1+z\mathbf{d}_-^1+\frac{1-z^2}{2!}\mathbf{G}_0^{-1}\mathbf{D}_1\mathbf{d}_-^0,\quad \mathbf{t}_z^1=\mathbf{t}_z^1+z\mathbf{t}_-^1=\mathbf{G}_0\mathbf{d}_-^1$$
$$+\mathbf{G}_\parallel^T\mathbf{D}^T\mathbf{d}_+^0-z\mathbf{D}\mathbf{G}_\parallel\mathbf{d}_-^0,$$
$$\mathbf{t}_x^1=\mathbf{G}_*\mathbf{G}_0^{-1}\mathbf{t}_+^1+(\mathbf{G}_\perp^T\mathbf{D}^T-\mathbf{G}_*\mathbf{N}^T)\mathbf{d}_+^0+z(\mathbf{G}_\perp^T\mathbf{D}^T \tag{7}$$
$$-\mathbf{G}_*\mathbf{G}_0^{-1}\mathbf{D}_1)\mathbf{d}_-^0\quad(\mathbf{N}\equiv\mathbf{D}\mathbf{G}_\parallel\mathbf{G}_0^{-1}).$$

For the dimensional quantities some of the relations (7) can be rewritten in the form

$$\mathbf{t}_-=(0-\mathbf{N}\mathbf{G}_0)\mathbf{d}_-,$$
$$\mathbf{d}_-=h(\mathbf{G}_0^{-1}\mathbf{t}_+-\mathbf{N}^T\mathbf{d}_+)\quad\text{or}\quad h\mathbf{t}_+=\mathbf{G}_0(\mathbf{d}_-+h\mathbf{N}^T\mathbf{d}_+). \tag{8}$$

Due to the order of the terms truncated these dimensional impedance boundary relations have the asymptotic error $O(\varepsilon^2)$. Note, that the values of boundary functions are included in (8) symmetrically by their half-sum and half-difference.

Other relations, more convenient to approximate the problem in terms of given stresses or of given displacements only, can be easily derived from formulas (8) as follows:

$$\mathbf{t}_-+h\mathbf{N}\mathbf{t}_+=0, \tag{9}$$

$$\mathbf{d}_-+h\mathbf{N}^T\mathbf{d}_+=h\mathbf{G}_0^{-1}(1\pm h\mathbf{N})\mathbf{t}_z^\pm. \tag{10}$$

Remark 2. The second-order relations (8)–(10) permit one to the express any component of the displacements and stresses on the surfaces through the boundary values given. But all of them are essentially quasistatic—no time derivatives appear. As shown in Refs. 8–10 and 12–16 this model may work for rather long-waves, but the more correct dynamic description needs the relations of high-order.

V. RELATIONS OF HIGH ORDER

The recurrent formulas (2) and (3) permit us to write the components of s-order as polynomials of the transverse coordinate z and to express the half-sum and half-difference of the boundary values

$$\mathbf{d}^s=\sum_{k=0}^{s+1}\frac{z^k}{k!}\mathbf{d}_k^s,\quad \mathbf{d}_k^s=-\mathbf{G}_0^{-1}\mathbf{D}_1\mathbf{d}_{k-1}^{s-1}+\mathbf{G}_0^{-1}\mathbf{A}_2\mathbf{d}_{k-2}^{s-2}(k\geqslant 2),$$
$$\mathbf{t}_z^s=\sum_{k=0}^{s}\frac{z^k}{k!}\mathbf{t}_{kz}^s,\quad \mathbf{t}_{kz}^s=\mathbf{G}_0\mathbf{d}_{k+1}^s+\mathbf{G}_\parallel^T\mathbf{D}^T\mathbf{d}_k^{s-1}, \tag{11}$$

$$\mathbf{d}_+^s=\sum_k\frac{1}{(2k)!}\mathbf{d}_{2k}^s,$$
$$\mathbf{d}_-^s=\sum_k\frac{1}{(2k+1)!}\mathbf{d}_{2k+1}^s\quad(\mathbf{d}\leftrightarrow\mathbf{t}). \tag{12}$$

Let $s=2$ now. Upon the recurrent chains (11) and (12) the components of the displacements and normal stresses are rewritten as follows:

$$\mathbf{d}^2=\mathbf{d}_0^2+z\mathbf{d}_1^2+\frac{z^2}{2!}\mathbf{d}_2^2+\frac{z^3}{3!}\mathbf{d}_3^2,\quad \mathbf{t}_z^2=\mathbf{t}_{0z}^2+z\mathbf{t}_{1z}^2+\frac{z^2}{2!}\mathbf{t}_{2z}^2,$$

$$\mathbf{d}_+^2=\mathbf{d}_0^2+\frac{1}{2!}\mathbf{d}_2^2=\mathbf{d}_0^2-\frac{1}{2!}\mathbf{G}_0^{-1}\mathbf{D}_1\mathbf{d}_-^1+\frac{1}{2!}\mathbf{G}_0^{-1}\mathbf{A}_2\mathbf{d}_+^0,$$

$$\mathbf{d}_-^2=\mathbf{d}_1^2+\frac{1}{3!}\mathbf{d}_3^2=\mathbf{d}_1^2+\frac{1}{3!}\mathbf{G}_0^{-1}\mathbf{B}_2\mathbf{d}_-^0$$
$$(\mathbf{B}_2\equiv\mathbf{A}_2+\mathbf{D}_1\mathbf{G}_0^{-1}\mathbf{D}_1),$$

$$\mathbf{t}_+^2=\mathbf{t}_{0z}^2+\frac{1}{2!}\mathbf{t}_{2z}^2=\mathbf{G}_0\mathbf{d}_-^2+\mathbf{G}_\parallel^T\mathbf{D}^T\mathbf{d}_+^1+\frac{1}{3}\mathbf{B}_2\mathbf{d}_-^0,$$

$$\mathbf{t}_-^2=\mathbf{t}_{1z}^2=-\mathbf{D}\mathbf{G}_\parallel\mathbf{d}_-^0+\mathbf{A}_2\mathbf{d}_+^0,$$

and after transformations they acquire the form

$$\mathbf{d}^2 = \mathbf{d}_+^2 + z\mathbf{d}_-^2 + \frac{1-z^2}{2!}\mathbf{G}_0^{-1}(\mathbf{D}_1\mathbf{d}_-^1 - \mathbf{A}_2\mathbf{d}_+^0) + \frac{z(z^2-1)}{2!}\mathbf{G}_0^{-1}\mathbf{B}_2\mathbf{d}_-^0,$$

$$\mathbf{t}_z^2 = \mathbf{t}_+^2 + z\mathbf{t}_-^2 + \frac{z^2-1}{2!}\mathbf{t}_{2z}^2 = \mathbf{G}_0\mathbf{d}_-^2 + \mathbf{G}_\parallel^T\mathbf{D}^T\mathbf{d}_+^1 + \frac{1}{3}\mathbf{B}_2\mathbf{d}_-^0 + z(-\mathbf{D}\mathbf{G}_\parallel\mathbf{d}_-^0 + \mathbf{A}_2\mathbf{d}_+^0) + \frac{z^2-1}{2!}(\mathbf{A}_2 + \mathbf{D}\mathbf{G}_\parallel\mathbf{G}_0^{-1}\mathbf{D}_1)\mathbf{d}_-^0. \tag{13}$$

To be brief, let us omit the expressions of stresses $\mathbf{t}_x$ here.

The symmetrical relations, similar to (8), look as follows

$$\{1 + h^2/3\mathbf{G}_0^{-1}\mathbf{B}_2\}\mathbf{d}_- = h\{\mathbf{G}_0^{-1}\mathbf{t}_+ - \mathbf{N}^T\mathbf{d}_+\},$$

$$\mathbf{t}_- = \{0 - \mathbf{N}\mathbf{G}_0\}\mathbf{d}_- + h\mathbf{A}_2\mathbf{d}_+, \tag{14}$$

or $h\mathbf{t}_+ = \mathbf{G}_0\{(1+h^2/3\mathbf{G}_0^{-1}\mathbf{B}_2)\mathbf{d}_- + h\mathbf{N}^T\mathbf{d}_+\}$.

As one can see the most significant difference from previous is the appearance of the second order wave operators $\mathbf{A}_2$ and $\mathbf{B}_2$. So, the relations (14) are dynamic. Let us also write the analogs of formulas (11) and (12), respectively

$$\mathbf{t}_- + h(\mathbf{N} \pm h\mathbf{L}_2)\mathbf{t}_+ = h\mathbf{L}\mathbf{d}^\pm, \tag{15}$$

$$(\mathbf{L} \equiv \mathbf{A}_2 + \mathbf{N}\mathbf{G}_0\mathbf{N}^T, \quad \mathbf{L}_2 \equiv \mathbf{L}\mathbf{G}_0^{-1})$$

$$\mathbf{d}_- + h(\mathbf{N}^T \pm h\mathbf{L}_2^T)\mathbf{d}_+ = h\mathbf{G}_0^{-1}(1 \pm h\mathbf{N} + h^2\mathbf{N}_2)\mathbf{t}_z^\pm, \tag{16}$$

$$\left(\mathbf{N}_2 \equiv \mathbf{N}^2 - \tfrac{1}{3}\mathbf{B}_2\mathbf{G}_0^{-1}\right)$$

Using recurrent formulas (2), (3), (11), and (12) similarly to (13) we relate $\mathbf{d}_k^s$ and $\mathbf{t}_{kz}^s$ for the order $s=3,4,5$ to the corresponding boundary values of displacements and transverse stresses. Neglecting the terms since the order $O(\varepsilon^6)$, represent the final dimensional relations for the half-sum and half-differences of the boundary values up to $s=5$

$$\mathbf{t}_- = (0 - \mathbf{N}\mathbf{G}_0 + h^2/3\mathbf{A}_3 - h^4/5\mathbf{A}_5)\mathbf{d}_- + (h\mathbf{A}_2 - h^3/3\mathbf{A}_4)\mathbf{d}_+,$$

$$h\mathbf{t}_+ = (\mathbf{G}_0 + h^2/3\mathbf{B}_2 - h^4/5\mathbf{B}_4)\mathbf{d}_- + (h\mathbf{G}_0\mathbf{N}^T - h^3/3\mathbf{B}_3 + h^5/6\mathbf{B}_5)\mathbf{d}_+, \tag{17}$$

where the operators of high order are presented in the Appendix.

The relations, which generalize the formulas (15), acquire the form

$$\mathbf{t}_- + h(0 + \mathbf{N} \pm h\mathbf{L}_2 + h^2\mathbf{S}_3 \pm h^3\mathbf{R}_4 + h^4\mathbf{P}_5)\mathbf{t}_+ = h(0 + 0 + \mathbf{L} \pm h\mathbf{L}\mathbf{N}^T + h^2\mathbf{M}_4 \pm h^3\mathbf{Q}_5)\mathbf{d}^\pm. \tag{18}$$

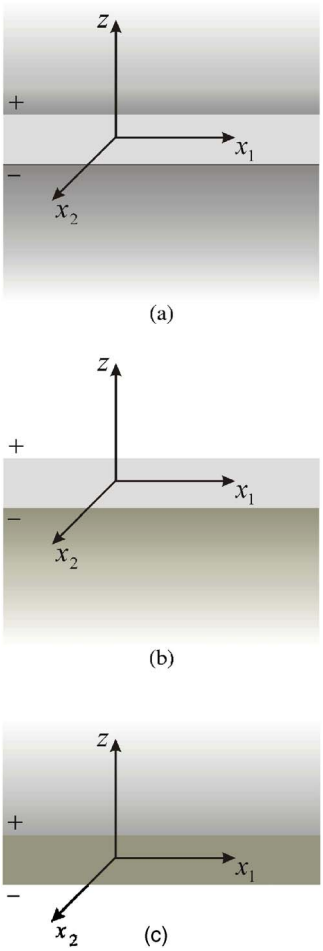


FIG. 1. (Color online) Sketch of a layer, embedded in between two thick solids (a), of a thick solid coated on the top (b), and of a thick solid coated on the bottom (c).

The relations (16) are generalized as well by the following formulas:

$$\mathbf{d}_- + h(0 + \mathbf{N} \pm h\mathbf{L}_2 + h^2\mathbf{S}_3 \pm h^3\mathbf{R}_4 + h^4\mathbf{P}_5)^T\mathbf{d}_+ = h\mathbf{G}_0^{-1}(1 \pm h\mathbf{N} + h^2\mathbf{N}_2 \pm h^3\mathbf{R}_3 + h^4\mathbf{T}_4 \pm \mathbf{h}^5\mathbf{Y}_5)\mathbf{t}_z^\pm, \tag{19}$$

where the high-order matrix differential operators are represented in the Appendix. From the first glance the operators look a bit complicated, but being calculated once using the appropriate symbolic code (MAPLE, MATHEMATICA, etc) they do lead to the problem of reduced dimension. The remarkable fact is, that matrix operators in the left hand side of (18) and (19) are related by transposition, i.e., possess some duality.

VI. HIGH ORDER THEORY OF THIN COATING OR INNER LAYER

Summarize now what can be extracted from the relations (17)–(19). Assume first, that a layer is placed *in between* two thick solids, where the size of each solid is large and its curvature can be neglected. So, we may deal with the elastic anisotropic half spaces, thick plates, shallow shells, etc. Let us mark each solid by the sign − and +, respectively [see Fig. 1(a)]. Each is described by the dynamic equations of anisotropic elasticity, complemented by the continuity

TABLE I. Parameters of materials.

	Aluminum (Al)	Polystyrene (Po)	E-glass (Eg)
Mass density ρ,[kg/m^3]	2700	1060	2000
Young modulus and Poisson's ratio, elastic modulii [n/m^2] (and some wave speeds [m/s])	$E=0.688\,618\times10^{11}$ $\nu=0.344\,260$ $\begin{pmatrix} c_P=6320 \\ c_S=3080 \end{pmatrix}$	$E=0.376\,413\times10^{10}$ $\nu=0.342\,56$ $\begin{pmatrix} c_P=2350 \\ c_S=1150 \end{pmatrix}$	$c_{11}=0.483\,059\times10^{11}$ $c_{13}=0.569\,995\times10^{10}$ $c_{33}=0.148\,662\times10^{11}$ $c_{55}=0.440\times10^{10}$ $(c=1483)$

conditions of normal stresses σ_{ZZ}, tangent stresses $\sigma_{\alpha Z}$ and displacements $\mathbf{U}$ and W on the interface $Z=\mp h$. The scaling procedure in solids differs from that of thin layer only by coordinate Z

$$Z=zL, \quad \mathbf{X}=\mathbf{x}L, \quad T=tT_0. \tag{20}$$

As seen, the timescale is assumed to be the same as in the layer, i.e., we do not consider a contrast situation when the ratio of speeds is comparable with ε-power. If doing so, the case of the low speed layer is nevertheless covered by the same description and the respective terms of inertia in the thick solids can be small; but when the speed in the layer is much greater than those in solids some additional difficulty may appear. Physically such *contrast* situation looks similarly to the motion of a stiff plate embedded into a soft medium and reminds the limit case of a stress free layer. The degenerate situation of a softer environment has to be developed further like the interaction of the carrying stiff layers with soft filler in a panel with contrasting properties.[22]

Hence, substituting (20) into the dynamic equations of elasticity and into Hooke's law we observe no small parameter to appear and these relations look identical both in the dimensional and dimensionless form. As a result, the expansion of each dimensionless displacement or stress in solid into the asymptotic series is caused by the contact with a layer only. By this reason the formulas (17) can be derived in a similar manner to approximate the action of a layer. The approximated solution to the total problem is subdivided into two: Solution to the boundary value problems for two solids with given perturbation (incident waves, different kind of concentrated sources, etc.) and with the effective IBC (17); and, when necessarily, recalculation of the layer stress and strain state [see e.g., the relations (7) in Sec. IV]. As one can see, the *limit case* $h\to0$ of IBC (17) leads to the familiar conditions of the full contact between solids.

When considering a *coated substrate* the choice of the IBC depends on what happens on the surface [see Figs. 1(b) and 1(c)]. Namely, when the *transversal stresses* on the surface are given the effective IBC are chosen in the form (18). The upper indices + or − correspond to the geometry shown in Fig. 1. The perturbation in the right hand side is the value of the displacement on the surfaces + or −. If the *displacements* on the faces are given one should choose the IBC (19).

Remark 3. The degree λ is not specified now and follows automatically from the dimensional IBC with taking account of the perturbation given.

Again, the subdivision of any problem is reduced to the solution of the boundary value problem to a solid with IBC (18) and (19) and to the further recalculation of the layer state. Note that the *limit case* $h\to0$ of these IBC leads to the equivalence of the interface stresses (and displacements) to those on the surface.

The implementation of IBC (17)–(19) can be various—not only for calculating the reflection/transmission coefficients but also for approximating the dispersion curves of Love, Stoneley, or Lamb waves in layered media. The Green function and the media response to an arbitrary load can be studied as well. However, one should note that when modeling a coating subjected to the *concentrated* load $\mathbf{t}_z^+$ or $\mathbf{t}_z^-$, which, in general, contradicts the initial asymptotic assumptions, the latter has to be differentiated. This may explain some accuracy loss shown in Ref. 16.

The frequency domain should not exceed the first quasi resonance frequency of a thin layer. Indeed, when the layer surfaces totally reflect energy and at least one surface is bound the dimensionless quasi resonance frequencies are determined by formula $\Omega_n\equiv\omega h/c=\pi n/2$ $(n=1,2,\ldots)$ where c runs the values of characteristic wavespeeds. So, if the frequency is smaller than Ω_1 we may expect the satisfactory accuracy. In the meantime the dimensional frequency ω might be large enough and be greater than a few cut off frequencies of a thick solid. Of course, the mode conversion between the thin layer and thick solid(s) may perturb the numerical value Ω_1. The spectrum of the associated boundary value problem may contain imaginary part due to the energy leak into a thick solid (e.g., a half-space), but the real part of such natural frequency is caused by the layer itself and resonance becomes finite.

VII. NUMERICAL VALIDATION

Since any problem for seeking modes and spectra usually begins with the wave decomposition the validation by the time-harmonic $e^{-i\omega T}$ partial waves seem quite reasonable and informative for testing purpose. Similarly to Refs. 8–10 we obtain an exact solution using the propagator matrix for the following cases: A contact of two elastic isotropic half spaces with an orthotropic layer in between and a case of an isotropic half-space, coated by such a layer. The materials chosen are: Aluminum (Al), Polystyrene (Po), and orthotropic composite E-glass (Eg). Their parameters are presented in the Table I. In what follows the incident P- or S-harmonic wave falls from the upper half space + to the interface and causes the reflected (and transmitted) waves. Thus, for any

incident wave the response in each half-space consists of two waves with complex magnitudes M, normalized by the magnitude of the incident wave M^{inc} and marked by $+$, $-$ and P, S, respectively. A comparison of these magnitudes, obtained using asymptotic IBC (17) and (18) or (19) against direct evaluation is carried out.

First, consider an orthotropic layer (Eg) in between the upper half-space (Al) and lower half-space (Po), i.e., the media contact Al-Eg-Po. As a criterion introduce the relative mean square error

$$e=\left\{\frac{1}{4}\sum_{[+,-;P,S]}\left|\frac{M^{\mathrm{as}}-M^{\mathrm{ex}}}{M^{\mathrm{ex}}}\right|^2\right\}^{1/2}\times 100\,\%\,, \tag{21}$$

where M^{as} runs the magnitudes by the asymptotic solution and M^{ex}—by the exact one. A few typical graphs are represented in Fig. 2 where the magnitudes and the error e are depicted by Figs. 2(a) and 2(b), respectively. The incident S-wave falls at the angle 80° to the interface. The indices $n\equiv s+1=1,2,\ldots,6$ indicate the curves with asymptotic error of the order $O(\varepsilon^n)$ in IBC (17); symbols Pr, Pt, and Sr, St denote the reflected (half-space +) and transition (half-space $-$) P- and S-waves. All the graphs are plotted against H/L, where L is the wavelength of the slowest shear wave in the layer: $L=cT_0$, $c=\sqrt{g_{55}/\rho}$, $T_0=2\pi/\omega$ (ω is a frequency). As seen, the low order approximations at $n=1,2$ are not very accurate, even for small H/L. The higher approximations $n=3,4$ and $n=5,6$ essentially improve the accuracy. Thus, the error e remains less than 1% till $H/L\approx 0.11$ for $n=3,4$ and $H/L\approx 0.18$ for $n=5,6$. We also may highlight one peculiarity of our IBC: the odd approximation brings a "new quality" and the next even approximation slightly improves this result in the previous interval of validity. For the P-incidence the picture is qualitatively similar but accuracy is slightly better. The numerical results for other subcritical, critical or supercritical angles of incidence look similarly and tell nothing new in context of validation. The physical limit of the asymptotic model is the first quasi resonance frequency of the layer, which corresponds to $H/L=1/2$ when the layer surfaces totally reflect energy. The deviation from one half is caused by the layer interaction with the half-spaces. We can observe new value $H/L\approx 0.255$ [Fig. 2(a)] with the explicit error increasing near it. Thus, in principle, we may guess the approximate quasi resonance frequencies from these data. For the case of contrast materials, e.g., two stiff thick solids—half-spaces—and a soft embedded layer, the limit value is expected to be closer to one half similarly to the case of a fluid layer in between two solids.[30] For a stiffer layer the mechanism of the energy exchange between elastic media is more difficult due to the simultaneous distortion and shear processes and their conversion on the interface.

To characterize the energy propagation introduce the relative error

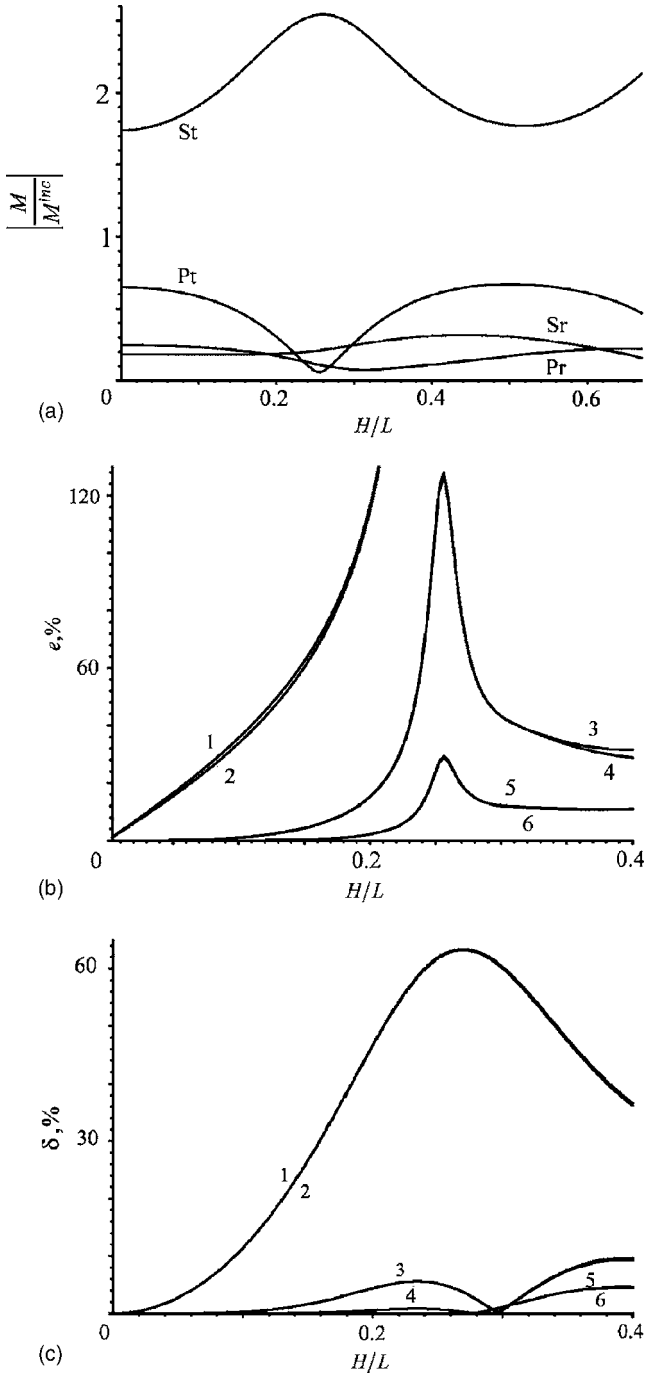


FIG. 2. Media contact Al-Eg-Po at the incident S-wave falling at 80° to the interface: (a) Exact solution; (b) relative error of asymptotics; (c) relative error of the power flow for the upper solid.

$$\delta=\left|\frac{\mathbf{\Pi}^{\mathrm{as}}-\mathbf{\Pi}^{\mathrm{ex}}}{\mathbf{\Pi}^{\mathrm{in}}}\right|\times 100\,\%\,, \tag{22}$$

where $\mathbf{\Pi}^{\mathrm{ex}}$ and $\mathbf{\Pi}^{\mathrm{as}}$ are vectors of the averaged power flow density of the exact and asymptotic solution, given by formulas

$$\mathbf{\Pi}\equiv(\Pi_1,\Pi_2,\Pi_3),\quad \Pi_\alpha\equiv-\frac{1}{T_0}\int_0^{T_0}\mathrm{Re}(\sigma_{\alpha\beta})\mathrm{Re}(\partial_t u_\beta)dt,$$

and Π^{in} is the power flow of the incident wave. For the simplest plane waves the power flow is located along the wave vector and acquires a scalar value

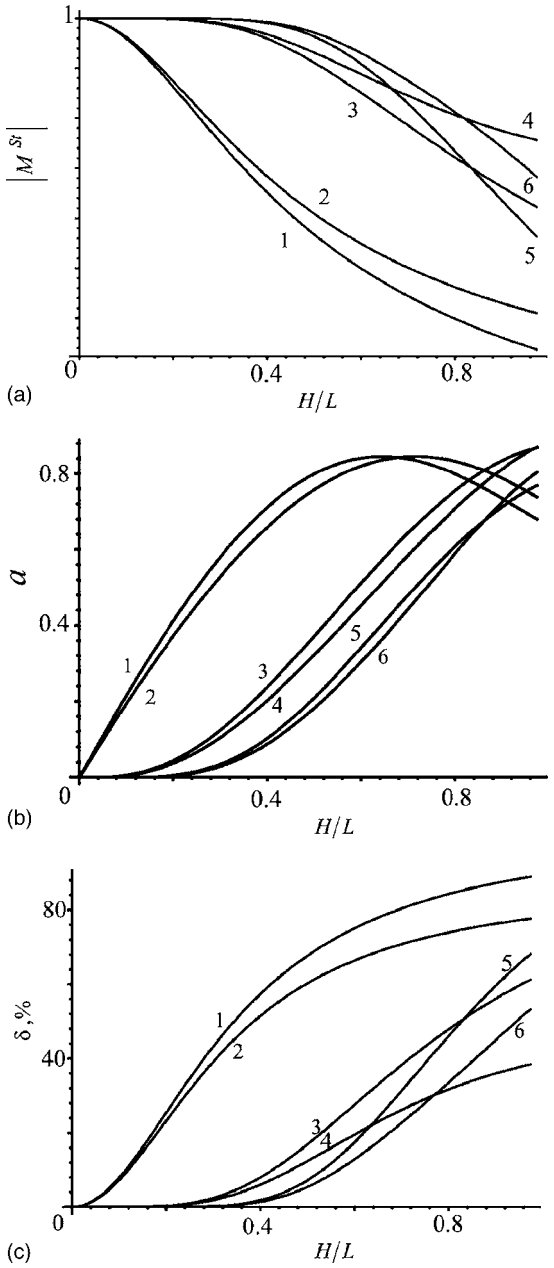


FIG. 3. Media contact Al-Al-Al at the incident S-wave falling at 80° to the interface: (a) Magnitude of St-wave; (b) absolute mean square error of asymptotics; (c) relative error of power flow.

$$\Pi^{P,S} = \frac{\omega^2}{2}\rho c_{P,S}|M^{P,S}|^2.$$

The error (22) for the upper solid is represented in Fig. 2(c). As seen this error is smaller than e. For the lower half-space the error δ is identical since the power flow *across* the layer is *balanced* for the approximation of any order $s = 0, 1, \ldots, 5$. Its value varies with s but remains the same over and under the layer. This property holds even for the case of identical materials—despite the fact that our assumptions were for *different* media. One example for the same material (A1) of a layer and half-spaces is shown in Fig. 3 with 80° of S-incidence. In the exact solution the only S-wave of unit magnitude propagates. The magnitudes of St-waves given by asymptotics are depicted in Fig. 3(a). In

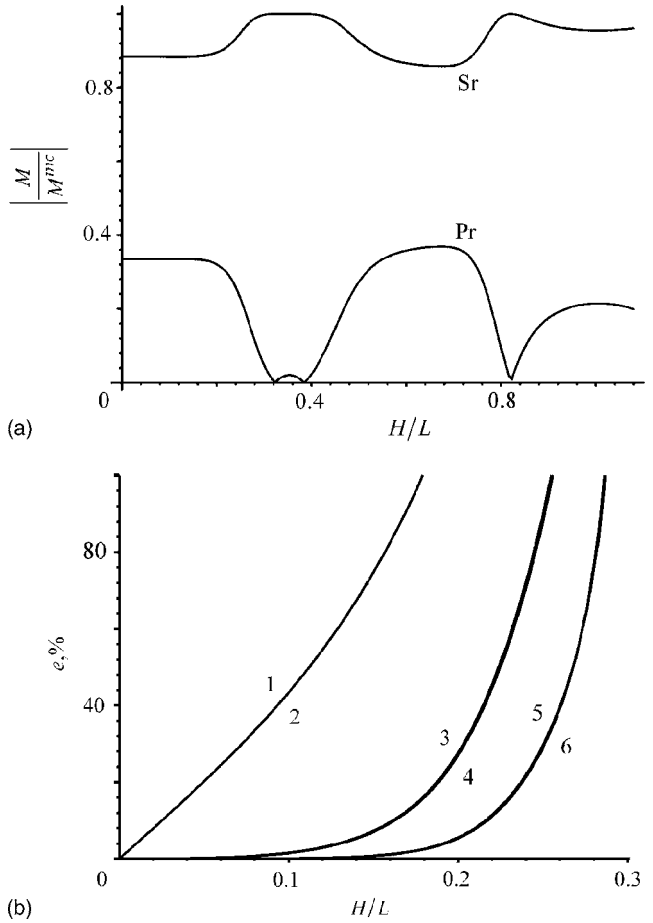


FIG. 4. Media contact Al-Eg with the stress free surface at the incident S-wave falling at 80° to the interface: (a) Exact solution; (b) relative error of asymptotics.

Fig. 3(b) the absolute mean square error a (which is similar to e but not in percents and without division) is plotted. The curves in Fig. 3(c) represent the error δ for the upper solid. Their behavior is monotonous since there are no physical resonances in this system.

Second, let us validate the asymptotic model of the coating and to begin with a *stress free* layer surface. This approximation is given by formulas (18). Consider the upper half-space (A1) coated by the orthotropic layer (Eg). The mean error is introduced similarly to (21) but now for two complex magnitudes of the refracted P- and S-waves. In Figs. 4(a) and 4(b) the exact magnitudes of the refracted waves and the error e are presented for the incident S-waves falling to the interface with the angle 80°. The error behaves similarly to the previous consideration. The coincidence of the curves $n = 1, 2$ is explained by the appearance of the displacements in formulas (18) since the third iteration only. Note that for S-incidence the neighboring curves for $n = 3, 4$ and $n = 5, 6$ are close to each other. This is the only difference from the similar situation with P-incidence where their divergence is more visible.

The same comparison is performed for the *clamped* surface of the coating which is approximated by IBC (19). The respective results for the Al-Eg media contact are represented in Figs. 5(a) and 5(b). As one can see, the error is slightly smaller than that of the stress free surface. Now the differ-

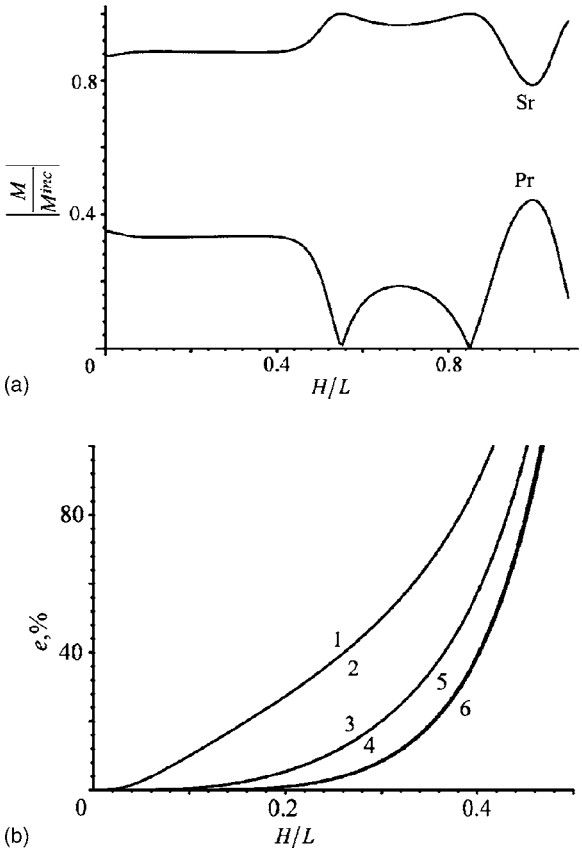


FIG. 5. Media contact Al-Eg with the clamped surface at the incident S-wave falling at 80° to the interface: (a) Exact solution; (b) relative error of asymptotics.

ence between the neighboring curves is also rather small for both S- and P-incidence (the latter looks very similar and omitted here). But in contrast to previous, the curves for $s=1,2$ do not coincide. For the critical and supercritical angles the situation does not change too much.

Finally let us mention that, in fact, when dealing with monochromatic waves $e^{i(\mathbf{kr}-\omega T)}$ the small parameter $\varepsilon=h/L$ should be replaced by $kh=\pi H/L$. As shown, our asymptotics work satisfactory till the values $H/L\approx 0.1$ (for $n=3,4$) and $H/L\approx 0.16$ (for $n=5,6$). The respective $kh\approx 0.3$ and $kh\approx 0.51$, i.e., is no longer small.

VIII. DISCUSSION AND COMPARISON WITH REFS. 24 and 25

The deduced asymptotic models are much more efficient than the approximate low order IBC obtained from the asymptotic analysis of the propagator matrix.[8–10,12–16] Indeed, the procedure of our asymptotic integration is direct, based on the 3D dynamic equations and boundary conditions of anisotropic elasticity and permits one to derive the model of any order using recurrent relations. The relative accuracy of each recurrent step is $O(\varepsilon)$ in contrast to $O(\varepsilon^2)$ when modeling thin plates[17–20] or fluid layer.[30] As one can see, the significant improvement of the model occurs since the third step. In can be explained by the appearance of the iterated $\mathbf{A}_2$- and $\mathbf{B}_2$-wave operators in formulas (17)–(19). Each of the iterations moves the right bound of the validity interval closer and closer to the first quasi resonance frequency of the elastic system. But note the *difference*—the iterations of the wave operators take place for $s=2,4,\ldots$ (respective curves $n=s+1=3,5,\ldots$ in our notations) and bring a *new quality*. For the iterations $s=3,5,\ldots$ ($n=4,6,\ldots$) the operators from previous step are differentiated with respect to the longitudinal coordinates (see the slight corrections from curve to curve in Figs. 2–5.

The possibility to replace a layer by IBC reflects the fundamental property of a layer with bound faces at low frequency—the absence of propagating modes till the first nonzero cut-off frequency. But due to the mode conversion now we should speak about the quasi resonance whose value is perturbed by the interface contact and the limit value H/L is smaller than one-half.

Comparing with the plate and shallow shell theories in our case each iteration of the high order requires more analytical efforts. For example, the sixth-order IBC look more cumbersome than its analogue for plates—the sixth-order Timoshenko-Reissner model,[31,32] but have a bit smaller numerical error. However, a good compromise give the iterations $s=2,3$ which hold the accuracy 1% till $H/L\approx 0.1$ and still look simple enough [see IBC (17)–(19) and Figs. 2–5]. The bound values H/L for different iterations satisfying the criterion $e\leqslant 1\%$ are represented in the Table II.

To sum up the discussion consider the approximate transfer matrix given by asymptotic IBC for waves propagating as $e^{i(\mathbf{kx}-\omega T)}$. For this purpose let us come back to the equations of motion and Hooke's law. Expressing normal derivatives of $\mathbf{d}$ and $\mathbf{t}_z$ from Eqs. (2) and (3) in the dimensional form (no more upper indices there) we arrive at the equality

$$\partial_z \boldsymbol{\xi} = \mathbf{A}_* \boldsymbol{\xi},$$

where

$$\boldsymbol{\xi} \equiv \begin{bmatrix} \mathbf{d} \\ \mathbf{t}_z \end{bmatrix}, \quad \mathbf{A}_* \equiv \begin{bmatrix} -\mathbf{N}^T & \mathbf{G}_0^{-1} \\ \mathbf{L} & -\mathbf{N} \end{bmatrix},$$

and matrix $\mathbf{A}_*$ differs from the fundamental matrix (fundamental acoustic tensor[33]) by factor i. The asymptotic IBC become dynamic since $s=2$. Having Eqs. (14)–(16) for $\mathbf{d}^{\pm}$ and $\mathbf{t}_z^{\pm}$ we obtain the approximation $\mathbf{B}_{II}$ of the transfer matrix $\mathbf{B}(h)=e^{2h\mathbf{A}_*}$ by relating the surface values of $\boldsymbol{\xi}$ as follows:

TABLE II. The IBC limit values H/L corresponding to the error 1% against the index $s=0,1,\ldots$.

	0	1	2	3	4	5
(17)	0.04	0.042	0.11	0.115	0.18	0.185
(18)	0.04	0.045	0.1	0.105	0.16	0.163
(19)	0.03	0.033	0.12	0.122	0.19	0.193

$$\mathbf{A}_{II}^{+}\boldsymbol{\xi}^{+}=\mathbf{A}_{II}^{-}\boldsymbol{\xi}^{-}\rightarrow\boldsymbol{\xi}^{+}=\mathbf{B}_{II}\boldsymbol{\xi}^{-},\quad \mathbf{B}_{II}\equiv(\mathbf{A}_{II}^{+})^{-1}\mathbf{A}_{II}^{-}, \tag{23}$$

where

$$\mathbf{A}_{II}^{\pm}(h)\equiv 1\mp h\mathbf{A}_{*}-h^{2}\mathbf{O}_{2}\mp h^{3}\mathbf{O}_{3}\quad\text{and}$$
$$(\mathbf{A}_{II}^{+})^{-1}\mathbf{A}_{II}^{-}=1+2h\mathbf{A}_{*}+O(h^{2}), \tag{24}$$

$$\mathbf{O}_{2}\equiv\begin{bmatrix}0 & \mathbf{G}_{0}^{-1}\mathbf{N}\\ 0 & 0\end{bmatrix},\quad \mathbf{O}_{3}\equiv\begin{bmatrix}0 & \mathbf{G}_{0}^{-1}\mathbf{N}_{2}\\ 0 & 0\end{bmatrix}. \tag{25}$$

Matrix $\mathbf{B}_{II}$ differs from that in Refs. 10, 24, and 25 by presence of square and cubic terms in $\mathbf{A}_{II}^{\pm}$. However, it also provides the balance of power flow across the layer even for the case of identical materials and, in particular, can be used for subdivision of a thick layer into N thin layers with the thickness $2h/N$ each. Then, similarly to Refs. 24 and 25 the total transfer matrix is approximated by

$$[\mathbf{B}_{II}(h/N)]^{N}=[1+2h/N\mathbf{A}_{*}+O(h/N)^{2}]^{N}\rightarrow e^{2h\mathbf{A}_{*}}\quad\text{as } N\rightarrow\infty,$$

which converges to an exact matrix $\mathbf{B}(h)$. The additional square and cubic terms do not change the limit but may affect the rate of convergence.

Thus, our next asymptotics for $s=3,4,\ldots$ cause approximations of the transfer matrix $\mathbf{B}(h)$ which also differ from $\mathbf{B}_{IV},\mathbf{B}_{VI},\ldots$ presented in Refs. 24 and 25 by high order terms (e.g., see Appendix for matrix $\mathbf{A}_{IV}^{\pm}$ at $s=4$). Hence, the efficiency of presented approach is comparable with the recursive algorithms[24,25] but the difference is responsible for the description of the internal stress and strain state of a layer in the uniform manner, i.e., when truncating the asymptotic series (1) for the displacements and stresses with the same relative error.

IX. CONCLUSION

The efficient direct asymptotic integration approach, applied to the 3D dynamic equations and boundary conditions of elasticity, results in the so-called 2D *impedance boundary conditions* (17)–(19) for the case of the inner layer or coatings. So, the action of the layer can be replaced by these conditions. Due to the recurrent nature of asymptotic procedure the IBC of any order can be found easily. Comparison with the alternative recursive algorithm[24,25] for calculating transfer or stiffness matrices exhibits the difference, responsible for the uniform asymptotic error of the displacements and stresses in our case. In this paper we restrict ourselves by the asymptotic accuracy $O(\varepsilon^{6})$. The numerical tests demonstrate that the frequency loop up to the first quasi resonance of the system thick solid(s)-layer can be described by this way. Since the dimensionless wavenumber can reach the value of one-half, the obtained results are valid at low frequency but possibly not for a long-wave process. In addition, the obtained IBC are applicable to the various substrates—anisotropic half-spaces, thick multilayered coated or sandwich plates, shallow shells, etc. Due to the problem addressed and to the frequency band a researcher may choose exact solution, approximation,[24,25] or our asymptotics as a compromise to the dimension of problem, accuracy, logical, and numerical efforts. As it is said in Sec. VII and in Refs. 24 and 25, the approximate IBC do not cause a "parasitic" energy scattering.

ACKNOWLEDGMENT

This work has been carried out in the frame of the EPRRC Grant No. GR/M31552/01, which is gratefully acknowledged.

APPENDIX: OPERATORS OF HIGH ORDER

(1) The high-order matrix differential operators in formulas (17) look as follows:

$$\mathbf{A}_{3}\equiv\mathbf{A}_{2}\mathbf{G}_{0}^{-1}\mathbf{D}_{1},\quad \mathbf{A}_{4}\equiv\mathbf{A}_{2}\mathbf{G}_{0}^{-1}\mathbf{A}_{2},$$

$$\mathbf{A}_{5}\equiv\tfrac{2}{3}\mathbf{A}_{4}\mathbf{G}_{0}^{-1}\mathbf{D}_{1}+\tfrac{1}{4}\mathbf{A}_{3}\mathbf{G}_{0}^{-1}\mathbf{B}_{2},$$

$$\mathbf{B}_{3}\equiv\mathbf{A}_{3}^{T},\quad \mathbf{B}_{4}\equiv\tfrac{2}{3}\mathbf{B}_{3}\mathbf{G}_{0}^{-1}\mathbf{D}_{1}+\tfrac{1}{4}\mathbf{B}_{2}\mathbf{G}_{0}^{-1}\mathbf{B}_{2},$$

$$\mathbf{B}_{5}\equiv\tfrac{4}{5}\mathbf{B}_{3}\mathbf{G}_{0}^{-1}\mathbf{A}_{2}+\tfrac{2}{15}\mathbf{B}_{2}\mathbf{G}_{0}^{-1}\mathbf{B}_{3}.$$

(2) The high-order matrix differential operators in formulas (18) are expressed using previous operators in the following form:

$$\mathbf{S}_{3}\equiv\left[\mathbf{L}\mathbf{N}^{T}-\tfrac{1}{3}(\mathbf{A}_{3}+\mathbf{N}\mathbf{B}_{2})\right]\mathbf{G}_{0}^{-1},$$

$$\mathbf{M}_{4}\equiv\mathbf{S}_{3}\mathbf{G}_{0}\mathbf{N}^{T}-\tfrac{1}{3}(\mathbf{A}_{4}+\mathbf{N}\mathbf{B}_{3}),$$

$$\mathbf{R}_{4}\equiv(\mathbf{M}_{4}-\tfrac{1}{3}\mathbf{L}_{2}\mathbf{B}_{2})\mathbf{G}_{0}^{-1},$$

$$\mathbf{Q}_{5}\equiv\mathbf{R}_{4}\mathbf{G}_{0}\mathbf{N}^{T}-\tfrac{1}{3}\mathbf{L}_{2}\mathbf{B}_{3},$$

$$\mathbf{P}_{5}\equiv\left[\mathbf{Q}_{5}-\tfrac{1}{3}\mathbf{S}_{3}\mathbf{B}_{2}+\tfrac{1}{5}(\mathbf{A}_{5}+\mathbf{N}\mathbf{B}_{4})\right]\mathbf{G}_{0}^{-1}.$$

(3) The operators in formulas (19) acquire the form

$$\mathbf{R}_{3}\equiv\mathbf{S}_{3}+\mathbf{N}_{2}\mathbf{N}-\mathbf{L}\mathbf{N}^{T}\mathbf{G}_{0}^{-1},$$

$$\mathbf{T}_{4}\equiv\mathbf{R}_{3}\mathbf{N}+\mathbf{N}\mathbf{R}_{3}-\mathbf{N}\mathbf{N}_{2}\mathbf{N}+\left(\tfrac{1}{5}\mathbf{B}_{4}+\tfrac{1}{9}\mathbf{B}_{2}\mathbf{G}_{0}^{-1}\mathbf{B}_{2}\right)\mathbf{G}_{0}^{-1},$$

$$\mathbf{Y}_{5}\equiv\mathbf{T}_{4}\mathbf{N}+\left[\tfrac{1}{5}(\mathbf{A}_{5}+\mathbf{N}\mathbf{B}_{4})-\tfrac{1}{3}(\mathbf{N}_{2}\mathbf{A}_{3}+\mathbf{R}_{3}\mathbf{B}_{2})\right]\mathbf{G}_{0}^{-1}.$$

(4) Matrix $\mathbf{A}_{IV}^{\pm}$ for $s=4$ is introduced similarly to (23) and given by formulas:

$$\mathbf{A}_{IV}^{\pm}(h)\equiv\mathbf{A}_{II}^{\pm}(h)-h^{2}\mathbf{O}_{2}'\mp h^{3}\mathbf{O}_{3}'-h^{4}\mathbf{O}_{4}'\mp h^{5}\mathbf{O}_{5}',$$

where

$$\mathbf{O}_{2}'\equiv\begin{bmatrix}0 & 0\\ \mathbf{L}\mathbf{N}^{T} & 0\end{bmatrix},\quad \mathbf{O}_{3}'\equiv\begin{bmatrix}-\mathbf{S}_{3}^{T} & 0\\ \mathbf{M}_{4} & -\mathbf{S}_{3}\end{bmatrix},$$

$$\mathbf{O}_{4}'\equiv\begin{bmatrix}0 & \mathbf{G}_{0}^{-1}\mathbf{R}_{3}\\ 0 & 0\end{bmatrix},\quad \mathbf{O}_{5}'\equiv\begin{bmatrix}0 & \mathbf{G}_{0}^{-1}\mathbf{T}_{4}\\ 0 & 0\end{bmatrix}.$$

As seen, some of these components appear in matrix $\mathbf{A}_{*}^{2}$

$$\mathbf{A}_{*}^{2}\equiv\begin{bmatrix}(\mathbf{L}_{2}+\mathbf{N}^{2})^{T} & -\mathbf{G}_{0}^{-1}\mathbf{D}_{1}\mathbf{G}_{0}^{-1}\\ -\mathbf{L}\mathbf{N}^{T}-\mathbf{N}\mathbf{L} & \mathbf{L}_{2}+\mathbf{N}^{2}\end{bmatrix},$$

but there is no coincidence between this square and additional high order terms in $\mathbf{A}_{IV}^{\pm}$ even with taking into account

the terms $\mathbf{O}_2$ and $\mathbf{O}_3$ from expressions (24) and (25).

[1]W. T. Thomson, "Transmission of elastic waves through a stratified solid medium," J. Appl. Phys. **21**, 89–93 (1950).
[2]N. A. Haskel, "The dispersion of surface waves on multilayered media," Bull. Seismol. Soc. Am. **43**, 17–34 (1953).
[3]A. L. Knopoff, "A matrix method for elastic wave problem," Bull. Seismol. Soc. Am. **54**(1), 431–438 (1964).
[4]F. Schwab and A. L. Knopoff, "Surface waves in multilayered anelastic media," Bull. Seismol. Soc. Am. **61**(4), 893–912 (1971).
[5]L. M. Brekhovskikh, *Waves in layered media* (Academic, New York, 1960).
[6]P. D. Jackins and G. C. Gaunaurd, "Resonance acoustic scattering from stacks of bonded elastic plates," J. Acoust. Soc. Am. **80**, 1762–1765 (1986).
[7]M. A. Leontovich, "On approximate boundary conditions for electromagnetic field on a surface of well conducting bodies," Invest. of Radio Waves Prop., USSR. Acad. Sci., part **2**, 5–12 (1948).
[8]S. L. Rokhlin and Y. J. Wang, "Equivalent boundary conditions for thin orthotropic layer between two solids: reflection, refraction and interface waves," J. Acoust. Soc. Am. **91**, 1875–1887 (1992).
[9]S. L. Rokhlin and Y. J. Wang, "Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids: exact and asymptotic-boundary-condition method," J. Acoust. Soc. Am. **92**, 1729–1742 (1992).
[10]S. L. Rokhlin and W. Huang, "Ultrasonic wave interaction with a thin anisotropic layer between two anisotropic solids: II. Second-order asymptotic boundary conditions," J. Acoust. Soc. Am. **94**(6), 3405–3420 (1993).
[11]J. Sadler, B. O'Neil, and R. G. Maev, "Ultrasonic wave propagation across a thin nonlinear anisotropic layer between two half-spaces," J. Acoust. Soc. Am. **118**(1), 51–59 (2005).
[12]P. Bovik, "On the modeling of thin interface layers in elastic and acoustic scattering problems," Q. J. Mech. Appl. Math. **47**, 17–42 (1994).
[13]P. Bovik, "A comparison between the Tiersten model and O(h) boundary conditions for elastic surface waves guided by thin layers," ASME J. Appl. Mech. **63**, 162–167 (1996).
[14]A. J. Niklasson, S. K. Datta, and M. L. Dunn, "On approximating guided waves in plates with thin anisotropic coating by means off effective boundary conditions," J. Acoust. Soc. Am. **108**(3), 924–933 (2000).
[15]A. J. Niklasson, S. K. Datta, and M. L. Dunn, "On ultrasonic guided waves in a thin anisotropic layer lying between two isotropic layers," J. Acoust. Soc. Am. **108**(5), 2005–2011 (2000).
[16]A. J. Niklasson and S. K. Datta, "Transient ultrasonic waves in multilayered superconducting plates," ASME J. Appl. Mech. **69**, 811–818 (2002).
[17]A. L. Goldenweiser, "The general theory of elastic bodies (shells, coatings and linings)," Mech. Solids [MTT] **3**, 3–17 (1992).
[18]A. L. Goldenweiser, "Derivation of the approximate theory of plate bending by asymptotic integration method," J. Appl. Math. Mech. [PMM] **26**, 668–686 (1962).
[19]K. O. Friedrichs and R. F. Dressler, "A boundary layer theory for elastic plates," Commun. Pure Appl. Math. **14**, 1–33 (1961).
[20]A. E. Green, "On the linear theory of the elastic shells," Proc. R. Soc. London, Ser. A **266**, 143–160 (1962).
[21]D. D. Zakharov, "2D dynamic equations of a thin nonsymmetrical layered elastic plate with anisotropy of a general form," Physics-Doklady **336**(5), 378–380 (1994).
[22]D. D. Zakharov, "The averaged dynamic equations for thin multilayered packets of arbitrary structure with contrasting directions of anisotropy in elastic layers," J. Appl. Math. Mech. [PMM] **63**(1), 93–100 (1999).
[23]G. Johansson and A. J. Niklasson, "Approximate dynamic boundary conditions for a thin piezoelectric layer," Int. J. Solids Struct. **40**, 3477–3492 (2003).
[24]L. Wang and S. I. Rokhlin, "Modeling of wave propagation in layered piezoelectric media by a recursive asymptotic method," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**(9), 1060–1071 (2004).
[25]L. Wang and S. I. Rokhlin, "Recursive geometric integrators for wave propagation in a functionally-graded multilayered elastic medium," J. Mech. Phys. Solids **52**(11), 2473–2506 (2004).
[26]B. A. Auld, *Acoustic Fields and Waves in Solids*, Vol. **1** (Krieger, New York, 1990).
[27]T. C. T. Ting, *Anisotrpic Elasticity: Theory and Applications* (Oxford University Press, Oxford, 1996).
[28]R. Leiderman, A. M. B. Braga, and P. E. Barbone, "Scattering of ultrasonic waves by defective adhesion interfaces in submerged laminated plates," J. Acoust. Soc. Am. **118**(4), 2154–2166 (2005).
[29]M. Destrade, "Surface acoustic waves in rotating orthorhombic crystals," Proc. R. Soc. London, Ser. A **460**, 653–665 (2004).
[30]D. D. Zakharov, "Approximate high order dynamic theory of a fluid layer in between two thick solids," J. Acoust. Soc. Am. **117**(2), 518–527 (2005).
[31]D. D. Zakharov, "Analysis of the acoustical edge flexural mode in a plate using refined asymptotics," J. Acoust. Soc. Am. **116**(2), 872–878 (2004).
[32]A. L. Goldenveizer, J. D. Kaplunov, and E. V. Nolde, "On Timoshenko-Reissner type theories of plates and shells," Int. J. Solids Struct. **30**, 675–694 (1993).
[33]A. H. Fahmy and E. D. Adler, "Propagation of acoustic waves in multilayers: A matrix description," Appl. Phys. Lett. **20**, 495–497 (1973).

Generation of the basis sets for multi-Gaussian ultrasonic beam models—An overview

Hak-Joon Kim[a)]
Center for NDE, Iowa Sate University, Ames, Iowa 50011

Lester W. Schmerr, Jr.[b)]
Center for NDE, Iowa State University, Ames, Iowa 50011 and Department of Aerospace Engineering, Iowa State University, Ames, Iowa 50011

Alexander Sedov[c)]
Department of Mechanical Engineering, Lakehead University, Thunder Bay, Ontario P7B 5E1, Canada

(Received 10 April 2005; revised 10 December 2005; accepted 8 January 2006)

By using a small number of Gaussian basis functions, one can synthesize the wave fields radiated from planar and focused piston transducers in the form of a superposition of Gaussian beams. Since Gaussian beams can be transmitted through complex geometries and media, such multi-Gaussian beam models have become powerful simulation tools. In previous studies the basis function expansion coefficients of multi-Gaussian beam models have been obtained by both spatial domain and *k*-space domain methods. Here, we will give an overview of these two methods and relate their expansion coefficients. We will demonstrate that the expansion coefficients that have been optimized for circular piston transducers can also be used to generate improved field simulations for rectangular probes. It will also be shown that because Gaussian beams are only approximate (paraxial) solutions to the wave equation, a multi-Gaussian beam model is ultimately limited in the accuracy it can obtain in the very near field. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2169921]

PACS number(s): 43.20.Bi, 43.20.El, 43.35.Cg [LLT] Pages: 1971–1978

I. INTRODUCTION

In a seminal paper, Wen and Breazeale[1] demonstrated that it was possible to accurately simulate the sound beam of a circular piston probe radiating into water by the superposition of only ten Gaussian beams. This produces a multi-Gaussian (MG) beam model that is orders of magnitude more numerically efficient than a direct numerical evaluation of a Rayleigh-Sommerfeld integral representation of the wave field. While there are other piston transducer models, such as a boundary diffraction wave model,[2] for example, that are also numerically efficient, a MG beam model retains its efficiency even when the beam of the transducer interacts with complex geometries and media since Gaussian beams can be analytically transmitted or reflected from curved interfaces and can be propagated through anisotropic elastic media. Thus, MG beam models are powerful tools for applications such as simulating ultrasonic nondestructive evaluation inspections.

A key element of a MG beam model is the determination of the expansion coefficients that define the basis set of Gaussians that are distributed on the face of the transducer. In Wen and Breazeale's original paper,[1] these coefficients were obtained by fitting the Gaussians to match the piston velocity distribution on the face of the transducer using a nonlinear least squares approach. Since this fitting is done in the real spatial coordinates of the transducer, we will refer to this procedure as a *spatial domain method*. Although the expansion coefficients obtained by Wen and Breazeale were only for a circular planar piston transducer, Ding *et al.*[3] have shown that the same coefficients can be used to simulate elliptical and rectangular transducers. By a simple modification of these same coefficients, it is also possible to model bi-cylindrically or spherically focused probes. In contrast, Sha *et al.*[4,5] obtained the expansion coefficients for a rectangular and elliptical planar piston transducers by using a nonlinear least squares fitting procedure for a set of Gaussians expressed in a two-dimensional (2D) spatial Fourier transform (*k*-space) domain. Since the Fourier transform of a Gaussian is itself a Gaussian, this *k-space method* can be easily converted into a MG beam model in real space coordinates. We will discuss the advantages/disadvantages of using either the spatial or *k*-space domains. The expansion coefficients obtained by spatial and *k*-space domain fitting methods are different, but we will show how these two sets of coefficients are related so that one can easily transform from one domain representation to another.

Wen and Breazeale[6] and others[7] have also examined other methods to obtain the expansion coefficients. In Ref. 6, for example, Wen and Breazeale obtained a set of fifteen coefficients for a circular piston transducer by also constraining those coefficients to produce the known on-axis field.

[a)]Now with: School of Mechanical Engineering, Sungkyunkwan University, 300 ChunChun-Dong, JangAn-Gu, Suwon, KyungGi-Do 440-746, Korea. Telephone: +82-31-290-7460; Fax: +82-31-290-5276. Electronic mail: hjkim21c@skku.edu
[b)]Telephone: +1-515-294-9746; Fax: +1-515-294-7771. Electronic mail: Ischmerr@iastate.edu
[c)]Telephone: +1-807-343-8268; Fax: +1-807-343-8928. Electronic mail: asedov@lakeheadu.ca

They showed that these fifteen "optimized" coefficients gave a better representation than their original ten coefficients for the transducer field in the very near field. Here, we will show that these same fifteen coefficients, even though they were specifically optimized for the circular case, also give significantly improved near field behavior when used to model rectangular transducers.

If fifteen expansion coefficients can give improved behavior over ten coefficients, one might naturally ask if one could continue to improve the agreement between the MG beam model and the exact solution in the very near field by using, say, 20–25 coefficients instead. Here, we will show that the answer is a qualified yes. However, we will also postulate that the amount of improvement that one can obtain by using more Gaussians is ultimately limited by the fact that Gaussian beams are only approximate (paraxial) solutions to the wave equation. This paraxial approximation begins to fail when one is about one transducer diameter from the face, so that it is unlikely that one can accurately model the wave field a distance less than approximately a transducer diameter by using more Gaussians in the beam representation. This limitation is supported by our numerical results.

II. SPATIAL DOMAIN METHODS

A. Multi-Gaussian beams

All the MG beam models discussed here model a piston transducer radiating into a fluid as a superposition of Gaussian beams [of $\exp(-i\omega t)$ time dependency]. A general MG beam model can be written in terms of Gaussians of elliptical cross section as[8]

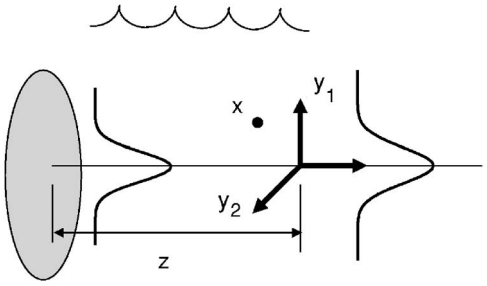


FIG. 1. A multi-Gaussian beam model uses a superposition of a small number of Gaussians to represent the velocity field on the face of the transducer. Each of these Gaussians then generates a Gaussian beam in the adjacent medium, as shown.

$$\frac{p(\mathbf{x},\omega)}{\rho_1 c_{p1} \nu_0} = \sum_{n=1}^{N} \frac{A_n}{\sqrt{\det[\mathbf{I} + c_{p1} z \mathbf{M}_n(0)]}} \times \exp[ik_{p1}z] \exp\left[i\omega \left(\frac{1}{2} \mathbf{y}^T \mathbf{M}_n(z) \mathbf{y} \right) \right], \tag{1}$$

where p is the pressure, ρ_1, c_{p1} are the density and wave speed of the fluid, and ν_0 is the velocity on the face of the transducer The point $\mathbf{x}$ at which the pressure is calculated is given by $\mathbf{x}=(y_1,y_2,z)$ (see Fig. 1) where $\mathbf{y}=(y_1,y_2)$ and $k_{p1}=\omega/c_{p1}$ is the wave number. The 2×2 matrices $\mathbf{M}_n(z)$, are given by

$$\mathbf{M}_n(z) = \frac{1}{\Delta_n} \begin{bmatrix} [M_n(0)]_{11} + c_{p1} z \det[\mathbf{M}_n(0)] & [M_n(0)]_{12} \\ [M_n(0)]_{21} & [M_n(0)]_{22} + c_{p1} z \det[\mathbf{M}_n(0)] \end{bmatrix}, \tag{2}$$

where

$$\Delta_n = 1 + c_{p1} z \operatorname{trace}[\mathbf{M}_n(0)] + (c_{p1} z)^2 \det[\mathbf{M}_n(0)]. \tag{3}$$

The expansion coefficients $A_n, \mathbf{M}_n(0)$ are determined so as to match the specified velocity field on the transducer face ($z=0$), which for a piston source is just the constant velocity ν_0 acting over the area of the transducer face. Once those expansion coefficients are obtained, the transducer wave field at any point in the fluid can be obtained directly from Eqs. (1)–(3). Thus, the determination of these coefficients is a key element of MG beam models. It should be noted that the Gaussian beams contained in Eq. (1) are only approximate paraxial solutions to the wave equation so that the MG beam model is also a paraxial solution.[9] Generally speaking, the paraxial approximation in this case assumes that the waves are all traveling in approximately the z-direction. Since an ultrasonic transducer is designed in the first place to produce a well collimated beam of ultrasound traveling normal to the transducer face, this paraxial approximation is likely well satisfied for a large portion of the transducer beam, a point we will discuss further later.

B. Circular piston transducers

Wen and Breazeale[1] used a special form of Eqs. (1)–(3) to model a circular plane piston transducer of radius a as a superposition of ten Gaussian beams. Because of the symmetry of the velocity field on the transducer face, they chose a set of circular cross section Gaussians given by

$$[M_n(0)]_{12} = [M_n(0)]_{21} = 0,$$

$$[M_n(0)]_{11} = [M_n(0)]_{22} = \frac{iB_n}{c_{p1} D_r}, \tag{4}$$

where $D_r = k_{p1} a^2/2$ is the *Rayleigh distance*. In this case, the MG beam solution, Eq. (1), becomes

$$\frac{p(\mathbf{x},\omega)}{\rho_1 c_{p1}\nu_0} = \sum_{n=1}^{10} \frac{A_n}{1+iB_n z/D_r} \times \exp[ik_{p1}z]\exp\left[i\omega\left(\frac{1}{2}\mathbf{y}^T\mathbf{M}_n(z)\mathbf{y}\right)\right], \quad (5)$$

and from Eqs. (2) and (3) we find

$$[M_n(z)]_{11} = \frac{[M_n(0)]_{11}}{1+c_{p1}z[M_n(0)]_{11}} = \frac{iB_n/c_{p1}D_r}{1+iB_n z/D_r},$$

$$[M_n(z)]_{22} = \frac{[M_n(0)]_{22}}{1+c_{p1}z[M_n(0)]_{22}} = \frac{iB_n/c_{p1}D_r}{1+iB_n z/D_r},$$

$$[M_n(z]_{12} = [M_n(z)]_{21} = 0. \quad (6)$$

Thus, A_n, B_n are the expansion coefficients that need to be determined to match the velocity field on the face of the transducer. For a circular planar piston transducer of radius a, this normalized velocity field, ν_z/ν_0, is given by the circ function, where

$$\frac{\nu_z(x,y,z=0,\omega)}{\nu_0} = \begin{cases} 1 & \rho^2/a^2 < 1 \\ 0 & \text{otherwise} \end{cases} \equiv \text{circ}(\rho^2/a^2), \quad (7)$$

and where $\rho^2 = y_1^2 + y_2^2$. From Eqs. (4)–(6) then it follows that we must have:

$$\text{circ}(\rho^2/a^2) = \sum_{n=1}^{10} A_n \exp\{i\omega[M_n(0)]_{11}y_1^2/2 + i\omega[M_n(0)]_{22}y_2^2/2\} = \sum_{n=1}^{10} A_n \exp[-B_n\rho^2/a^2]. \quad (8)$$

In Wen and Breazeale's original paper,[1] the expansion coefficients were chosen to minimize an objective function, J, where

$$J(A_n,B_n) = \int_0^\infty \left[\text{circ}(\rho^2/a^2) - \sum_{n=1}^{10} A_n \exp[-B_n\rho^2/a^2]\right]^2 d\rho. \quad (9)$$

Wen and Breazeale also gave explicitly the ten complex expansion coefficients, A_n, B_n obtained in this manner.[1] These ten Gaussians do a remarkably good job of representing the transducer wave field, even in the near field. Figure 2, for example, compares the exact on-axis behavior of a piston probe to the 10-coefficient multi-Gaussian (MG) beam model for a 6 mm radius circular transducer radiating into water at a frequency of 5 MHz. It can be seen that the MG accurately predicts the wave field to approximately a distance of 17 mm from the transducer face. Other authors have used these 10-coefficients and the propagation/transmission/reflection laws for Gaussians for curved interfaces and/or anisotropic solids to solve some very complex problems.[8–13] However, Wen and Breazeale in a subsequent paper that is not as well known[6] obtained even

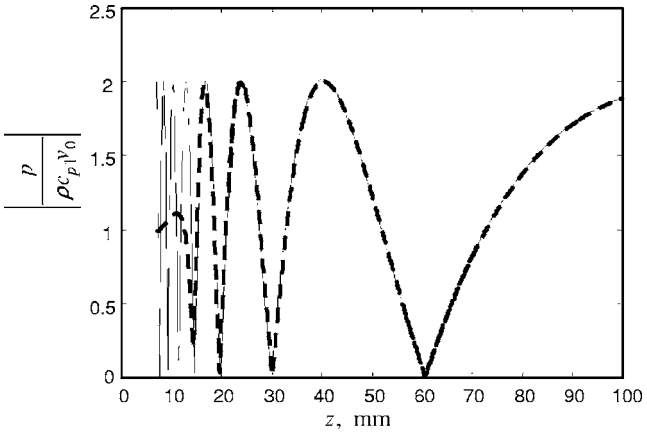


FIG. 2. Comparison of the magnitude of the on-axis pressure for a circular, planar 6 mm radius transducer radiating into water at 5 MHz. Solid line—exact solution, dashed line—MG beam solution using the 10 coefficients of Wen and Breazeale (Ref. 1).

better results with a slightly larger number of coefficients and a modified solution procedure. In Ref. 6 they used 15 coefficients and noted that on the axis of a circular piston probe the exact pressure can be obtained explicitly as

$$\tilde{p}_{\text{exact}}(z,\omega) = \frac{p(z,\omega)}{\rho_1 c_{p1}\nu_0} = \exp(ik_{p1}z) - \exp[ik_{p1}\sqrt{a^2+z^2}], \quad (10)$$

while the MG beam model gives

$$\tilde{p}_{\text{MG}}(z,\omega) = \frac{p(z,\omega)}{\rho_1 c_{p1}\nu_0} = \sum_{n=1}^{15} \frac{A_n}{1+iB_n z/D_r} \exp[ik_{p1}z]. \quad (11)$$

They then used these expressions to define a modified objective function given by

$$J(A_n,B_n) = \int_0^\infty \left[\text{circ}(\rho^2/a^2) - \sum_{n=1}^{15} A_n \exp[-B_n\rho^2/a^2]\right]^2 d\rho + \lambda_w \int_{z_1}^{z_2} [|\tilde{p}_{\text{exact}}(z,\omega)| - |\tilde{p}_{\text{MG}}(z,\omega)|]^2 dz, \quad (12)$$

where λ_w is a weighting constant to give a weight to the on-axis matching conditions relative to the boundary matching conditions, and z_1 and z_2 are near field values that define the range where matching is to take place. Figure 3 shows the results of using these 15 coefficients obtained in this "optimized" fashion. These fifteen coefficients now gives accurate results down to approximately 10 mm for the same 6 mm radius transducer case considered in Fig. 2, picking up the behavior of two more near field on-axis pressure oscillations. The fifteen expansion coefficients were given explicitly in Ref. 6 but they were also published in Ref. 14 which is a more readily accessible reference.

With these 15 "optimized" coefficients one has a remarkably efficient MG beam model that is applicable over a wide range of points in the transducer wave field. Although similar coefficients can be obtained for other velocity profiles, the piston model is often able to simulate many commercial ultrasonic transducers, so that it may not be necessary to consider this type of generalization. However,

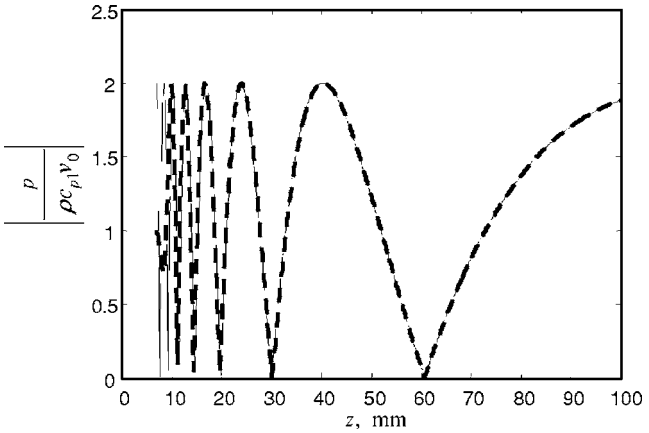


FIG. 3. Comparison of the magnitude of the on-axis pressure for a circular, planar 6 mm radius transducer radiating into water at 5 MHz. Solid line—exact solution, dashed line—MG beam solution using the 15 "optimized" coefficients of Wen and Breazeale (Refs. 5 and 13).

focused transducers are commonly used and they can also easily be modeled with these same coefficients by noting that in the paraxial approximation, focusing can be simulated by including a complex exponential term that has a quadratic spatial variation. For example, for a spherically focused piston transducer of radius a having geometrical focal length, F, the normalized velocity on the face of the transducer then becomes

$$\nu_z(y_1,y_2,z=0,\omega)/\nu_0 = \mathrm{circ}(\rho^2/a^2)\exp(-ik_{p1}\rho^2/2F). \quad (13)$$

Comparing Eqs. (8) and (13) it follows that to include the effects of spherical focusing one merely needs to modify the expansion coefficients B_n for the planar case by making the replacement

$$B_n \rightarrow B_n + \frac{ik_{p1}a^2}{2F}. \quad (14)$$

C. Elliptical and rectangular piston transducers

Recently, Ding *et al.*[3] showed that by clever use of the circ function representation of Eq. (8) one can use the same ten A_n, B_n expansion coefficients obtained by Wen and Breazeale[1] to also accurately simulate the wave fields of elliptical and rectangular transducers. Consider first a planar elliptical piston transducer with semimajor axes of the transducer face given by (a_1, a_2) in the (y_1, y_2) directions, respectively. In this case the velocity field on the transducer surface is given by

$$\frac{\nu_z(x,y,z=0,\omega)}{\nu_0} = \begin{cases} 1 & y_1^2/a_1^2 + y_2^2/a_2^2 < 1 \\ 0 & \text{otherwise} \end{cases} \equiv \mathrm{circ}(y_1^2/a_1^2 + y_2^2/a_2^2). \quad (15)$$

Thus, from Eq. (8) one finds

$$\begin{aligned} &\mathrm{circ}(y_1^2/a_1^2 + y_2^2/a_2^2) \\ &= \sum_{n=1}^{10} A_n \exp\{i\omega[M_n(0)]_{11}y_1^2/2 + i\omega[M_n(0)]_{22}y_2^2/2\} \\ &= \sum_{n=1}^{10} A_n \exp[-B_n y_1^2/a_1^2 - B_n y_2^2/a_2^2], \end{aligned} \quad (16)$$

from which it follows that:

$$[M_n(0)]_{11} = iB_n/c_{p1}D_{r1}, [M_n(0)]_{22} = iB_n/c_{p1}D_{r2},$$

$$[M_n(0)]_{12} = [M_n(0)]_{21} = 0, \quad (17)$$

with $D_{rm}=k_{p1}a_m^2/2$ $(m=1,2)$.

The pressure wave field in this case is given by

$$\frac{p(\mathbf{x},\omega)}{\rho_1 c_{p1}\nu_0} = \sum_{n=1}^{10} \frac{A_n}{\sqrt{1+iB_n z/D_{r1}}\sqrt{1+iB_n z/D_{r2}}} \times \exp[ik_{p1}z]\exp\left[i\omega\left(\frac{1}{2}\mathbf{y}^T\mathbf{M}_n(z)\mathbf{y}\right)\right], \quad (18)$$

where $\mathbf{M}_n(z)$ is given by forms similar to those in Eq. (6).

On the other hand, for a rectangular piston transducer with sides of lengths $(2a_1, 2a_2)$ in the (y_1, y_2) directions, respectively, the velocity field on the transducer surface is given by

$$\frac{\nu_z(x,y,z=0,\omega)}{\nu_0} = \begin{cases} 1 & |y_1/a_1| < 1, |y_2/a_2| < 1 \\ 0 & \text{otherwise} \end{cases} \equiv \mathrm{circ}(y_1^2/a_1^2)\mathrm{circ}(y_2^2/a_2^2). \quad (19)$$

Using Eq. (8) again in a product form, we find

$$\begin{aligned} &\mathrm{circ}(y_1^2/a_1^2)\mathrm{cir}(y_2^2/a_2^2) \\ &= \sum_{m=1}^{10}\sum_{n=1}^{10} A_m A_n \exp\{i\omega[M_{mn}(0)]_{11}y_1^2/2 \\ &\quad + i\omega[M_{mn}(0)]_{22}y_2^2/2\} \\ &= \sum_{m=1}^{10}\sum_{n=1}^{10} A_m A_n \exp[-B_n y_1^2/a_1^2 - B_m y_2^2/a_2^2], \end{aligned} \quad (20)$$

where

$$[M_{mn}(0)]_{11} = iB_n/c_{p1}D_{r1}, \quad [M_{mn}(0)]_{22} = iB_m/c_{p1}D_{r2},$$

$$[M_{mn}(0)]_{12} = [M_{mn}(0)]_{21} = 0. \quad (21)$$

The wave field of the rectangular piston transducer then is also given in product form as

$$\frac{p(\mathbf{x},\omega)}{\rho_1 c_{p1}\nu_0} = \sum_{m=1}^{10}\sum_{n=1}^{10} \frac{A_n}{\sqrt{1+iB_n z/D_{r1}}} \frac{A_m}{\sqrt{1+iB_m z/D_{r2}}} \times \exp[ik_{p1}z]\cdot\exp\left[i\omega\left(\frac{1}{2}\mathbf{y}^T\mathbf{M}_{mn}(z)\mathbf{y}\right)\right], \quad (22)$$

where $\mathbf{M}_{mn}(z)$ is obtained from $\mathbf{M}_{mn}(0)$ using the same forms as given in Eq. (6).

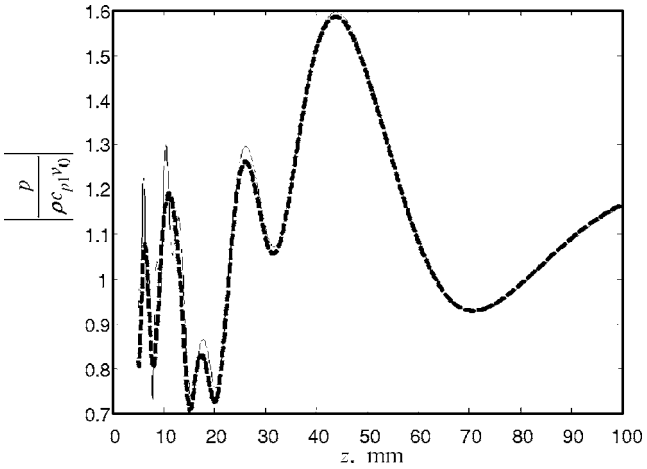


FIG. 4. Comparison of the magnitude of the on-axis pressure for a rectangular, 12 mm×6 mm planar transducer radiating into water at 5 MHz. Solid line—"exact" solution from a numerical integration of a boundary diffraction wave model, dashed line—MG beam solution using the 10 coefficients of Wen and Breazeale (Ref. 1) obtained for a circular piston transducer.

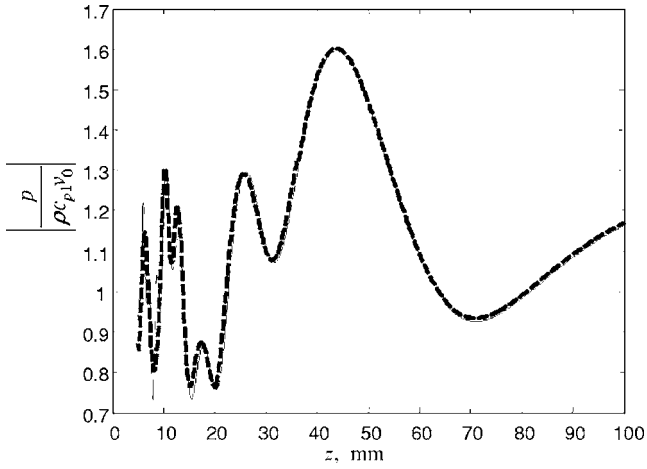


FIG. 5. Comparison of the magnitude of the on-axis pressure for a rectangular, 12 mm×6 mm planar transducer radiating into water at 5 MHz. Solid line—"exact" solution from a numerical integration of a boundary diffraction wave model, dashed line—MG beam solution using the 15 "optimized" coefficients of Wen and Breazeale (Refs. 5 and 13) obtained for a circular piston transducer.

Again, it is easy to add focusing to these results. For example, in the case of a bi-cylindrically focused rectangular transducer, one merely has to modify the B_n, B_m coefficients by making the replacements

$$B_n \rightarrow B_n + \frac{ik_{p1}a_1^2}{2F_1},$$

$$B_m \rightarrow B_m + \frac{ik_{p1}a_2^2}{2F_2}, \tag{23}$$

where (F_1, F_2) are the geometrical focal lengths in the (y_1, y_2) directions, respectively.

We used the original ten expansion coefficients of Wen and Breazeale[1] to compare the results of Eq. (22) with a highly accurate numerical integration of a boundary diffraction wave model[2] of a 12×6 mm ($a_1=6$ mm, $a_2=3$ mm) rectangular piston transducer radiating into water at 5 MHz. The results are shown in Fig. 4. It can be seen from that figure that the MG beam model accurately models the near field of the transducer to within approximately 15 mm of the transducer face. This is the same type of agreement found by Ding *et al.*[3] However, we also used the 15 expansion coefficients of Wen and Breazeale[6,14] in Eq. (22) (with 15 terms in the summations) to see how well those coefficients, which were optimized to match the on-axis fields of a circular transducer, could model this rectangular transducer. The results are shown in Fig. 5. Remarkably, those 15 coefficients match the near field structure of the rectangular transducer much better than the 10 coefficients, agreeing with the exact results to within a distance less than 6 mm from the transducer face. We should note that in these and subsequent numerical studies we will show comparisons only of on-axis values. However, the off-axis comparisons follow the same behavior of these on-axis cases so that these multi-Gaussian beam model do represent the entire transducer wave field well except very close to the transducer (or at large angles from the central axis of the transducer where the responses are very small anyway).

Since the 15 optimized expansion coefficients of Wen and Breazeale[6,14] gave better near field results than their original 10 coefficients for both circular and rectangular probes, one might expect this improvement to continue if one used a larger number of optimized coefficients. However, the fact that the MG beam model relies on the paraxial approximation limits such improvements. To understand this, consider the following simple argument. The pressure in a spherical wave generated by a point source is given by $p=\exp(ikr)/r$, where r is the distance from the source. Consider a particular radial direction as the z-axis, and make the paraxial approximation for this spherical wave in the neighborhood of that direction. Then the pressure is given approximately by $p=\exp[ik(z+\rho^2/2z)]/z$, where $\rho=\sqrt{r^2-z^2}$ is the distance from the z-axis to the point at which the pressure is being evaluated as measured in a plane perpendicular to the z-axis. If one plots the exact phase in the spherical wave versus the phase in this paraxial approximation as a function of θ, where $\tan\theta=\rho/z$ one finds that the paraxial approximation begins to lose accuracy when $\theta\cong20°$. Since a piston transducer can be modeled exactly as a superposition of such point sources acting over the face of the transducer (Rayleigh/Sommerfeld model), a paraxial approximation of those sources will begin to lose accuracy when (see Fig. 6) the angle $\theta<20°$ or at a distance approximately one diameter from the transducer face. This suggests that our MG beam model will also be unable to accurately model the wave fields at distances much less than about a diameter away from the transducer, regardless of the number of expansion coefficients. We recently conducted a series of numerical tests that have confirmed this fact. Figure 7, for example, shows a case where starting from an initial guess of A=B=1 we generated a set of 25 optimized coefficients for the 6 mm radius, 5 MHz circular transducer considered previously. This set of coefficients was able to model partially

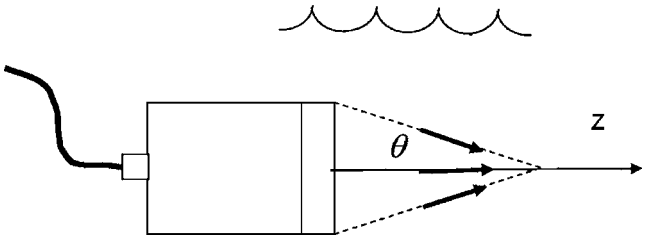


FIG. 6. A transducer radiating into water where the angle θ is defined as shown. For angles $\theta > 20°$ the phase terms of spherical waves emanating from the transducer edge will not be accurately described in the paraxial approximation.

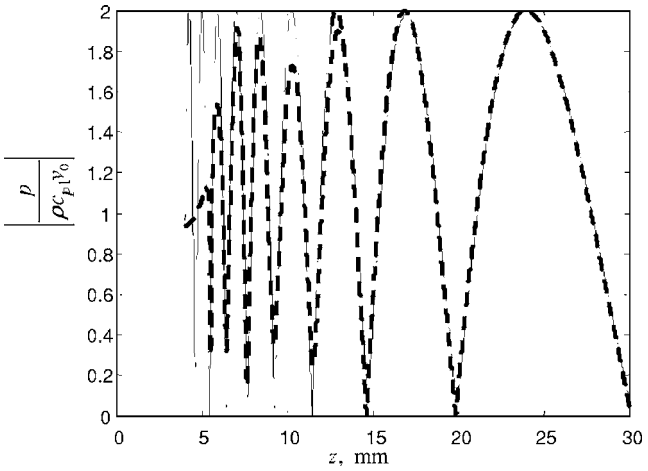


FIG. 7. Comparison of the magnitude of the on-axis pressure for a circular, planar 6 mm radius transducer radiating into water at 5 MHz. Solid line—exact solution, dashed line—MG beam solution using the 25 "optimized" coefficients of Table I.

all of the near field oscillations down to a distance of approximately 7 mm, as shown in Fig. 7, but again those near field oscillations began to lose accuracy at distances less than about a transducer diameter. These 25 expansion coefficients are listed explicitly in Table I.

It is difficult to meaningfully compare relative errors for the 10-, 15-, and 25-coefficient Gaussian beam models at single frequencies because of the highly oscillatory nature of the wave field in the near field. Thus, we have instead examined errors by evaluating the on-axis response of each beam model at many frequencies, multiplying that response by a Gaussian-shaped window function and then inverting the resulting frequency domain response into the time domain with an fast Fourier transform (FFT). The relative error was then defined in terms of differences of peak-to-peak values of the time domain waveform of the approximate model from the peak-to-peak values of the "exact" waveform generated from the Rayleigh-Sommerfeld integral. This type of comparison is also useful from a practical standpoint since transducers are often used in a pulse mode rather than a single frequency mode. Figure 8 shows a plot of the relative peak-to-peak errors versus distance obtained for a Gaussian window having a center frequency of 5 MHz and a 1 MHz bandwidth, using either 10-, 15-, or 25-coefficients. If one examines the individual curves shown in Fig. 8 carefully one can see that 25-coefficients does generally produce smaller relative errors than the 10- or 15-coefficient cases, but even the 25-coefficient case begins to exceed a relative error of 0.05 at about a transducer diameter. The same behavior was found (but not shown here) with Gaussian windows having the same band width but with center frequencies of 2.25 or

TABLE I. Twenty-five optimized spatial domain coefficients for a circular piston transducer obtained from minimizing the objective function of Eq. (12).

A_1	−0.051 932+0.074 854 i	B_1	1.9598−68.491 i
A_2	−0.001 932+0.133 38 i	B_2	2.2259−62.801 i
A_3	0.2038+0.156 04 i	B_3	2.6482−51.148 i
A_4	0.493 13−0.054 592 i	B_4	3.0329−39.309 i
A_5	−0.014 04−0.017 898 i	B_5	0.972 96−79.031 i
A_6	0.751 46−0.7956 i	B_6	3.4613−27.379 i
A_7	−4.6458−6.3564 i	B_7	4.4953−9.5366 i
A_8	17.899−9.5721 i	B_8	5.344+6.6418 i
A_9	0.020 451+0.4854 i	B_9	4.2603+45.77 i
A_{10}	−4.2364−3.8044 i	B_{10}	4.9867+17.935 i
A_{11}	0.1324−0.040 58 i	B_{11}	3.8823+60.869 i
A_{12}	−0.064 179−10.45 i	B_{12}	5.1546+12.172 i
A_{13}	−0.240 48+0.966 24 i	B_{13}	4.9921+36.952 i
A_{14}	−3.1798+0.057 147 i	B_{14}	4.8183+23.873 i
A_{15}	0.245 24+0.145 56 i	B_{15}	3.6069+52.531 i
A_{16}	−1.308+1.0953 i	B_{16}	4.658+30.099 i
A_{17}	−0.000 527−0.020 896 i	B_{17}	0.830 72+68.991 i
A_{18}	0.642 27−0.321 08 i	B_{18}	3.2362−33.351 i
A_{19}	0.086 431+0.165 29 i	B_{19}	2.4479−57.008 i
A_{20}	−0.054 28+0.011 274 i	B_{20}	1.5995−73.994 i
A_{21}	14.433+29.229 i	B_{21}	5.3897+1.3751 i
A_{22}	0.689 06−1.6732 i	B_{22}	3.7246−21.401 i
A_{23}	−0.111 12−3.4071 i	B_{23}	4.0538−15.436 i
A_{24}	−21.03+3.9134 i	B_{24}	5.0621−3.8817 i
A_{25}	0.342 17+0.090 409 i	B_{25}	2.84−45.245 i

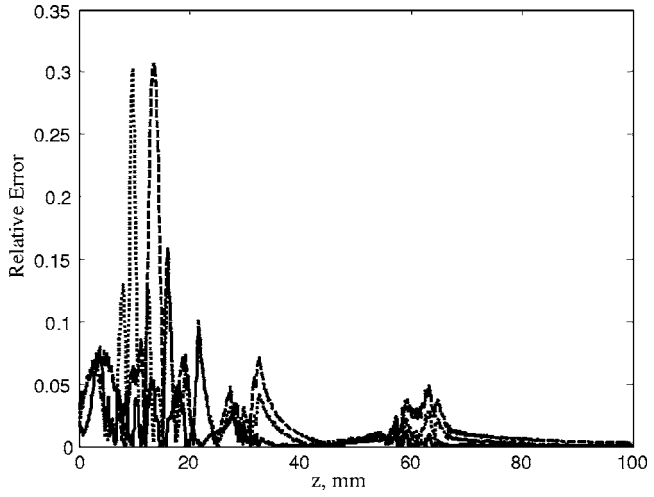


FIG. 8. Relative error of calculated peak-to-peak amplitudes vs distance for a 6 mm radius circular transducer with center frequency of 5 and 1 MHz bandwidth using 10-, 15-, and 25-optimized Gaussian beam coefficients. Dashed line—the 10 optimized coefficients, dotted line—the 15 optimized coefficients and solid line—the 25 optimized coefficients.

10 MHz. These results support our contention that while using more coefficients in a multi-Gaussian beam model can reduce errors, it is losing battle to try to gain those improvements at distances much less than a transducer diameter because of the inherent paraxial approximation errors in that region.

III. *K*-SPACE DOMAIN METHODS

Sha *et al.*[4] also have recently considered a MG beam model for rectangular piston transducers. Unlike Ding *et al.*,[3] however, they developed a set of expansion coefficients obtained by taking a 2D spatial Fourier transform of the velocity field on the transducer face and fitting a set of Gaussians in this spatial Fourier transform (*k*-space) domain. Because the Fourier (and Inverse Fourier) transform of a Gaussian is still a Gaussian, Sha *et al.*[4] were able to generate a MG beam model for the rectangular transducer of the same form shown in Eq. (21). Here, we outline briefly their procedure and results. First, we take a 2D spatial Fourier transform of the velocity field on the face of the transducer, i.e., one computes

$$\bar{\nu}_z(k_x,k_y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \nu_z(x,y,z=0,\omega) \times \exp(ik_x x)\exp(ik_y y)dxdy. \tag{24}$$

For the rectangular planar piston transducer, this spatial Fourier transform can be performed to give

$$\bar{\nu}_z(k_x,k_y) = 4a_1a_2\nu_0 \frac{\sin(k_x a_1)}{k_x a_1}\frac{\sin(k_y a_2)}{k_y a_2}. \tag{25}$$

Sha *et al.*[4] then expanded the sinc functions appearing in Eq. (25) in terms of ten *k*-space Gaussians, i.e.,

$$\frac{\sin(k_x a_1)}{k_x a_1} = \sum_{n=1}^{10} A_n^k \exp(-B_n^k k_x^2 a_1^2),$$

$$\frac{\sin(k_y a_2)}{k_y a_2} = \sum_{m=1}^{10} A_m^k \exp(-B_m^k k_y^2 a_2^2), \tag{26}$$

and the *k*-space coefficients, A_n^k, B_n^k were obtained by minimizing the objective function

$$J(A_n^k,B_n^k) = \int_0^{\infty} \left| \frac{\sin u}{u} - \sum_{n=1}^{10} A_n^k \exp(-B_n^k u^2) \right| du, \tag{27}$$

where $u=k_x a_1$ or $u=k_y a_2$. They gave in a table the 10 complex *k*-space coefficients obtained in this manner.[4] This approach works since, as mentioned previously, one can invert the *k*-space Gaussians back into the spatial domain and obtain an equivalent spatial domain Gaussian representation of the wave field on the surface of the transducer as

$$\frac{\nu_z(x,y,z=0,\omega)}{\nu_0} = \frac{1}{\pi}\sum_{n=1}^{10}\sum_{m=1}^{10} \frac{A_n^k}{\sqrt{B_n^k}}\frac{A_m^k}{\sqrt{B_m^k}} \exp(-y_1^2/4B_n^k a_1^2) \times \exp(-y_2^2/4B_m^k a_2^2). \tag{28}$$

They then wrote the corresponding pressure wave field in terms of these *k*-space coefficients. However, if one compares Eqs. (20) and (28), one can relate these *k*-space coefficients directly to a corresponding set of spatial domain coefficients given by

$$A_n = \frac{A_n^k}{\sqrt{\pi B_n^k}} \quad B_n = \frac{1}{4B_n^k}, \tag{29}$$

with similar definitions for A_m,B_m. These relations are not unexpected since similar relations exist between a Gaussian and its Fourier transform.[2] Note, however, that the spatial domain coefficients obtained in this fashion are not the Wen and Breazeale coefficients[1] since the fitting procedure is different and there is no uniqueness to the coefficients obtained by these methods. In general, these ten *k*-space coefficients produce very similar on-axis results for a rectangular transducer to those obtained when using the 10 Wen and Breazeale coefficients.

The *k*-space approach would appear to have an advantage over the spatial domain method in that the *k*-space function being approximated by the Gaussians expansions $[\sin(u)/u]$ is smooth, well behaved function whereas the circ function is discontinuous. However, this advantage is offset by the fact that in order to reproduce the near field behavior accurately one needs to approximate the *k*-space function to large values in *k*-space, where the function becomes small. This can be seen in Fig. 9, which shows the fitting of the $\sin(u)/u$ function for a rectangular transducer case to 10 *k*-space Gaussians. To obtain accurate near field values one needs to use a range of *k*-space values from approximately 0 to 50 and at the largest values it becomes increasingly difficult to match the small amplitude oscillatory values of the *k*-space function.

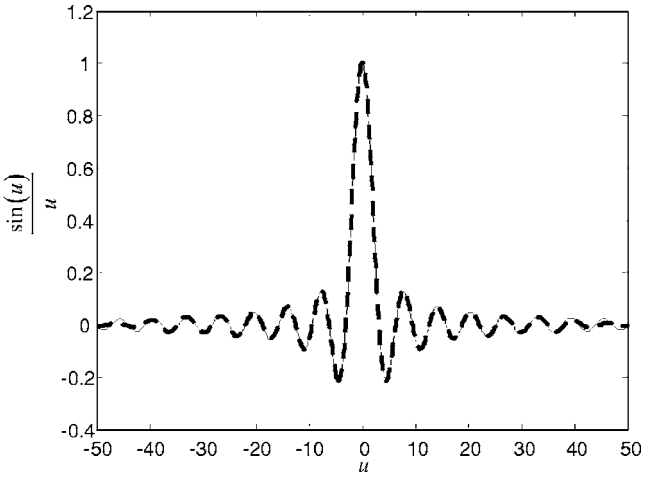


FIG. 9. Fitting of the $\sin(u)/u$ function in k-space for a rectangular planar piston transducer. Solid line—exact function, dashed line—MG beam solution using ten k-space Gaussian functions.

IV. SUMMARY AND OVERVIEW

A multi-Gaussian beam model is a remarkably fast and effective simulation model since it can reproduce the wave field of circular, elliptical, and rectangular piston transducers radiating into a fluid everywhere except in the very near field by simply adding up a number of Gaussian beams. The number of terms needed is only 10–15 for circular and elliptical probe shapes and 100–225 for rectangular transducers. Even with these modest number of Gaussians the pressure of the transducer can be accurately predicted everywhere in the wave field except within a region approximately one diameter distance from the transducer face. Spherically and bicylindrically focused transducers can also be easily modeled with the same number of terms and also give accurate field predictions as long as the focusing is not too severe. The Gaussian beams described here all have simple circular or elliptical cross sections described by 2×2 **M** matrices [see Eq. (1)] where $[\mathbf{M}]_{12}=[\mathbf{M}]_{21}=0$. When these beams cross curved interfaces or are transmitted or reflected in anisotropic elastic media, the Gaussian beams change to the more general form given in Eq. (1) where $[\mathbf{M}]_{12}=[\mathbf{M}]_{21}\neq 0$, but one can obtain explicit, analytical transmission/reflection and propagation laws that transform the **M** matrices given here into the more general forms needed in those cases.[7] Thus, a multi-Gaussian beam model retains the efficiency demonstrated here for a single fluid medium in a much more general setting.

The expansion coefficients needed for a multi-Gaussian beam model can be obtained for circular, elliptical and rectangular transducers using either spatial domain methods or k-space domain methods. As shown here, there is a simple relationship [see Eqs. (29)] between the coefficients obtained in either domain. However, the explicit functions and forms involved in the minimization procedures used to obtain the expansion coefficients are quite different in the two domains. In general, it appears that the spatial domain approach where one can obtain "optimized" expansion coefficients through an objective function such as the one given in Eq. (12) offers the most effective approach. The 25 "optimized" coefficients of Table I or the 15 "optimized" Wen and Breazeale coefficients[6,14] both give excellent wave field predictions for both circular and rectangular transducers to within the very near field. Even in that very near field these coefficients come very close to giving results up to the limits of the paraxial approximation itself, as discussed previously. Thus, at present either the 25 "optimized" expansion coefficients given here or the 15 "optimized" coefficients of Wen and Breazeale[6,14] appear to offer the best choices for generating a multi-Gaussian beam model.

ACKNOWLEDGMENTS

For H. J. Kim and L. W. Schmerr this work was supported by the National Science Foundation Industry/University Cooperative Research Program at the Center for NDE, Iowa State University. A. Sedov would like to acknowledge the support of the past Natural Sciences and Engineering Research Council of Canada.

[1] J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756 (1988).
[2] L. W. Schmerr, *Fundamentals of Ultrasonic Nondestructive Evaluation—A Modeling Approach* (Plenum, New York, 1998).
[3] D. Ding, Y. Zhang, and J. Liu, "Some extensions of the Gaussian beam expansion: Radiation fields of rectangular and elliptical transducers," J. Acoust. Soc. Am. **113**, 3043–3048 (2003).
[4] K. Sha, J. Yang, and W. S. Gan, "A complex virtual source approach for calculating the diffraction beam field generated by a rectangular planar source," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 890–896 (2003).
[5] J. Yang, K. Sha, W. S. Gan, and J. Tian, "Nonlinear wave propagation for a parametric loudspeaker," IEICE Trans. Fundamentals **E87-A**, 2395–2400 (2004).
[6] J. J. Wen and M. A. Breazeale, "Computer optimization of the Gaussian beam description of an ultrasonic field," in *Computational Acoustics; Scattering, Gaussian Beams, and Aeroacoustics*, Vol. 2, edited by D. Lee, A. Cakmak, and R. Vichnevetsky (Elsevier Science, North Holland, 1990), pp. 181–196.
[7] D. Ding and Y. Zhang, "Notes on the Gaussian beam expansion," J. Acoust. Soc. Am. **116**, 1401–1405 (2004).
[8] V. Cerveny, *Seismic Ray Theory* (Cambridge University Press, Cambridge, U. K., 2001).
[9] A. E. Siegman, *Laser* (University Science Books, Sausalito, California, 1986).
[10] A. N. Norris, "A theory of pulse propagation in anisotropic elastic solids," Wave Motion **9**, 509–532 (1987).
[11] M. Spies, "Transducer field modeling in anisotropic media by superposition of Gaussian basis functions," J. Acoust. Soc. Am. **105**, 633–638 (1999).
[12] M. Rudolph, "Ultrasonic beam models in anisotropic media" (PhD dissertation, Iowa State University, IA, 1999).
[13] L. W. Schmerr, "A multi-Gaussian ultrasonic beam model for high performance simulations on a personal computer," Mater. Eval. **58**, 882–888 (2000).
[14] D. Huang and M. A. Breazeale, "A Gaussian finite-element method for description of sound diffraction," J. Acoust. Soc. Am. **106**, 1771–1781 (1999).

On using polynomial chaos for modeling uncertainty in acoustic propagation

Dennis B. Creamer
U.S. Naval Research Laboratory, Washington, D.C. 20375

(Received 20 June 2005; revised 5 January 2006; accepted 18 January 2006)

The use of polynomial chaos for incorporating environmental variability into propagation models is investigated in the context of a simplified one-dimensional model, which is relevant for acoustic propagation when the random sound speed is independent of depth. Environmental variability is described by a spectral representation of a stochastic process and the chaotic representation of the wave field then consists of an expansion in terms of orthogonal random polynomials. Issues concerning implementation of the relevant equations, the accuracy of the approximation, uniformity of the expansion over the propagation range, and the computational burden necessary to evaluate different field statistics are addressed. When the correlation length of the environmental fluctuations is small, low-order expansions work well, while for large correlation lengths the convergence of the expansion is highly range dependent and requires high-order approximants. These conclusions also apply in higher-dimensional propagation problems. [DOI: 10.1121/1.2173523]

PACS number(s): 43.20.Bi, 43.20.Mv, 43.30.Re [JBS] Pages: 1979–1994

I. INTRODUCTION

Sound-speed variability due to oceanic heterogeneities leads to distortion and variability of acoustic signals propagating in such an environment. This temporal and spatial variability is caused by oceanic processes with dynamics that is inherently nonlinear. This nonlinearity induces decorrelation (in time) and, by the ergodic hypothesis, synoptic snapshots can be treated as drawn from a stochastic ensemble. Depending on the acoustic frequency and the particular environment considered, this acoustic variability imposes physical limits on sonar detection systems. The theoretical basis for understanding acoustic propagation through an unbounded, fluctuating ocean is presented in Flatté *et al.*[1] In unbounded environments, wave propagation through random media (WPRM) is qualitatively understood in both strong and weak scattering regimes. In both these cases, the stochastic behavior of the wave field is well described.

The predicative capabilities of modeling and computational efforts that incorporate this variability depend crucially on a quantitative description of the environment. Uncertainties in this description may affect the acoustic modeling and there is usually no simple way to determine changes to the acoustic field. The modeling work of Flatté and others characterizes these uncertainties in the propagation medium by a few gross measures, typically medium fluctuation strength and medium correlation lengths. For example, the $\Lambda\Phi$ diagrams used by Flatté *et al.* allow the prediction of the pressure field, acoustic intensity, intensity variance, and intensity cross correlation in various asymptotic regimes (usually determined by the relative size of the propagation range). This has had a certain utility, e.g., in work on tomographic inversion.[2] The unexpectedly high-intensity variance due to oceanic internal waves for a certain acoustic ray led to the conclusion that this ray was really the sum of two rays with arrival times that were close together but wandered. However, these measures are generally not that useful; most experimental situations are not in an asymptotic region of the $\Lambda\Phi$ diagram, while even the data addressed by Flatté *et al.* do not fall into one of these regions.

In a general modeling framework, this approach is less fruitful in that no simple quantitative estimate of the uncertainty is simply calculable. The $\Lambda\Phi$ parameters could be calculated (after somewhat lengthy analysis) and still not give an unambiguous estimate of the errors in the physical quantities, because the problem is not usually in one of the asymptotic regions of the $\Lambda\Phi$ diagram. Strictly *deterministic* equations can be derived for the important physical quantities (mean field, intensity, scintillation, etc.). However, it is usually necessary to invoke some other approximation[3] (such as closure, the Markov approximation, etc.) in order to secure partial differential equations that can then be solved numerically. Extensions beyond these simplifications or use of other measures of environmental uncertainty in the modeling process are not available and thus the modeler cannot readily infer the propagation uncertainty due to environmental variability.

Contemporary modeling efforts often use Monte Carlo simulations to calculate the statistics of the acoustic field,[4] which give all the relevant results, in principle. The environmental variables are drawn as samples from the appropriate stochastic process and deterministic propagation is computed. Statistical characteristics of the acoustic field are determined by sampling the probability distributions appropriately. There are two implicit assumptions in using this approach: (1) the probability distribution describing environmental variables is known or can be modeled with a few parameters; and (2) the sampling appropriate for the acoustic statistics can be done. As there are many independent sources of environmental variability, assumption (1) usual requires tremendously simplifying the environment with little ability to determine how these simplifications affect the acoustic field. Assumption (2) is sometimes problematic as you cannot easily take appropriate Monte Carlo samples

when the acoustic field is intermittent[5] (i.e., there are very large intensity fluctuations). Even when these assumptions are valid the size of the sampling ensemble is necessarily quite large. Looking at the effect of changes in the stochastic process underlying the environment requires separate Monte Carlo simulations for each change.

Finette[6] has suggested that polynomial chaos might be useful for incorporating oceanic variability in underwater acoustic propagation and has outlined how it might be used. Polynomial chaos describes the observation that any function of a stochastic variable can be written as an orthogonal polynomial expansion in terms of the underlying set of random variables (called the chaotic representation[7] in the stochastics literature). In this representation, ensemble averages over the probability distribution of the random variables correspond to inner products of functions, treating the random variables as deterministic integration variables. The orthogonal polynomials form a Hilbert space and the expansions are mean-square convergent in this space. The inner products naturally incorporate the probability distribution function as the weighting function defining the family of orthogonal polynomials. Any member of the Wiener-Askey class of multivariate orthogonal functions[7] can be basis functions for this Hilbert space, with each different choice of basis functions associated with different probability distributions. For Gaussian random variables the appropriate polynomials are versions of the Hermite polynomials, $H_n(x)$, with a weighting function $\exp(-x^2/2)$, discussed in the Appendixes.

Use of the expansion represented by the chaotic representation in wave propagation problems corresponds to a reordering of the perturbation series, where the stochastic variables are treated as the perturbation. The reordering occurs by applying the stochastic averaging process and rearranging the contributions in each order. For example, the second-order (i.e., two stochastic variables) contribution is rewritten as a number (the stochastic average) plus other pieces that vanish on averaging (in this case by adding and subtracting the stochastic average). For the case of a Gaussian distribution, this latter term would just be the second Hermite polynomial. In the diagrammatic language of Codona *et al.*,[8] the terms that vanish (on averaging) correspond to connected diagrams.

Polynomial chaos has seen a resurgence of interest in the engineering community,[9] where it can be used in conjunction with finite-element modeling to extract the effect of uncertainties in the description of structures. There has been some work on using these expansions in stochastic differential equations. If the weighting function, defining the orthogonal polynomials, corresponds to the probability distribution function describing the uncertainty, then the expansion is actually exponentially convergent (i.e., errors decrease exponentially with increasing order of approximation). However, for stochastic differential equations this expansion is not guaranteed to be uniform over the propagation range, i.e., longer ranges may require more terms in the expansion to achieve the same relative accuracy. At any propagation range the expansion is convergent, i.e., given any error there exists a number N (the number of terms in the expansion) where the difference between the expansion and the exact result is within this error tolerance. Mathematically then, the expansion is uniform over a finite range interval (just take the maximum value of N over all of the points in the range interval). However, increasing range intervals require higher truncation levels and an important computational issue, not addressed in the literature, is how the number of terms in the chaotic representation needed to achieve good results scales with propagation range.

Finette[6] has sketched the derivation for underwater acoustic propagation. The point of the present work is to implement the chaotic representation in detail and to investigate the issue of scaling of the number of terms of the expansion with the propagation range and how well the various approximations work. Rather than deal with a realistic environment, I look at a simple one-dimensional stochastic differential equation. This "toy" problem will be useful to provide detail on implementing the chaotic expansion, allow the issue of uniformity of the expansion to be quantitatively addressed, determine how well various acoustic statistics are approximated, and investigate the complexity of numerical implementation. These issues are common whether one, two, or three-dimensional stochastic differential equations are considered, and this toy problem provides valuable insight into realistic acoustic problems. The one-dimensional case is simplest and the exact answer is analytically solvable so that the approximation errors are easily determined. In higher dimensions, the main differences will be a dramatic increase in the number of stochastic components and the inclusion of a Laplacian in the space variables orthogonal to the propagation direction. The main effect of the Laplacian is to diffuse energy and thus ameliorate any singularities that might appear. The other main effect in higher dimensions that does not appear in one dimension is the existence of multipaths, which leads to multiple interference phenomena and a nonzero scintillation index. This interference does not exist in one dimension, implying the scintillation index is zero. Finally, a higher-dimensional problem of recent interest[10] is one is which the random perturbation is only range dependent (i.e., there is no dependence of the sound-speed perturbation on the coordinates orthogonal to the propagation direction). This work[11] is the extension of Sec. III of this paper to higher dimensions. For this case, the wave field factorizes into a product of a deterministic wave function and a stochastic one. The stochastic piece obeys the equation of the toy model and so the considerations of this paper are directly applicable to such a problem.

If the medium correlation length is short (e.g., as necessary for the Markov approximation to be valid), then quite good approximations can be had by including just a few terms in the expansion. This is not too surprising since, if the medium fluctuations are decorrelated one range step to the next, the problem reduces to a sequence of phase screens. If the medium correlation length is larger than the other physical scales, the expansion is not convergent over all range intervals, i.e., larger ranges will require larger number of terms.

In the sequel, I present the derivation of the chaotic representation for a simple stochastic differential equation. The preliminaries and notation are given in Sec. II. In Sec. III I

look at the case where the medium correlation length is very large (i.e., medium fluctuations act coherently) and find that the issue of how many terms of the expansion are needed at large ranges is important. The scaling of the order of approximation (to achieve a given error tolerance) with range is determined. The intensity variance (fourth-order field statistic) is completely unreliable when calculated in the most straightforward way (as the intensity squared) from the truncated expansion. This will also be true in higher-dimensional problems. I propose a modification that is correct if the exact results are used in the expression and thus converges quickly to the correct result. Section IV presents the derivation of the propagation equation in the general case, while Sec. V looks at the simplest truncations in the case where the medium correlation function decays exponentially. Section VI summarizes the results.

II. PRELIMINARIES

A. Exact results

The basic stochastic propagation equation for the field ψ is

$$\frac{d\psi}{dr} \equiv \dot{\psi} = i\xi(r)\psi; \quad \psi(0) = 1, \tag{1}$$

where r is the propagation distance and i is the square root of -1. The function $\xi(r)$ is a weighted sum of zero-mean random variables each governed by the normal (Gaussian) distribution. It is a zero-mean stochastic process and has the correlation function

$$\langle \xi(r)\xi(r')\rangle = \bar{D}(r-r') \equiv \sigma^2 D(r-r'), \tag{2}$$

where $\sigma^2 = \bar{D}(0)$ is a measure of the fluctuation strength. Note this form implies $D(0)=1$. The brackets $\langle \ldots \rangle$ denote an ensemble average over the Gaussian random variables. There are at least two physical quantities naturally contained in the function $\bar{D}$: (1) the fluctuation strength, σ, which has dimensions of length^{-1} and (2) the correlation length, L, of the propagation medium. The quantity $1/\sigma$ is a length scale over which the standard deviation of the phase of the wave function changes significantly. The correlation length is the scale separating coherent action of the fluctuations from the incoherent action, i.e., for $|r-r'| \ll L$ then $D(r) \approx D(0) = 1$, while for $|r-r'| \gg L$ then $D(r) \to 0$. The quantity $|r-r'|$ *is* the distance between fluctuations. There are many ways to define this L; however, the physical meaning is clear. In the propagation problem we have three quantities representing a length scale: L, $1/\sigma$, and r (the propagation range). We are interested in determining quantities as a function of r, so the only dimensionless parameter available for characterizing solutions is σL, depending only on medium parameters.

The solution of the propagation equation is straightforwardly determined as

$$\psi(r) = \exp\left[i\sigma \int_0^r \xi(r')dr' \right]. \tag{3}$$

With $\xi(r)$ a zero-mean Gaussian process, expectation values are simply obtained, e.g.,

$$\langle \psi(r)\rangle = \exp\left(-\frac{\sigma^2}{2} \int_0^r \int_0^r D(r'-r'')dr'dr'' \right)$$
$$= \exp\left(-\sigma^2 \int_0^r (r-r')D(r')dr' \right) \tag{4}$$

for the average field, while the field covariance is

$$C(r-r') \equiv \langle \psi^*(r)\psi(r')\rangle$$
$$= \exp\left(-\sigma^2 \int_0^{|r-r'|} (|r'-r| - r'')D(r'')dr'' \right). \tag{5}$$

Note the intensity, $I = \psi^*(r)\psi(r)$, is identically 1 as is I^2. The latter leads to the conclusion that the scintillation index (i.e., intensity variance) is zero. Any putative approximation scheme should respect these conditions. When the length L, characterizing the function D, is large or small compared to the arguments appearing in these integrals, the behavior of the correlation function noted above gives a limiting form for the mean field

$$\langle \psi(r)\rangle \approx \exp\left(-\sigma^2 (r-r_0) \int_0^\infty D(r')dr' \right) = e^{-\sigma^2 L(r-r_0)} \tag{6}$$

for $r \gg L$ (i.e., small correlation length as compared to propagation range) and

$$\langle \psi(r)\rangle \approx \exp(-\sigma^2 r^2/2), \tag{7}$$

for $r \ll L$. Here, $r_0 = \int_0^\infty r' D(r')dr' / \int_0^\infty D(r')dr'$, and the quantity $L \equiv \int_0^\infty D(r')dr'$ is taken as a definition of the medium correlation length. The behavior of these limiting forms can be quite different depending on the size of the dimensionless parameter σL.

The relevant scale for variation in r of the mean field is $1/(\sigma^2 L)$ for small correlation lengths, Eq. (6), and $1/\sigma$ for large correlation lengths, Eq. (7). The dimensionless parameter, σL, relates these two scales and is used to classify the expected solution for the mean field. A small or large value of this parameter does not imply that the correlation length, L, is correspondingly smaller or larger than the propagation range. However, effectively this link can be made because the situations where it is not true does not change the corresponding behavior of the mean field.

For large σL (i.e., $L \gg 1/\sigma$), Eq. (7) gives the mean field for $r \ll L$. In principle, when $r \gg L$ (i.e., the correlation length is not large), the mean field transitions to Eq. (6) but is already vanishingly small at these ranges, e.g., for $r=L$ the mean field is $\exp(-\sigma^2 L^2/2)$, which is small when σL is large. While the propagation scale $(\sigma^2 L)^{-1}$ is much smaller than σ^{-1}, it is irrelevant in this case. For the practical purposes of characterizing the behavior of the solutions, large values of σL then imply large correlation lengths (meaning larger than all the other physical scales). Section III investigates the use

of the polynomial chaos expansion in the limit $\sigma L \to \infty$, where the analysis and numerics are simpler than for the general case.

When σL is small the mean field behaves as Eq. (6) for $r \gg L$ (i.e., correlation length smaller than the propagation range). When $r \ll L$ Eq. (7) is the relevant equation but in this regime $\sigma r \ll 1$ as well and thus the mean field is unity anyway, as it is in Eq. (6). For small σL the propagation scale σ^{-1} satisfies $L \ll \sigma^{-1} \ll (\sigma^2 L)^{-1}$ but is not physically relevant as there is not any transitional behavior (at this scale) in Eq. (6); it is unity until the scale $(\sigma^2 L)^{-1}$ is reached. Example approximations of the polynomial chaos solutions for different values of σL are studied in Sec. V, where general solutions of the polynomial chaos expansions need to be used.

B. Medium description

The stochastic function $\xi(r)$ is best described as an expansion in terms of a set of independent stochastic variables, ξ_n, each of which is a zero-mean, complex Gaussian random variable. These stochastic modes represent the underlying cause of the variability of the sound speed, e.g., in Dozier and Tappert,[4] who refer to the expansion as the Karhunen-Loève representation, these represent internal wave fluctuations. In the engineering literature this goes by the name of empirical orthogonal eigenfunction (EOF). We use the spectral expansion

$$\xi(r) = \sigma \sum_{k \in K} f_k e^{-ikr} \xi_k, \tag{8}$$

where K is an appropriate set of wave numbers. For a finite Fourier series these would the $2N$ numbers $\{-(1-1/N)k_{\max}, \ldots, -k_{\max}/N, 0, k_{\max}/N, \ldots, k_{\max}\}$. Notice that ξ_k does not depend functionally on the index k, which only serves as a label for the different random variables. Thus, there is a set of $2N$ statistically independent, zero-mean, complex Gaussian random variables (stochastic modes) with unit variance in the case of a finite Fourier series expansion. These random variables obey the orthogonality relation $\langle \xi_p^* \xi_k \rangle = \delta_{pk}$. Because $\xi(r)$ is real, take $\xi_p^* = \xi_{-p}$ and $f_k^* = f_{-k}$. The quantity $\phi_k = |f_k|^2$ is the spectrum of statistical fluctuations (normalized by σ^2) given at the set of discrete values $k \in K$. The quantity f_k is chosen real and is just $\sqrt{\phi(k)}$. The medium correlation function is

$$D(r) = \sum_{k \in K} \phi(k) e^{ikr}. \tag{9}$$

The theoretical considerations in this paper are valid for any D; however, in the numerical work we compare with the correlation function

$$D(r) = e^{-r/L}. \tag{10}$$

The length L is just the correlation length discussed earlier. For this correlation function the mean field, $\langle \psi(r) \rangle$, and covariance, $C(r)$, are both given by

$$\exp(-\sigma^2 L^2 [r/L - (1 - e^{-r/L})]). \tag{11}$$

Notice that Eq. (11) is a function of r/L and depends on the dimensionless parameter σL. This correlation function has the Fourier transform

$$D(r) = \int_{-\infty}^{\infty} \phi(k) e^{ikr} dk, \tag{12}$$

where the fluctuation spectrum is

$$\phi(k) = \frac{2/L}{k^2 + 1/L^2} = \frac{2L}{1 + k^2 L^2}. \tag{13}$$

This correlation function does not result from a discrete expansion as in Eq. (8). In order to compare with numerical results, the integral in Eq. (12) must be approximated by a discrete set, K, of wave number values to a sufficient accuracy, using $\phi_k = \Delta k \phi(k)$ for $k \in K$, where Δk is the successive spacing between k values. For the purposes of this paper, this was accomplished by requiring a relative error less than 10^{-3}. Note that both ϕ_k and f_k are dimensionless as they should be. Because the spectrum, Eq. (13), has such a slow falloff in k space this decimation will require a very large set K in order to accurately represent Eq. (12), and provides a severe computational test for the polynomial chaos expansion.

C. Chaotic representation

The chaotic representation of the field $\psi(r)$ is

$$\psi(r) = \sum_{m=0}^{\infty} \sum_{\{k\}_m} y_m(r; \{k\}_m) H_m(\{\xi_k\}_m), \tag{14}$$

where $\{k\}_m$ denotes a set of m labels $\{k_1, \ldots, k_m\}$ which is to be summed over, e.g., k_1 is to be summed over all elements of K as is k_2, etc. In the case where the representation of the variability is as in Eq. (8), these labels correspond to the spectral parameters. The quantity $\{\xi_k\}_m$ represents the set of m random variables $\{\xi_{k_1}, \ldots, \xi_{k_m}\}$. Note that this set is not a set of independent random variables as the indices k_i can be identical [as is definitely the case in the sum of Eq. (14)]; depending on the values of the labels k_i, any particular random variable can appear multiple times in the set. The $H_m(\{\xi_k\}_m)$'s are the multivariate Hermite polynomials discussed in Appendix B; by definition they can contain multiple occurrences of the same random variable. The first few terms in the expansion are

$$y_0(r) + \sum_k y_1(r;k) H_1(\xi_k) + \sum_{k_1,k_2} y_2(r;k_1,k_2) H_2(\xi_{k_1}, \xi_{k_2}) + \sum_{k_1,k_2,k_3} y_3(r;k_1,k_2,k_3) H_3(\xi_{k_1}, \xi_{k_2}, \xi_{k_3}). \tag{15}$$

The summations are to be understood as summing over elements of K, i.e., $k \in K$, $k_1 \in K$, etc.

For N total spectral components, the first sum has N functions $y_1(r;k)$, while the second term sums over N functions $y_{\{2\}}^2(r;k) (\equiv y^2(r;k,k))$ and $N(N+1)/2$ distinct functions $y_2(r;\{k_1,k_2\})$, with $k_1 \neq k_2$. When there is a large number of

spectral components for the stochastic perturbation, the number of functions increases dramatically in the higher-order sums.

The method of polynomial chaos inserts the chaotic representation, Eq. (14), into the differential equation, Eq. (1), and projects onto the basis functions given by the Hermite polynomials (by performing expectation values). This gives an infinite set of differential equations relating the different coefficient functions y_m. Note that the stochastic quantity $\xi(r)$ in Eq. (8) is already in the form of a chaotic representation since $\xi_k = H_1(\xi_k)$.

Numerical use of the chaotic representation requires that the infinite series in Eq. (14) be approximated. In the polynomial chaos literature, the expansion itself is truncated, similar to approximating functions by polynomials. This makes sense in the use of polynomial chaos expansions in finite-element applications. For the stochastic differential equations used in wave propagation problems, this unfortunately acts both as an expansion truncation for the stochastic field and as a closure approximation to the differential equations for the coefficient functions. This leads to a subtlety concerning the sources of error in the approximation scheme. If the series is truncated with the exact results, y_m, retained as the coefficients, then the series is guaranteed to be exponentially convergent. But, these functions are not known *a priori* and are usually determined by truncating the infinite set of differential equations determining them. In the turbulence literature, the corresponding set of differential equations is often truncated by a closure approximation, usually by expressing a high-order moment in terms of products of lower-order ones. This could be done for Eq. (20) at the expense of turning a set of linear equations into nonlinear ones and losing the direct correspondence between the expansion and orthogonal polynomials. Truncating the expansion as in Eq. (14) implies a closure of the differential equations.

This distinction is important in this paper. The mean field is just the y_0 of Eq. (14) and the errors of its approximants are really errors due to closure; even if these are small (for y_0) the series, Eq. (14), might not be that good of an approximation. Other statistical quantities, such as the scintillation index, do include all of the y_m up to a certain order. As will be shown, the truncated expression for the scintillation index can be wildly wrong (depending on range) even when the exact y_m's are inserted to the truncated expression. A modified expression for the scintillation index is introduced which has the property that it vanishes when the exact y_m are used in the expression; thus, the errors in approximating this modified scintillation index are due to closure.

III. LIMIT OF LARGE CORRELATION LENGTHS

In the limit of an infinitely large medium correlation length ($\sigma L \to \infty$), the independent fluctuations act coherently and are represented by a single, range-independent random variable. It is more instructive to derive the appropriate polynomial chaos equations using this coherence

$$\dot{\psi} = i\sigma\xi\psi, \tag{16}$$

where ξ is a real zero-mean (real) Gaussian random value with unit variance.

A. Exact results

The solution and its average value are, respectively,

$$\psi(r) = e^{i\sigma\xi r} \tag{17}$$

and

$$\langle\psi(r)\rangle = \exp(-\sigma^2 r^2/2). \tag{18}$$

The intensity is identically 1 while the scintillation index is zero.

The chaotic representation for the field ψ is

$$\psi(r) = \sum_{m=0}^{\infty} y_m(r) H_m(\xi), \tag{19}$$

where the y_m's are the moment components of the field. Inserting this into the differential equation for ψ and performing a decomposition (Galerkin projection) on to the basis functions defined by the Hermite polynomials [i.e., $\langle\dot{\psi}(r)\xi\rangle = \langle\xi RHS_{\text{Eq.(16)}}\rangle$] yields the infinite set of coupled equations for the field coefficients

$$\begin{aligned} \dot{y}_0 &= i\sigma y_1, \\ \dot{y}_m &= i\sigma(\sqrt{m}\,y_{m-1} + \sqrt{m+1}\,y_{m+1}), \quad m = 1, \ldots, \infty \end{aligned} \tag{20}$$

with initial conditions $y_i(0) = \delta_{i0}$ for $i = 0, \ldots, \infty$ and solutions

$$y_m = (i\sigma r)^m \exp(-\sigma^2 r^2/2)\sqrt{m!}. \tag{21}$$

The average intensity is given by

$$\langle I\rangle = \langle\psi^*(r)\psi(r)\rangle = \sum_{m=0}^{\infty} |y_m(r)|^2 \tag{22}$$

and is unity, as can be explicitly checked from the solution set. The covariance of the field is

$$C(r-r') = \sum_{m=0}^{\infty} \frac{(\sigma^2 r r')^m}{m!} e^{-\sigma^2(r^2+r'^2)/2} = e^{-\sigma^2(r-r')^2/2}, \tag{23}$$

as expected. The deviation of the intensity from its mean is

$$\begin{aligned} I - \langle I\rangle &= \psi^*(r)\psi(r) - \sum_{m=0}^{\infty} |y_m|^2 \\ &= \sum_{m=1}^{\infty} |y_m|^2 (H_m^2 - 1) + 2 \sum_{\substack{m,n=0 \\ m>n}}^{\infty} \mathrm{Re}(y_m^* y_n) H_m H_n \\ &= \sum_{m=1}^{\infty} |y_m|^2 \sum_{j=1}^{m} \varepsilon_{mm(2j)} H_{2j} \\ &\quad + \sum_{\substack{m,n=0 \\ m>n}}^{\infty} 2\,\mathrm{Re}(y_m^* y_n) \sum_{\substack{j=m-n \\ \text{by } 2}}^{m+n} \varepsilon_{mnj} H_j, \end{aligned} \tag{24}$$

where the connection coefficients, $\varepsilon_{mnj} = \langle H_m H_n H_j\rangle$, are

given in Appendix A. From the initial condition and the form of the differential equations, note that the y_i^M's are real/imaginary functions for $\boldsymbol{i}$ even/odd, respectively. Then, $\mathrm{Re}(y_m^* y_n)=0$ if m and n are relatively odd. Thus, m-n is even, implying that only even index H_j appear. Thus,

$$I-\langle I\rangle=\sum_{m=1}^{\infty}|y_m|^2\sum_{j=1}^{m}\varepsilon_{mm(2j)}H_{2j}+2\sum_{m=0}^{\infty}\sum_{n=1}^{\infty}(y_{2n+m}^* y_m)\sum_{j=n}^{m+n}\varepsilon_{(2n+m)m(2j)}H_{2j}. \tag{25}$$

This quantity should be identically zero, i.e., the coefficient of each H_{2j} vanishes. Rearranging the summations secures

$$I=\sum_{j=0}^{\infty}H_{2j}g_j, \tag{26}$$

with

$$g_l=\sum_{m=j}^{\infty}\left[\varepsilon_{mm(2j)}|y_m|^2+2\sum_{n=1}^{j}\varepsilon_{(n+m)(m-n)(2j)}(y_{n+m}^* y_{m-n})\right]. \tag{27}$$

It is straightforward (using binomial identities) to show that g_j vanishes for any nonzero j.

B. Truncated expansions

In order to be numerically tractable, the infinite set of equations, Eq. (20), needs to be truncated. The truncated expansion is

$$\psi^M(r)=\sum_{n=0}^{M}y_n^M(r)H_n(\xi), \tag{28}$$

where the dependence of the functions on M is explicitly indicated. What this means is that given any truncation level (the integer M) the functions y_n^M satisfy a set of $M+1$ equations (to be determined later) that represents a closure of Eq. (20). These functions are approximations of the y_n of Eq. (21) that are then added together in Eq. (28) to yield an approximation for $\psi(r)$. For example, the mean field y_0 can be studied by looking at the approximations y_0^M as a function of the truncation level M. While important, this only gives information on the closure approximation for the set of differential equations. Other quantities, e.g., the coherence function or the scintillation index, involve all available stochastic components $\boldsymbol{y_n}$. Studying these quantities as a function of the truncation level involves both closure approximation and truncating the level of the expansion. We do not look at these separately but investigate the combined effect.

Projecting out the components in Eq. (28) gives a set of $M+1$ linear equations with constant coefficients, $\dot{Y}^M=i\sigma A^M Y^M$. The solution vector is $Y^M=(y_0^M,y_1^M,\ldots,y_M^M)^T$ and the coefficient matrix is symmetric and tridiagonal

$$A^M=\begin{pmatrix}0&1&0&0&\cdots&0\\1&0&\sqrt{2}&0&\cdots&0\\0&\sqrt{2}&0&\sqrt{3}&\ddots&\vdots\\0&0&\sqrt{3}&\ddots&\ddots&0\\\vdots&\vdots&\ddots&\ddots&0&\sqrt{M}\\0&0&\cdots&0&\sqrt{M}&0\end{pmatrix}. \tag{29}$$

These are linear equations with constant coefficients, and thus solutions are of the form

$$y_n^M(r)=\sum_{m=1}^{M+1}a_m^n e^{i\sigma\lambda_m r}, \tag{30}$$

where the λ_m's are the $M+1$ eigenvalues of the coefficient matrix. Since A is symmetric, all of the eigenvalues are real. It is straightforward to show that the characteristic equation associated with A, $\det(A-\lambda I)=0$, is a function of λ^2 when M is odd or λ times a similar function when M is even. Therefore, eigenvalues occur in oppositely signed pairs except when M is even, in which case there is an additional zero eigenvalue. The y_m^M's can be written as a sum of cosine, sine terms, and a possible constant.

The average intensity in the truncated expansion is

$$\langle I^M\rangle=\langle|\psi^M(r)|^2\rangle=[Y^M]^t Y^M, \tag{31}$$

and is seen to be unity since A is symmetric,

$$\partial\langle I^M\rangle/\partial r=i\sigma[Y^M]^t(A-A^T)Y^M=0. \tag{32}$$

Thus, the truncation automatically satisfies constant intensity. The field covariance is

$$C^M(r,r')=\langle(\psi^M(r))^*\psi^M(r')\rangle=[Y^M(r')]^t Y^M(r), \tag{33}$$

and it follows that

$$\frac{\partial}{\partial r}C^M(r,r')=-\frac{\partial}{\partial r'}C^M(r,r'). \tag{34}$$

Thus, the approximate expansion for the covariance is translationally invariant.

In the truncated expansion the intensity is

$$I^M=\sum_{j=0}^{M}H_{2j}\bar{g}_j^M, \tag{35}$$

with

$$\bar{g}_j^M=\sum_{m=j}^{M}\left[\varepsilon_{mm(2j)}|y_m^M|^2+2\sum_{n=1}^{\min(j,M-m)}\varepsilon_{(n+m)(m-n)(2j)}(y_{n+m}^{M*}y_{m-n}^M)\right]. \tag{36}$$

The change in the upper limit of the n summation is required so that no y_m^M appears in the formula with $m>M$. As M gets large the y_m^M converge to the y_m quickly (this is the closure approximation), so that the $\bar{g}_j^M$ should converge to 0 for nonzero j. However, this never happens for all r, no matter how large M is (as long as it is finite). To see this, insert the exact solutions, Eq. (21), for the y_m^M's, yielding

$$\bar{g}_j^M = \sum_{m=M-j+1}^{M} \varepsilon_{mm(2j)} |y_m|^2 \approx (\sigma r)^M e^{-\sigma^2 r} F(r), \tag{37}$$

where $F(r)$ is of order 1 for large r. This function has a maximum at $r=M$ where $\bar{g}_j^M$ is very large. This problem is a result of the truncation of the series mixing up closure approximations and series approximations. However, $\bar{g}_j^M$ can be modified to a form that vanishes when the exact solutions are inserted, thus separating the errors due to series expansion and to closure. With

$$g_j^M = \sum_{m=j}^{M-j} \left[\varepsilon_{mm(2j)} |y_m^M|^2 + 2 \sum_{n=1}^{j} \varepsilon_{(n+m)(m-n)(2j)} (y_{n+m}^{M*} y_{m-n}^M) \right], \tag{38}$$

the intensity converges to the mean intensity, since the g_j^M vanish when the exact solutions for y_j's are inserted. Thus, they converge to 0 (for $j>0$) as M gets large. Note that if $j>[M/2]$, where $[\ldots]$ denotes integer part, then g_j^M vanishes. The modified scintillation index is then

$$S^M = \sum_{j=1}^{M} (g_j^M)^2. \tag{39}$$

As examples, solutions for $M=1$ are

$$y_0^1(r) = \cos(\sigma r); \quad y_1^1(r) = i \sin(\sigma r), \tag{40}$$

with the covariance

$$C^1(r-r') = \cos(\sigma(r-r')). \tag{41}$$

Defining $\lambda_+ = \sqrt{3+\sqrt{6}}$ and $\lambda_- = \sqrt{3-\sqrt{6}}$, then solutions for $M=3$ are

$$\begin{aligned}
y_0^3(r) &= \frac{1}{6}[\lambda_-^2 \cos(\sigma\lambda_+ r) + \lambda_+^2 \cos(\sigma\lambda_- r)];\\
y_1^3(r) &= \frac{i\sqrt{3}}{6}[\lambda_- \sin(\sigma\lambda_+ r) + \lambda_+ \sin(\sigma\lambda_- r)];\\
y_2^3(r) &= \frac{\sqrt{3}}{6}[\cos(\sigma\lambda_+ r) - \cos(\sigma\lambda_- r)];\\
y_3^3(r) &= i\frac{\sqrt{3}}{6}[\lambda_- \sin(\sigma\lambda_+ r) - \lambda_+ \sin(\sigma\lambda_- r)],
\end{aligned} \tag{42}$$

with the covariance $C^3(r-r')$, where

$$C^3(x) = [\lambda_-^2 \cos(\sigma\lambda_+ x) + \lambda_+^2 \cos(\sigma\lambda_- r)]/6. \tag{43}$$

The scintillation index $S^{M=3}$ can be obtained from the formulas, but the analytic expression is messy and not very illuminating.

In Fig. 1(a) the approximants y_0^1, y_0^3, and y_0^5 are compared to the exact answer y_0, while Fig. 1(b) shows the differences of the approximants from the y_0. These figures illustrate that the convergence of the closure approximations for Eq. (20) induced by the polynomial chaos expansion is quite good at small, fixed r but requires more terms in the expansion for larger values of r. Thus, any finite expansion will not converge at all ranges and the number of terms needed is expected to grow (in range). Note that the expansion is uniformly convergent over any finite interval of the range. A numerical experiment to determine the minimum value of M needed to achieve an accuracy of 10^{-8} over the whole interval $[0, r_{\max}]$ for various $r_{\max}$ gives the results in Table I. It indicates that the number of terms in the expansion roughly scales as the range (i.e., linearly) to achieve any desired accuracy.

The comparison of the approximants of the covariance to the exact result is similar to the mean field, with similar approximation errors. In Fig. 2 it can be seen that approximations with $M>10$ are needed in order to be able to propagate to ranges where the field is well decorrelated. The approximants to the scintillation index (which should be zero) are depicted in Fig. 3. At the larger propagation ranges the convergence seems to be very slow. As done with the mean field, I conduct a numerical experiment to determine the minimum value of M needed to achieve an accuracy of 10^{-8} over the whole interval $[0, r_{\max}]$ for various $r_{\max}$ with results given in Table II. Convergence for the scintillation index is much slower than for the mean intensity. The necessary approximation level scales roughly quadratically for the longer ranges.

In any use of the chaotic representation one should carefully treat the convergence of the expansion over the entire propagation range. An especially severe constraint is the scintillation index, which requires a large number of terms in the chaotic representation of the field to be included. In practice, this may not be a problem as many modeling efforts invoke the Markov approximation (i.e., fluctuations are delta correlated) and therefore propagation ranges are always larger than the medium correlation lengths (which are effectively zero). To treat this case the full equations for a general stochastic field are needed.

IV. GENERAL CASE

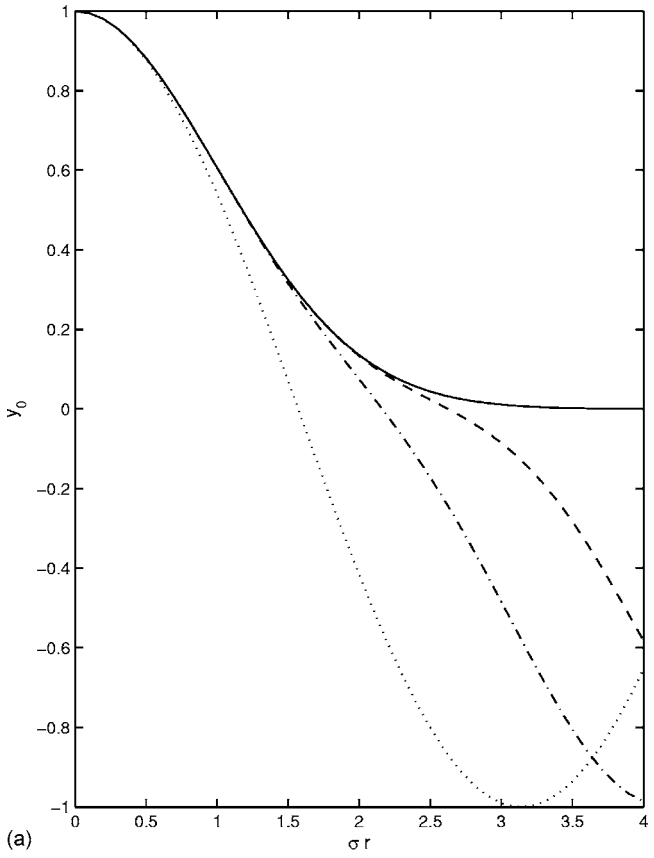
The chaotic expansion for the field $\psi(r)$, Eq. (14), and the spectral expansion for the stochastic process can be inserted into the differential equation satisfied by $\psi(r)$ giving, with the appropriate projection, the set of coupled equations

$$\begin{aligned}
\dot{y}_0 &= i\sigma \sum_k f_{-k} e^{-ikr} y_1(r;k),\\
\frac{\partial}{\partial r} y_n(r;\{k_n\}) &= \frac{i\sigma}{\sqrt{n}} \sum_{i=1}^{n} f_{-k_i} e^{-ik_i r} y_{n-1}(r;\{k_n\}_n') \\
&\quad + i\sigma\sqrt{n+1} \sum_k f_k e^{ikr} y_{n+1}(r;\{k_n\},k),
\end{aligned} \tag{44}$$

where $\{k_n\}_i'$ denotes a subset of $\{k_n\}$ with k_i deleted. The average intensity has the expansion

$$\langle \psi^*(r)\psi(r)\rangle = \sum_{m=0}^{\infty} \sum_{\{k\}_m} |y_m(r;\{k_m\})|^2 \tag{45}$$

and is constant as is easily checked. The variance of intensity suffers from the same problems as in Sec. III; the truncated expansions will not converge unless the intensity formulas are modified. There is no simple formula in this case.



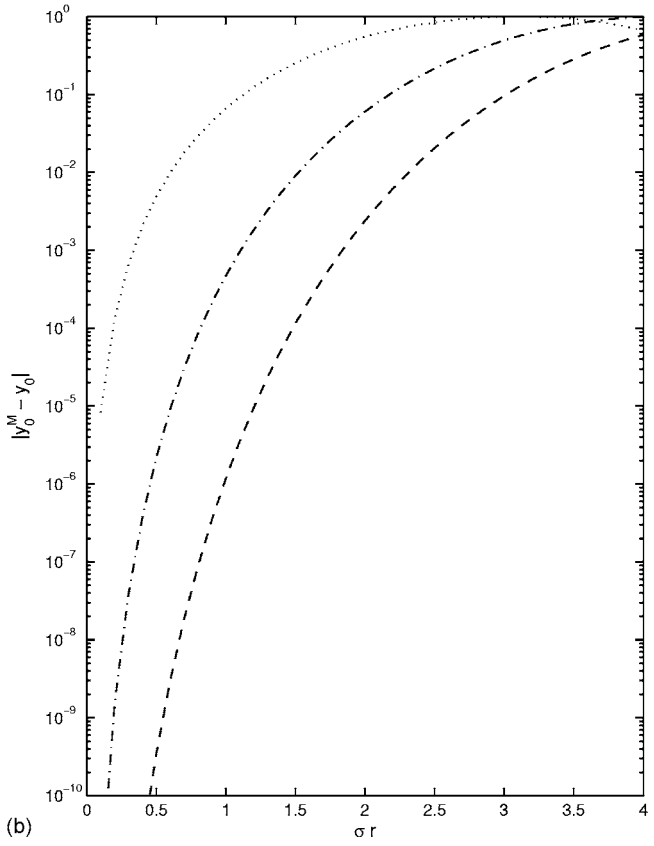


FIG. 1. (a) In limit $\sigma L \to \infty$, the exact answer, Eq. (21), for the mean field, y_0 (—), and the approximants y_0^1 (…), y_0^3 (-·-), and y_0^5 (---) from Eq. (30). (b) The absolute error in the approximants y_0^i, $i=1$ (…), 3 (-·-), and 5 (---).

TABLE I. Minimum value of M needed to achieve an accuracy of 10^{-8} over the interval $[0, r_{\max}]$ for the approximants to the mean field, Eq. (30) with $n=0$.

$\sigma r_{\max}$	0.5	1.0	2.0	4.0	6.0
M	4	6	11	25	40

Explicitly, the equations for $n=1$ and 2 are

$$\dot{y}_1(r;k) = i\sigma f_k e^{ikr} y_0(r) + i\sigma\sqrt{2}\sum_{k'} f_{-k'} e^{-ik'r} y_2(r;k,k') \tag{46}$$

and

$$\dot{y}_2(r;k_1,k_2) = i\sigma\sqrt{3}\sum_{k'} f_{-k'} e^{-ik'r} y_3(r;k_1,k_2,k') + \frac{i\sigma}{\sqrt{2}}(f_{k_1} e^{ik_1 r} y_1(r;k_2) + (k_1 \leftrightarrow k_2)). \tag{47}$$

The equations become somewhat unwieldy as n becomes large. For example, if 100 spectral components for the stochastic field are included, in the next order y_3 has over 150 000 independent components. Numerically integrating this number of equations becomes difficult. For the exponential correlation function these equations are solvable for small n. General expressions for the intensity are not easily

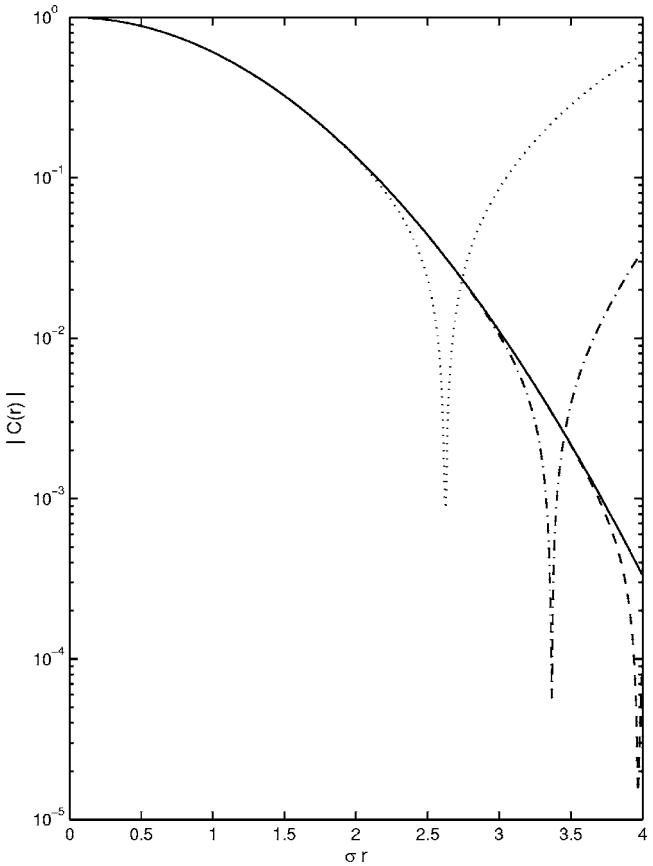


FIG. 2. In the limit $\sigma L \to \infty$, the covariance function, $C(r)$ (—), Eq. (23), and the absolute value of its approximants, Eq. (36), for $M=5$, 9, and 13: $C_5(r)$ (…), $C_9(r)$ (-·-), and $C_{13}(r)$ (---). The dips result from a sign change in the approximants.

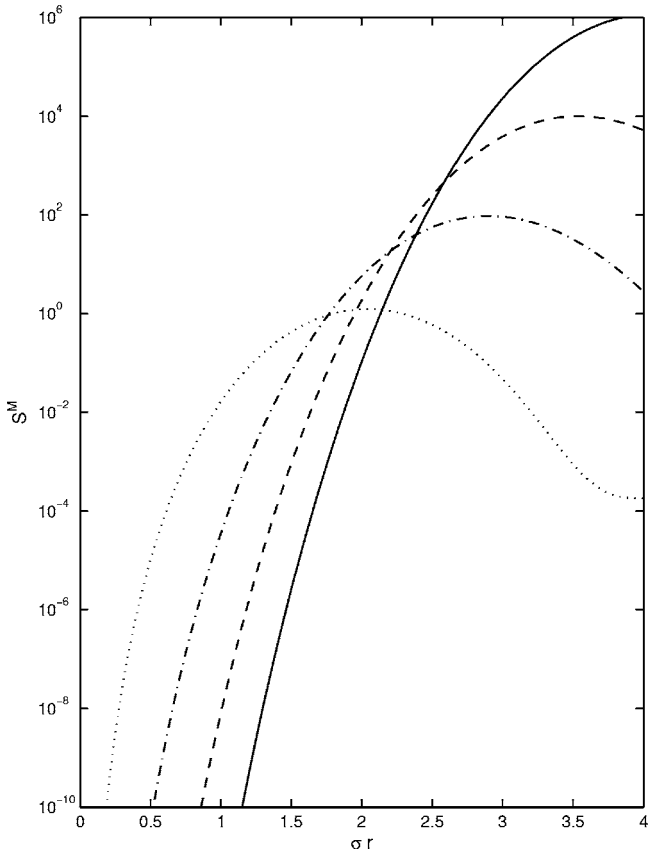


FIG. 3. Approximants for the modified scintillation index Eq. (39), in limit $\sigma L \to \infty$, for $M=3$, 7, 11, and 15: S^3 (…), S^7 (---), S^{11} (---), and S^{15} (—).

written, but are necessary for calculating the scintillation index. Low-order approximations give some feel for the form of the expressions.

General results for the different approximations are not easily derived for any correlation function. However, for $n=1$ it is quite easy to derive the equation for y_0 using the truncated form of Eq. (46) (i.e., setting y_2 to zero),

$$\begin{aligned}\dot{y}_0 &= -\sigma^2 \sum_k f_{-k} e^{-ikr} \int_0^r f_k e^{ikr'} y_0(r')dr' \\ &= -\sigma^2 \int_0^r D(r-r')y_0(r')dr', \end{aligned} \tag{48}$$

which is solvable by Laplace transform methods for the kernel G and its Laplace transform,

$$\tilde{y}_0 = \tilde{G}(s) = \frac{1}{s+\sigma^2 \tilde{D}(s)}. \tag{49}$$

The intensity is

TABLE II. Minimum value of M needed to achieve an accuracy of 10^{-8} over the interval $[0, r_{\max}]$ for the modified approximants to the scintillation index, Eq. (39).

$\sigma r_{\max}$	0.25	0.5	1.0	2.0	3.0	4.0
M	3	6	11	28	54	92

$$\begin{aligned} I &= \left| y_0 + \sum_k \xi_k y_1(r;k) \right|^2 \\ &= |y_0|^2 + \sum_{k,k'} \xi_k \xi_{k'}^* y_1^*(r;k) y_1(r;k') + 2\sum_k \mathrm{Re}[\xi_k y_0 y_1(r;k)]. \end{aligned} \tag{50}$$

The Re[…] term vanishes for similar reasons as in Sec. III, i.e., y_0 is real while y_1 is imaginary. The mean intensity is unity, as can be checked using

$$\begin{aligned} \sum_k |y_1(r;k)|^2 &= \sigma^2 \sum_k \phi(k) \left| \int_0^r dr' e^{ikr'} y_0(r') \right|^2 \\ &= \sigma^2 \int_0^r dr' \int_0^r dr'' D(r'-r'') y_0(r') y_0(r'') \\ &= 1 - y_0^2(r). \end{aligned} \tag{51}$$

The deviation of the intensity from its mean is then

$$I - \langle I \rangle = \sqrt{2} \sum_{k,k'} H_2(k,k') y_1^*(r;k) y_1(r;-k'), \tag{52}$$

giving the scintillation index as

$$\tilde{S}^1 = 2(1-|y_0|^2)^2. \tag{53}$$

This is nonzero even if y^0 were the exact mean-field result. The modification corresponding to Eq. (39) has only up to linear terms in ξ_k for the deviation from the intensity, Eq. (50), so that this modified scintillation index is identically zero.

The next order approximation keeps terms up to $n=2$, giving the equations

$$\dot{y}_0 = i\sigma \sum_k f_{-k} e^{-ikr} y_1(r;k), \tag{54}$$

$$\begin{aligned} \dot{y}_1(r;k) &= i\sigma f_k e^{ikr} y_0(r) + i\sigma\sqrt{2} \sum_{k'} f_{-k'} e^{-ik'r} y_2(r;k,k'), \\ \dot{y}_2(r;k_1,k_2) &= \frac{i\sigma}{\sqrt{2}} [f_{k_1} e^{ik_1 r} y_1(r;k_2) + (k_1 \leftrightarrow k_2)]. \end{aligned} \tag{55}$$

It is convenient to write $y_1(r;k) = i\sigma f_k z_1(r;k)$ and $y_2(r;k,k') = (i\sigma)^2 f_k f_{k'} z_2(r;k,k')/\sqrt{2}$, giving

$$\begin{aligned} \dot{y}_0 &= -\alpha^2 \sum_k \phi(k) e^{-ikr} z_1(r;k), \\ \dot{z}_1(r;k) &= e^{ikr} y_0(r) - \sigma^2 \sum_{k'} \phi(k') e^{-ik'r} z_2(r;k,k') \end{aligned} \tag{56}$$

and

$$\dot{z}_2(r;k_1,k_2) = [e^{ik_1 r} z_1(r;k_2) + (k_1 \leftrightarrow k_2)]. \tag{57}$$

Performing a Laplace transform gives the set of equations

$$s\tilde{y}_0(s) = 1 - \sigma^2 \sum_k \phi(k) \tilde{z}_1(s+ik;k), \tag{58}$$

$$s\tilde{z}_1(s;k) = \tilde{y}^0(s-ik) - \sigma^2 \sum_{k'} \phi(k')\tilde{z}_2(s+ik';k,k'), \quad (59)$$

and

$$s\tilde{z}_2(s;k,k') = \tilde{z}_1(s-ik';k) + \tilde{z}_1(s-ik;k'). \quad (60)$$

The last equation can be inserted into the second equation and, with the identity $\Sigma_k \phi(k)(s+ik)^{-1} = \tilde{D}(s)$ for the Laplace transform of the correlation function, gives

$$\tilde{G}^{-1}(s+ik)\tilde{z}_1(s+ik;k) = \tilde{y}_0(s) - \sigma^2 \times \sum_{k'} \frac{\phi(k')}{s+ik+ik'}[\tilde{z}_1(s+ik';k')]. \quad (61)$$

For a general correlation function $D(r)$ further analytical progress cannot be made; however, it can for the exponential correlation function that we consider in the next section.

The ***n=2*** expansion for the intensity gives

$$I = \left| y_0 + i\sigma \sum_k \xi_k f_k z_1(r;k) - \sigma^2/\sqrt{2} \sum_{k_1,k_2} H_2(\xi_{k_1},\xi_{k_2}) f_{k_1} f_{k_2} z_2(r;k_1,k_2) \right|^2, \quad (62)$$

so that the deviation from the mean intensity is

$$I - \langle I \rangle = \sqrt{2}\sigma^2 \sum_{k_1,k_2} H_2(\xi_{k_1},\xi_{k_2}) f_{k_1} f_{k_2} \bar{\Delta}_{12}(r) + \frac{\sigma^4\sqrt{6}}{2} \sum_{k_1,k_2,k_3,k_4} H_4(\xi_{k_1},\xi_{k_2},\xi_{k_3},\xi_{k_4}) \times f_{k_1} f_{k_2} f_{k_3} f_{k_4} \bar{\Delta}_{1234}(r), \quad (63)$$

where

$$\bar{\Delta}_{12} = \{z_1(r;k_1)z_1^*(r;-k_2) - y_0 z_2(r;k,k') + \sigma^2 \sum_q z_2(r;k_1,q) z_2^*(r;-k_2,q)\} \quad (64)$$

and

$$\bar{\Delta}_{1234} = z_2(r;k_1,k_2) z_2(r;k_3,k_4). \quad (65)$$

The modification corresponding to Eq. (39) has $\Delta_{1234}=0$ and $\Delta_{12} = \bar{\Delta}_{12}|_{\sigma=0}$, giving

$$S_{\mathrm{mod}}^2 = 2\left[\sigma^2 \sum_{k_1} \phi(k_1)|z_1(k_1)|^2\right]^2 + 4y_0^2\left[1 - y_0^2 - \sigma^2 \sum_{k_1} \phi(k_1)|z_1(k_1)|^2\right] - 4\sigma^4 y_0 \sum_{k_1,k_2} \phi(k_1)\phi(k_2) z_1(k_1) z_1(k_2) z_2^*(k_1,k_2), \quad (66)$$

where the r dependence is not explicitly noted.

V. EXAMPLE SOLUTIONS

We determine the $n=1$ and 2 approximations of the solutions for the mean field and the scintillation indices for the case of the correlation function given by Eq. (10).

A. *n*=1 approximation

The simplest approximation keeps only terms up to $n=1$, for which y_0 satisfies Eq. (49),

$$\dot{y}_0 = -\sigma^2 \int_0^r D(r-r') y_0(r') dr', \quad (67)$$

and is solved using Laplace transform methods

$$\tilde{y}_0(s) \equiv \tilde{G}(s) = \frac{s+1/L}{s(s+1/L)+\sigma^2}. \quad (68)$$

The solution is then

$$y_0(r) = e^{-(r/2L)}\left[\cosh\left(\frac{r}{2L}\sqrt{1-4\sigma^2L^2}\right) + \frac{1}{\sqrt{1-4\sigma^2L^2}} \sinh\left(\frac{r}{2L}\sqrt{1-4\sigma^2L^2}\right)\right], \quad (69)$$

while $y_1(r;k)$ is given by Eq. (60). The exact solution, Eq. (4), obeys

$$\frac{\partial}{\partial r}\langle \psi(r) \rangle = -\sigma^2\left(\int_0^r D(r-r')dr'\right)\langle \psi(r) \rangle, \quad (70)$$

which is an important difference, e.g., Eq. (70) is a simple differential equation while Eq. (67) is an integro-differential equation. Under the Markov approximation (i.e., small medium correlation lengths) the two solutions are equivalent.

The equations for the $n=1$ approximation can be numerically evaluated. Although not necessary since we already have an analytic result, doing so gives an estimate of the numerical requirements for computations using polynomial chaos. From sampling theory, the spacing between the spectral components k, denoted dk, satisfies

$$dkL < \frac{L}{r_{\max}} = \frac{(\sigma L)^2}{\sigma^2 L r_{\max}}, \quad (71)$$

in terms of the scaled range variable $\sigma^2 Lr$. For the largest values of k needed, note that effectively we have integrals of the form

$$F(k_{\max}) = \int_0^{k_{\max}} \frac{dk}{1+k^2} = F(\infty) - \frac{A}{k_{\max}}, \quad (72)$$

for A constant. The last term is the error implying that good approximations need somewhat large $k_{\max}$. It is not necessary to uniformly sample the space of spectral parameters, as fine sampling for small k and large sample spacing for the larger values of k will reduce the computational burden and achieve the same accuracy. Using convergence acceleration methods, such as Richardson extrapolation, will also reduce the computations necessary to achieve a desired accuracy. For the exponential correlation function, the minimum number of k values needed to achieve an accuracy of 10^{-3} for a propagation range of $\sigma^2 Lr=6$ is numerically estimated: less than 100 for $\sigma L=0.05$ and 1000 to 2000 for $\sigma L=0.5$. For the case of $\sigma L=5.0$ over 2000 spectral values are needed. In this latter case the ratio of the decorrelation scale for the field ($\sigma^2 L$) to the

integration scale (L) is 2.5×10^{-3}, indicating that the equations are stiff.

B. $n=2$ approximation

For the $n=2$ approximation take Eqs. (58) and (61)

$$s\tilde{y}_0(s) = 1 - \sigma^2\sum_k \phi(k)\tilde{z}_1(s+ik;k),$$
$$\tilde{G}^{-1}(s+ik)\tilde{z}_1(s+ik;k) = \tilde{y}_0(s) - \sigma^2\sum_{k'}\frac{\phi(k')}{s+ik+ik'}[\tilde{z}_1(s+ik';k')]. \tag{73}$$

For the exponential correlation function and its Laplace transform (LT), we have

$$\sum_{k'}\phi(k')\tilde{H}(s+ik') \overset{\text{ILT}}{\rightarrow} \sum_{k'}\phi(k')e^{-ikr}H(r) = e^{-r/L}H(r) \overset{\text{LT}}{\rightarrow} \tilde{H}(s+1/L), \tag{74}$$

for a function $H(r)$. Here, ILT denotes an inverse Laplace transform. A bit of algebra gives

$$\sum_{k'}\frac{\phi(k)}{s+ik+ik'}[\tilde{z}_1(s+ik';k')] = \tilde{D}(s+ik)\tilde{F}(s)\tilde{y}_0(s), \tag{75}$$

where

$$\tilde{F}(s) = \frac{\tilde{G}(s+1/L)}{(1+\sigma^2\tilde{G}(s+1/L)\tilde{D}(s+1/L))} = \frac{(s+2/L)}{(s+1/L)(s+2/L)+2\sigma^2}. \tag{76}$$

Inserting into Eq. (75) gives

$$\tilde{z}_1(s+ik;k) = \tilde{y}_0(s)\tilde{G}(s+ik)\{1-\sigma^2\tilde{D}(s+ik)\tilde{F}(s)\}, \tag{77}$$

while

$$\sum_k \phi(k)\tilde{z}_1(s+ik;k) = \tilde{F}(s)\tilde{y}_0(s). \tag{78}$$

These expressions secure the solution for the Laplace transform, $\tilde{y}_0(s)=K(s)$, where

$$\tilde{K}(s) = (s+\sigma^2\tilde{F}(s))^{-1} = \frac{(s+1/L)(s+2/L)+2\sigma^2}{s(s+1/L)(s+2/L)+\sigma^2(3s+2/L)}. \tag{79}$$

The solution is straightforwardly determined. Define

$$A \equiv \frac{1}{2}\left[\left(\frac{\sigma^2L^2}{2}+\sqrt{-d}\right)^{1/3}+\left(\frac{\sigma^2L^2}{2}-\sqrt{-d}\right)^{1/3}\right] \tag{80}$$

and

$$B \equiv \frac{i\sqrt{3}}{2}\left[\left(\frac{\sigma^2L^2}{2}+\sqrt{-d}\right)^{1/3}-\left(\frac{\sigma^2L^2}{2}-\sqrt{-d}\right)^{1/3}\right], \tag{81}$$

in terms of the discriminant, d,

$$d = \frac{(1-3(\sigma L)^2)^3}{27} - \frac{(\sigma L)^4}{4}. \tag{82}$$

The solution for $y_0(r)$ is then

$$y_0(r) = \frac{e^{-(r/L)}}{9A^2-B^2}\left[2(A(1+2A)+\sigma^2L^2)e^{(2Ar/L)}+e^{(-Ar/L)}\times\left\{(2A(4A-1)-(1-\sigma^2L^2))\cosh(Br/L) - (A(6A+1+3\sigma^2L^2)-1)\frac{\sinh(Br/L)}{B}\right\}\right]. \tag{83}$$

When $d>0$, the solution is a sum of decaying exponentials, while for $d<0$ it oscillates.

Figure 4 shows the comparison of these two approximations to the mean field to that of the exact answer for several values of σL. In Fig. 4(a) it is seen that for the short correlation lengths (small σL) the successive approximations converge quite quickly. Even the simplest approximation ($n=1$) is indistinguishable from the exact answer on the scale of the figure; what is plotted is the relative error. The scale $r=L$ corresponds to the point 2.5×10^{-3} on the x axis. Even when σL is of order unity [Fig. 4(b)], the second-order approximant already does quite well. When the correlation length is large [Fig. 4(c)], the convergence is quite slow and highly dependent on range as shown in Sec. III. The scale $r=L$ corresponds to the point 5.0 on the x axis in this figure [again note the scale of the x axis in Fig. 4(c) is different than in Figs. 4(a) and 4(b)].

The scintillation index for $n=1$, Eq. (53), is easily evaluated, while the modified version is automatically zero. For $n=2$, the modified index, Eq. (66), requires sums involving $z_1(r;k)$. Equation (78) shows that $-\sigma^2F(s)K(s)$ is the Laplace transform of $\dot{y}_0$, so that Eq. (77) gives

$$z_1(r;k) = \int_0^r dr' e^{ikr'}\{G(r-r')y_0(r')+G'(r-r')\dot{y}_0(r')\}, \tag{84}$$

where the Laplace transforms of G' is $\tilde{G}'=\tilde{G}\tilde{D}(s)$, giving

$$G'(r) = \frac{2Le^{-(r/2L)}}{\sqrt{1-4\sigma^2L^2}}\sinh\left(\frac{r}{2L}\sqrt{1-4\sigma^2L^2}\right)\Bigg]. \tag{85}$$

One summation that is needed is

$$\sigma^2\sum_k \phi(k)|z_1(r;k)|^2 = 2\sigma^2\int_0^r dr'\int_0^{r'} dr''A(r,r')D\times(r'-r'')A(r,r''), \tag{86}$$

where

$$A(r,r') = \{G(r-r')y_0(r')+G'(r-r')\dot{y}_0(r')\}. \tag{87}$$

Further analytical progress is not very illuminating. Fortunately, Eq. (86) is readily evaluated numerically. Another sum that is needed is

$$\sigma^4 \sum_{k_1,k_2} \phi(k_1)\phi(k_2) z_1(k_1) z_1(k_2) z_2^*(k_1,k_2)$$
$$= 2\sigma^4 \int_0^r dr' \sum_{k_1} \phi(k_1) z_1(r,k_1) z_1^*(r',k_1) \times \sum_{k_2} \phi(k_2) z_1(r,k_2) e^{-ik_2 r'}. \quad (88)$$

From Eq. (84)

$$\sum_k \phi(k) e^{-ikr'} z_1(r;k)$$
$$= \int_0^r dr_2 \sum_k \phi(k) e^{-ik(r'-r_2)} A(r,r_2)$$
$$= \int_0^r dr_2 A(r,r_2)[\Theta(r'-r_2) D(r'-r_2) + \Theta(r_2-r') D(r_2-r')], \quad (89)$$

and

$$\sum_k \phi(k) z_1^*(r';k) z_1(r;k)$$
$$= \int_0^r dr_2 \int_0^{r'} dr_3 \sum_k \phi(k) e^{-ik(r_3-r_2)} A(r,r_2) A(r',r_3)$$
$$= \int_0^r dr_2 \int_0^{r'} dr_3 A(r,r_2) A(r',r_3) \times [\Theta(r_3-r_2) D(r_3-2_2) + \Theta(r_2-r_3) D(r_2-r_3)], \quad (90)$$

with Θ the Heaviside function. These are not easily evaluated analytically but numerical computation of these expressions is straightforward.

Figure 5(a) shows the straightforwardly determined scintillation index for the $n=1$ approximation. It is seen that this approximation is not very good for any value of σL. The modified scintillation is identically zero as noted earlier. For the $n=2$ approximation, evaluating the above expressions gives the results shown in Fig. 5(b). For σL large the modified scintillation at this order of approximation is not very good. For moderate σL it provides a reasonable approximation. For small σL, the $n=2$ truncation provides a good approximation (within a few percent) to the scintillation index if the modified form, Eq. (66), is used. This can be understood from Eq. (86) and the answer for $\sigma^2 \Sigma_k \phi(k) |z_1(r;k)|^2$; this quantity is small over the whole propagation range (for $\sigma L = 0.05$ the maximum value is roughly 0.3). Since y_0 decays exponentially for large r and because the mean intensity is identically 1, this implies that $(\sigma^4/2) \Sigma_{k_1,k_2} \phi(k_1)\phi(k_2) |z_2(r;k_1,k_2)|^2$ asymptotes to unity as r goes to infinity. This type of term is neglected when using the modified form of the scintillation index and thus has a better chance of being small. The main conclusion to be drawn is that the modified form of the scintillation index provides a reasonable approximation to the exact scintillation index when the corresponding y_0 provides a good approximation to the mean field.

As mentioned earlier, many modeling efforts invoke the Markov approximation with medium correlation lengths much shorter than other physical scales of interest. For this case small-order approximants of the polynomial chaos expansion should give quite good results for the mean field and field covariance and provide a reasonable approximation to the scintillation index. However, in experiments where the Markov approximation is not valid, use of the polynomial chaos expansions might be problematic to implement numerically as high-order approximations may be needed.

VI. CONCLUSIONS

In stochastic propagation problems one of the more important physical parameters is the correlation length of the medium, L. The relative size of this length scale (to other length scales in the problem) often determines the qualitative behavior of the wave function; if L is much longer than the other scales (and also the propagation range), the components of the stochastic perturbations act coherently; if it is much less than the other physical scales then successive perturbations are effectively incoherent. In the toy problem of this paper, the only other length-scale parameter for the medium is $1/\sigma$, where σ is the standard deviation of the random field. As discussed in Sec. II, σL is a dimensionless parameter that can be used to characterize which propagation regime is relevant. It is the only dimensionless parameter available when we view the propagation range r as a dependent variable, and we use it to characterize solution behavior. When σL is small, propagation behaves as if the correlation length is small, e.g., the mean field follows Eq. (6) except for values of the range where it is effectively unity. Similarly, if σL is large, propagation behaves as if the correlation length is large, e.g., the mean field follows Eq. (7) except at ranges where it is negligible anyway. When σL is approximately unity description of the range behavior is done numerically.

The polynomial chaos method was implemented to calculate various statistics of the field in order to: (1) determine how well a truncated chaotic representation of the wave field approximated the different field statistics; (2) address computational concerns about convergence of the expansion as a function of the range variable; and (3) assess the computational burden necessary for numerical evaluation of the terms in the expansion. The conclusions for these different issues depend on the value of the dimensionless parameter σL.

For large σL, any expansion with a finite number of terms ceases to converge at and beyond a certain range, i.e., the longer ranges require higher-order approximants. For the mean field, the order of the approximation needed to achieve convergence roughly scales as the range, while for the scintillation index it roughly scales as the square of the range. Numerically this case is quite easy to evaluate as the coefficients of the expansion satisfy linear, constant-coefficient differential equations.

For small values of σL, lower-order approximations for the mean field work quite well (the smaller σL is, the better the approximation) over all propagation ranges. Numerically in this case, the first-order linear differential equations are quite difficult: the equations are stiff and the number of equations grows dramatically with the order of the approxima-

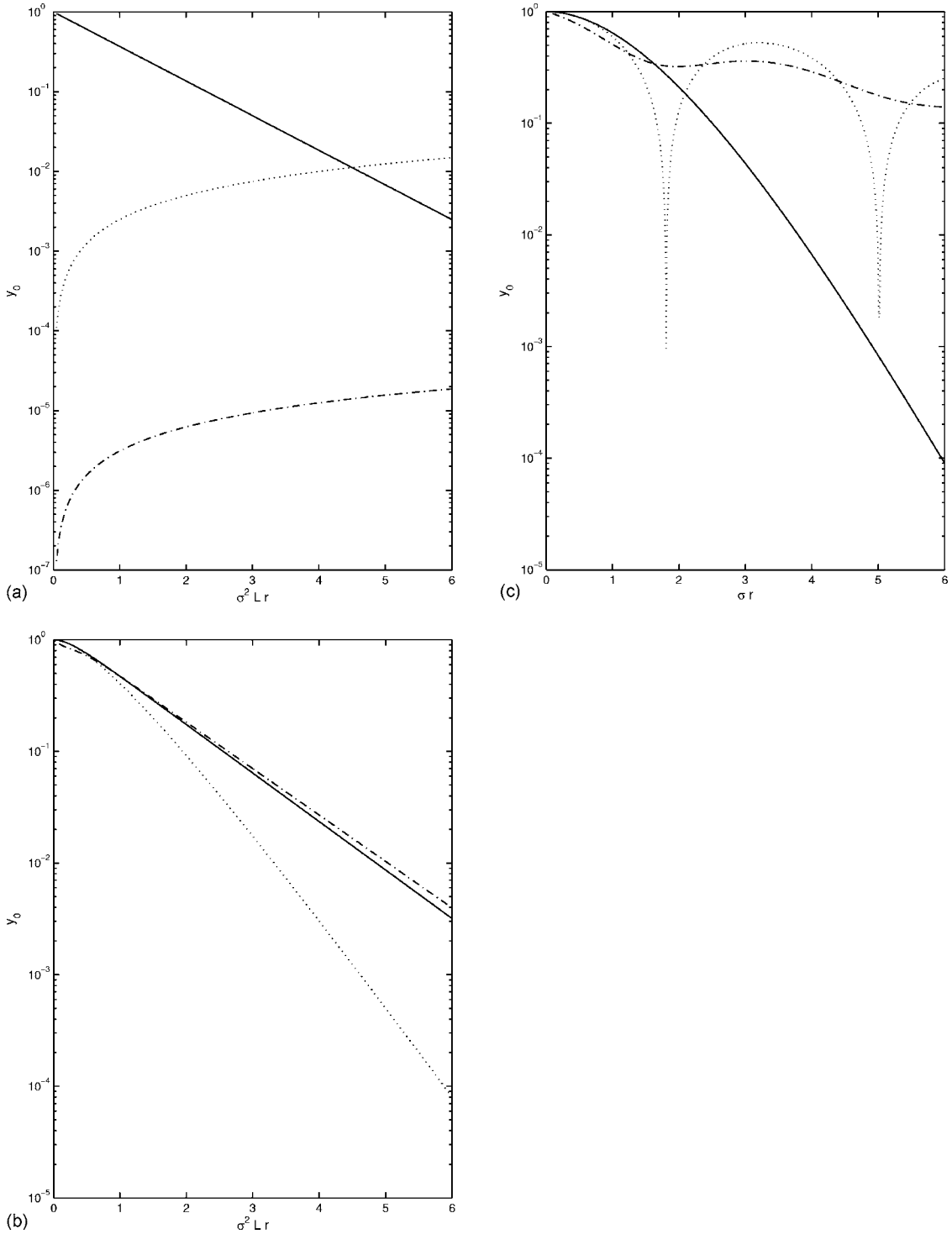


FIG. 4. (a) The exact mean field, Eq. (12), for $\sigma L=0.05$ and their relative error in the first and second-order approximants [Eqs. (69) and (83), respectively]: y_0 (—), $|y_0-y_0^1|/y_0$ (…), and $|y_0-y_0^2|/y_0$ (-·-). Note the x axis is scaled in terms of the length $(\sigma^2 L)^{-1}$ as is appropriate for small σL, cf. Eq. (6). (b) The first and second-order approximants to the mean field for $\sigma L=0.5$: y_0 (—), y_0^1 (…), and y_0^2 (-·-). The x axis is scaled in terms of the length $(\sigma^2 L)^{-1}$. (c) The first and second-order approximants to the mean field for $\sigma L=5.0$: y_0 (—), y_0^1 (…), and y_0^2 (-·-). Note the x axis is scaled in terms of the length σ^{-1} as is appropriate for large σL, cf. Eq. (7).

tion. This is because of the large number of stochastic components necessary to approximate a power-law medium (such as in the toy model): if N is the number of such components, then the Mth-order approximation has roughly N^M equations. For the toy model of this paper, the computational burden was too much for a personal computer, even for $N=2$.

This seems to give a pessimistic assessment for the utility of the method in ocean acoustics situations. However, for problems addressing the uncertainty associated with simpler stochastic contributions, the situation may be much better. For example, if only a few parameters of the stochastic contributions to propagation are relevant, or if one looks at a localized (in range) stochastic perturbation, the number of

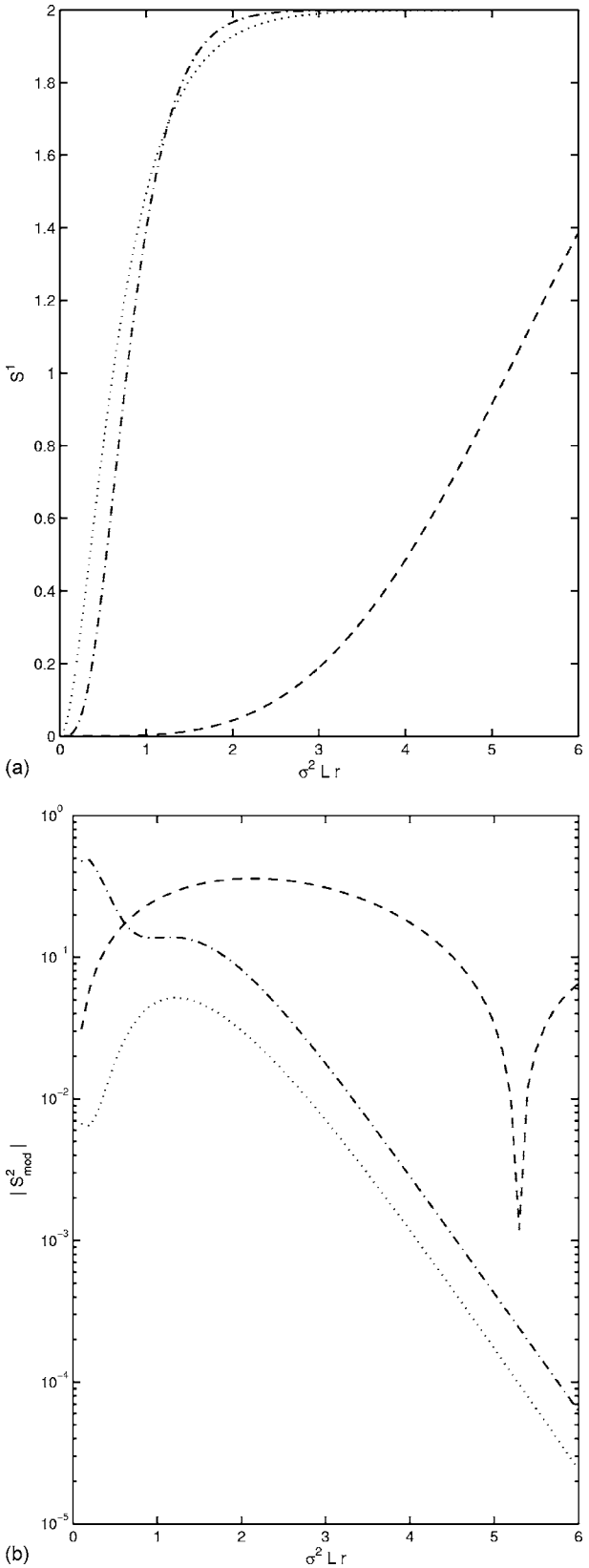


FIG. 5. (a) The first-order approximant to the scintillation index for $\sigma L = 0.05$ (…), 0.5 (---), and 5.0 (---). (b) The second-order approximant to the scintillation index for $\sigma L = 0.05$ (…), 0.5, (---), and 5.0 (---).

terms in the expansion can be significantly reduced. The problem with the toy model of this paper is that the correlation function of the stochastic variables is long-tailed, meaning that many different spectral components have to be included in order to accurately represent this correlation function.

The calculation of the intensity variance (scintillation index) using polynomial chaos is problematic. For large σL, a modified scintillation index was introduced which had the advantage of being correct when the exact answers for the coefficient functions were inserted into the corresponding expression. Even with this modification, higher-order approximants are needed than are required for the mean field. For small σL, when the mean field is well approximated using low-order expansions, the scintillation index is well approximated by using the modified scintillation index. This problem with the scintillation index may be symptomatic of the dimensionality; in one dimension it vanishes so any nonzero result of the calculation is incorrect. Still, a scintillation index near unity is much worse than much smaller values. As seen in Fig. 5(b), quite small scintillation values are obtained for small and moderate σL values over the propagation range displayed.

Many of the problems associated with using polynomial chaos probably result from the fact that it is basically a moment expansion (moments of the stochastic variable, not the wave field). In this paper, the probability distribution function of the stochastic variable is Gaussian and this moment expansion is guaranteed to converge. However, in the stochastics literature[7] it is known that use of moment expansion techniques is not as useful as desired when looking at approximations of quantities involving the probability distributions (again of the stochastic variable). There, this problem is addressed by the use of Stein's method. It would be useful to understand how this method might be incorporated into wave propagation problems.

ACKNOWLEDGMENTS

I thank Steven Finette for conversations concerning the use of polynomial chaos in propagation problems and for piquing my interest in the subject. I express my gratitude to Professor James Cole of the University of Tsukuba for his gracious hospitality during plum blossom time when most of this work was done. I also acknowledge discussions with Professor Wayne Lawton of the National University of Singapore, who made many seminal comments.

APPENDIX A: HERMITE POLYNOMIALS

The usual Hermite polynomials, $\bar{H}_m$, are associated with the weighting factor $\exp(-x^2)$ and an inner product,

$$\langle f^* g \rangle \equiv \int_{-\infty}^{\infty} f(x) g(x) e^{-x^2} dx. \tag{A1}$$

Since the Gaussian probability distribution function is of the form $\exp(-x^2/2)$, slightly different functions are needed

$$H_m(x) = \frac{2^{-m/2}}{\sqrt{m!}} \bar{H}_m(x/\sqrt{2}), \tag{A2}$$

with the normalization $\langle H_n H_m \rangle = \delta_{nm}$, where δ_{nm} is the Kronecker delta. The change in the normalization is for conve-

nience. These Hermite polynomials satisfy the three-term recursion relation

$$xH_m(x) = \sqrt{m}H_{m-1}(x) + \sqrt{m+1}H_{m+1}(x). \tag{A3}$$

For evaluation of powers of the field, the higher connection coefficients are needed. The intensity variance (fourth moment of the field) needs the coefficient $\varepsilon_{ijmn} = \langle H_i H_j H_m H_n \rangle$, for which there is no simple formula. However, an algorithm can be developed. Repeated use of the recursion relation shows that $H_m H_n$ is a combination of H_j's, where the index j is one of the integers from the set $\{m+n, m+n-2, \ldots, |m-n|\}$, i.e.,

$$H_m H_n = \sum_{\substack{j=|m-n| \\ \text{by } 2}}^{m+n} \varepsilon_{mnj} H_j, \tag{A4}$$

with

$$\varepsilon_{mnj} = \langle H_m H_n H_j \rangle = \frac{\sqrt{j!m!n!}}{\left(\frac{m+n-j}{2}\right)!\left(\frac{m+j-n}{2}\right)!\left(\frac{j+n-m}{2}\right)!} \tag{A5}$$

subject to the restrictions on m, n, and j. The expectation value in Eq. (A5) is determined by evaluating the corresponding integral. For the coefficient ε_{ijmn}, consider the pairs $H_i H_j$ and $H_m H_n$. These pairs will connect to H_l's where l is a member of the set $A=\{i+j, i+j-2, \ldots, |i-j|\}$ and $B=\{m+n, m+n-2, \ldots, |m-n|\}$, respectively. For ε_{ijmn} to be nonzero this l must be the same, i.e., it is a member of both sets A and B, which is true only if the intersection of the sets A and B is nonzero.

Our polynomials are generated from the formula

$$H_m(x) = \frac{(-1)^m}{\sqrt{m!}} e^{x^2/2} \frac{\partial^m}{\partial x^m} e^{-x^2/2}, \tag{A6}$$

so that the first few H's are

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = \frac{x^2-1}{\sqrt{2}}, \quad H_3(x) = \frac{x^3-3x}{\sqrt{6}}. \tag{A7}$$

APPENDIX B: MULTIVARIATE POLYNOMIALS

For the multivariate distributions the polynomials describing combinations of the N-independent random variables, $\{x_1, \ldots, x_N\}$ are needed. These variables satisfy $\langle x_i x_j \rangle = \delta_{ij}$. Let the set of n integers $L_n=\{l_1, \ldots, l_n\}$ be a set of not necessarily distinct integers (each chosen from $1, 2, \ldots, N$) and the set X_{L_n} be the corresponding set of variables $\{x_{l_1}, \ldots, x_{l_n}\}$. Note that if the integers are distinct then X_{L_n} is a subset of the total number of random variables; otherwise, some random variables appear multiple times in X_{L_n}. Rather than treat cases of the Hermite polynomials with repetition of some random variables differently, this notation has the advantage of treating both cases together; for a simple example see Eq. (B2) below. The multivariate Hermite polynomials are generated by

$$H_n(X_{L_n}) = \frac{(-1)^n}{\sqrt{n!}} (-1)^n e^{\mathbf{x}^T \cdot \mathbf{x}/2} \prod_{i=1}^{n} \left(\frac{\partial}{\partial x_{l_i}} \right) e^{-\mathbf{x}^T \cdot \mathbf{x}/2}, \tag{B1}$$

where $\mathbf{x}=[x_1, \ldots, x_N]^T$ is the appropriate column vector of the independent random variables. The x_i's are usually in the sums over the i's so there is no need to specify the ordering, e.g., $H_2(x_i, x_j) = H_2(x_j, x_i)$ for any integers i and j. These polynomials are symmetric in all of their arguments. An example is

$$H_2(x_i, x_j) = (x_i x_j - \delta_{ij})/\sqrt{2}. \tag{B2}$$

For independent random variables ($i \neq j$), Eq. (B2) is the scaled product of $H_1(x_i)H_1(x_j)$, while for $i=j$, it is the H_2 function of Eq. (A7).

These polynomials have the orthonormality property

$$\langle H_n(X_{L_n}) H_m(X_{K_m}) \rangle = \frac{\delta_{nm}}{n!} \sum_{P(K_m)} \prod_{i=1}^{n} \delta_{l_i k'_i}, \tag{B3}$$

where $P(K_m)$ is a permutation of the set $\{k_1, k_2, \ldots, k_m\}$ into $\{k'_1, k'_2, \ldots, k'_m\}$. All $m!$ different permutations are summed over. Intuitively, this formula represents the ideas that the patterns, including the multiplicities, of the integers in the sets L_n and K_m should correspond and that all such correspondences need to be included. The recursion relations for these polynomials are

$$x_{l_{m+1}} H_m(X_{L_m}) = \sqrt{m+1} H_{m+1}(X_{L_{m+1}}) + \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \delta_{l_i l_{m+1}} H_{m-1}(X_{L'_m}), \tag{B4}$$

where L'_m denotes the set obtained by deleting the index l_i from L_m. After $H_2(x_i, x_j)$ the next few polynomials are

$$H_3(x_i, x_j, x_k) = \frac{1}{\sqrt{6}} \{x_i x_j x_k - S[x_i \delta_{jk}]\}, \tag{B5}$$

and

$$H_4(x_i, x_j, x_k, x_l) = \frac{1}{\sqrt{6}} \{x_i x_j x_k x_l - S[x_i x_j \delta_{kl}] + S[\delta_{ij} \delta_{kl}]\}, \tag{B6}$$

where $S[\ldots]$ is an operator that forms a symmetric combination of the indices i, j, etc. in its argument, e.g.,

$$S[x_1 y_2 y_3] = x_1 y_2 y_3 + x_2 y_1 y_3 + x_3 y_1 y_2. \tag{B7}$$

[1] S. M. Flatté, R. Dashen, W. H. Munk, K. M. Watson, and F. Zachariasen, *Sound Transmission Through a Fluctuating Ocean* (Cambridge University Press, New York, 1979).

[2] D. B. Creamer and S. M. Flatté, unpublished (1984).

[3] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).

[4] L. B. Dozier and F. D. Tappert, "Statistics of normal-mode amplitudes in a random ocean. II. Computations," J. Acoust. Soc. Am. **64**, 533–547 (1978).

[5] D. B. Creamer, "Theoretical analysis of scintillating lossy waveguides," in

Shallow Water Acoustics, Proceedings of SWAC'97, edited by R. Zhang (1997), pp. 241–246.
[6]S. Finette, "Embedding uncertainty into ocean acoustic propagation models," J. Acoust. Soc. Am. **117**, 997–1000 (2005).
[7]W. Schoutens, *Stochastic Processes and Orthogonal Polynomials*, Lecture Notes in Statistics 146 (Springer, New York, 2000).
[8]J. L. Codona, D. B. Creamer, S. M. Flatté, R. G. Frehlich, and F. S. Henyey, "Moment-equation and path-integral techniques for wave propagation in random media," J. Math. Phys. **27**, 171–177 (1986).
[9]R. Ghanem and P. Spanos, *Stochastic Finite Elements: A Spectral Approach* (Springer, Berlin, 1991).
[10]S. Finette, "Uncertainty propagation in ocean acoustic waveguides," J. Acoust. Soc. Am. **117**, 2461 (2005).
[11]S. Finette, "A dynamical formulation for environmental uncertainty for acoustic propagation in ocean waveguides," J. Acoust. Soc. Am. (submitted).

Wave propagation in two-dimensional periodic lattices

A. Srikantha Phani, J. Woodhouse, and N. A. Fleck[a)]
Department of Engineering, Cambridge University, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

(Received 4 December 2005; revised 2 February 2006; accepted 3 February 2006)

Plane wave propagation in infinite two-dimensional periodic lattices is investigated using Floquet-Bloch principles. Frequency bandgaps and spatial filtering phenomena are examined in four representative planar lattice topologies: hexagonal honeycomb, Kagomé lattice, triangular honeycomb, and the square honeycomb. These topologies exhibit dramatic differences in their long-wavelength deformation properties. Long-wavelength asymptotes to the dispersion curves based on homogenization theory are in good agreement with the numerical results for each of the four lattices. The slenderness ratio of the constituent beams of the lattice (or relative density) has a significant influence on the band structure. The techniques developed in this work can be used to design lattices with a desired band structure. The observed spatial filtering effects due to anisotropy at high frequencies (short wavelengths) of wave propagation are consistent with the lattice symmetries. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2179748]

PACS number(s): 43.20.Bi, 43.20.Mv, 43.40.At, 43.40.Dx [RLW] Pages: 1995–2005

I. INTRODUCTION

Lattice materials are reticulated, cellular structures obtained by tessellating a unit cell comprising a few beams or bars (Gibson and Ashby, 1997). Engineering applications include sandwich beams, panels (Wicks and Hutchinson, 2001) and space trusses (Noor *et al.*, 1978). The geometric periodicity of these materials governs their static and dynamic responses. The effective elastic properties of two-dimensional lattice materials and their dependence on relative density are well documented (Christensen, 2000; Gibson and Ashby, 1997; Torquato *et al.*, 1998). Recent studies have also addressed the static elastic buckling and yielding phenomena of lattice materials (Hutchinson, 2004). Although considerable literature exists on their static properties, comparatively little is known about their wave propagation behavior.

The wave-bearing properties of a lattice are governed by its geometry. It has been observed by both physicists (e.g., Brillouin 1953; Kittel, 1962) and structural engineers (e.g., Langley *et al.*, 1997; Mead, 1973, 1996) that wave motion in periodic structures, such as lattice materials, exhibits characteristic pass and stop bands. These pass bands and stop bands are frequency intervals over which wave motion can or cannot occur, respectively. There may be an opportunity to tailor lattice materials to achieve desired band gap characteristics, such that wave propagation is prevented in the specified frequency regimes. This behavior has direct application to photonic devices, and a systematic approach based on topology optimization procedures has been used to design photonic crystals with optimized bandgap properties (Sigmund and Jensen, 2003). They considered a square arrangement of inclusions of arbitrary shape, and they optimized the width of bandgap by varying the proportion of inclusion to the matrix, with the modulus and density of each phase held fixed. They observed the widest bandgap for the high contrast case, but limited their study to structures with square symmetry. Wave beaming also occurs in periodic structures: the direction of wave propagation at any given frequency is restricted to preferred directions (Langley *et al.*, 1997; Ruzzene *et al.*, 2003). Consequently, lattice materials behave as *frequency and spatial filters*.

Our objective in the present study is to explore wave propagation phenomena, such as bandgaps and wave directionality, in three regular honeycombs—hexagonal, square and triangular and in the semiregular Kagomé lattice. In all four geometries, the constituent beams are of uniform length L and depth d, and two-dimensional (2D) prismatic topologies of unit thickness into the page are considered. The hexagonal and triangular honeycombs and the Kagomé lattice have isotropic in-plane effective properties, while the square honeycomb is strongly anisotropic (Gibson and Ashby, 1997; Torquato *et al.*, 1998). The effective elastic properties of these microstructures are summarized in Table I. It is anticipated that the results based upon effective medium theory should agree with the asymptotes to the dispersion curves in the long-wavelength limit at zero frequency. Clearly, the four lattices have different long-wavelength deformation limits. It is of particular interest to know whether the finite frequency short-wavelength deformation behavior of the lattices show similar differences. To explore this, dispersion curves of the four lattices are computed using Floquet-Bloch principles. The Floquet-Bloch principles used in the present study have been widely employed in a number of research fields: in solid-state physics to investigate wave propagation in crystal structures, Bragg gratings and in photonic crystals (Brillouin, 1953; Kittel, 1962), and in mechanical systems such as stiffened panels to study their in-plane and out-of-plane vibration behavior (Mead, 1973, 1996).

The present study is structured as follows. Floquet-Bloch's principles are summarized in Sec. II in the context of lattice materials. The unit cell of each lattice is modeled as a network of beams using the finite element method, as de-

a)Electronic mail: naf1@eng.cam.ac.uk

TABLE I. Effective properties of the four lattice topologies: E=Young's modulus of the solid material, K^* and G^* are the effective bulk modulus and shear modulus of the cellular solid, respectively. $\bar{\rho} \equiv \rho^*/\rho$ is relative density of the cellular solid, where ρ and ρ^* are the density of the solid material and lattice material, respectively. d and L are the thickness and length of the cell walls, respectively. λ is the slenderness ratio defined as $\lambda = 2\sqrt{3}L/d$.

Topology	Relative density $\bar{\rho}$	Relative bulk modulus $\bar{K} = \frac{K^*}{E}$	Relative shear modulus $\bar{G} = \frac{G^*}{E}$	Poisson's ratio ν^*	Isotropic
Triangular honeycomb	$2\sqrt{3}\left(\frac{d}{L}\right) = \frac{12}{\lambda}$	$\frac{1}{4}\bar{\rho}$	$\frac{1}{8}\bar{\rho}$	$\frac{1}{3}$	Yes
Hexagonal honeycomb	$\frac{2}{\sqrt{3}}\left(\frac{d}{L}\right) = \frac{4}{\lambda}$	$\frac{1}{4}\bar{\rho}$	$\frac{3}{8}\bar{\rho}^3$	1	Yes
Kagomé lattice	$\sqrt{3}\left(\frac{d}{L}\right) = \frac{6}{\lambda}$	$\frac{1}{4}\bar{\rho}$	$\frac{1}{8}\bar{\rho}$	$\frac{6-\bar{\rho}^2}{18+\bar{\rho}^2} \approx \frac{1}{3}$	Yes
Square honeycomb	$2\left(\frac{d}{L}\right) = \frac{4\sqrt{3}}{\lambda}$	$\frac{1}{4}\bar{\rho}$	$\frac{1}{16}\bar{\rho}^3$	0	No

scribed in Sec. III and the governing equations of motion are derived. The propagation of free harmonic waves is analysed in Sec. IV by applying Bloch's theorem to the governing equations of motion of the unit cell. A linear algebraic eigenvalue problem is formulated that involves the two components of the wave vector, called "phase constants," and the frequency of the plane wave. Dispersion surfaces are constructed by specifying the two phase constants and solving for the frequency. The dispersion surfaces for the four lattices and their long-wavelength asymptotes are discussed in Sec. V. Finally, the spatial filtering of waves due to anisotropic effects at high frequencies (corresponding to deformations of short wavelength) are considered with reference to the symmetries of the parent lattice.

II. BLOCH'S THEOREM

Before proceeding to Bloch's theorem, it is worth reviewing relevant concepts from solid state physics. The joints of any lattice structure, such as the hexagonal honeycomb, can be envisioned as a collection of points, called lattice points, and these are associated with a set of basis vectors. The lattice point system together with the basis is usually referred to as a direct lattice.

Upon selecting a suitable unit cell, the entire direct lattice can be obtained by tesselating the unit cell along the basis vectors $\mathbf{e}_i$. Denote the lattice points in a unit cell by $\mathbf{r}_j$: these will correspond to a subset of the nodes of the finite element model of the unit cell. Let $q(\mathbf{r}_j)$ denote the displacement of a lattice point in the reference unit cell. If a plane wave solution is admitted, then $q(\mathbf{r}_j)$ is of the form

$$q(\mathbf{r}_j) = q_j e^{(i\omega t - \mathbf{k}\cdot\mathbf{r}_j)}, \tag{1}$$

where q_j is the amplitude, ω is frequency in rad/s, and $\mathbf{k}$ is the wave vector of the plane wave. With reference to the chosen unit cell, let the integer pair (n_1, n_2) identify any other cell obtained by n_1 translations along the $\mathbf{e}_1$ direction and n_2 translations along the $\mathbf{e}_2$ direction. The point in the cell (n_1, n_2), corresponding to the jth point in the reference

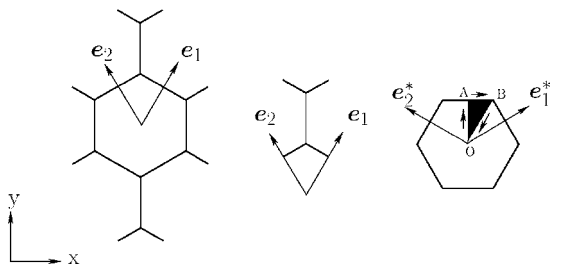


FIG. 1. Hexagonal honeycomb (left) with a selected primitive unit cell (middle) and the first Brillouin zone (right) in k space. The basis vectors of the direct lattice ($\mathbf{e}_i$) and the reciprocal lattice ($\mathbf{e}_i^*$) are also shown. The points O, A, and B are as defined in Table III.

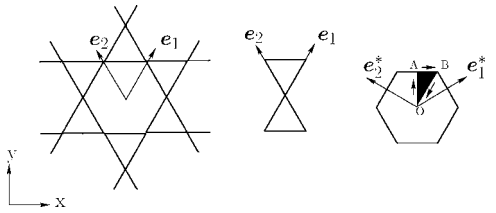


FIG. 2. Kagomé lattice (left) with selected primitive unit cell (middle) and the first Brillouin zone (right) in k space. The basis vectors of the direct lattice ($\mathbf{e}_i$) and the reciprocal lattice ($\mathbf{e}_i^*$) are also shown. The points O, A and B are as defined in Table III.

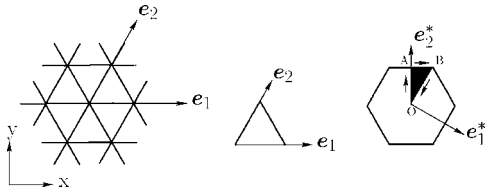


FIG. 3. Triangular honeycomb (left) with selected primitive unit cell (middle) and the first Brillouin zone (right) in k space. The basis vectors of the direct lattice ($\mathbf{e}_i$) and the reciprocal lattice ($\mathbf{e}_i^*$) are also shown. The points O, A, and B are as defined in Table III.

unit cell, is denoted by the vector $\mathbf{r}=\mathbf{r}_j+n_1\mathbf{e}_1+n_2\mathbf{e}_2$. According to Bloch's theorem, the displacement at the jth point in any cell identified by the integer pair (n_1,n_2) in the direct lattice basis is given by

$$q(\mathbf{r}) = q(\mathbf{r}_j)e^{\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}_j)} = q(\mathbf{r}_j)e^{(k_1n_1+k_2n_2)}. \tag{2}$$

Here, $k_1=\delta_1+i\epsilon_1$ and $k_2=\delta_2+i\epsilon_2$ represent the components of the wave vector $\mathbf{k}$ along the $\mathbf{e}_1$ and $\mathbf{e}_2$ vectors, that is $k_1=\mathbf{k}\cdot\mathbf{e}_1$ and $k_2=\mathbf{k}\cdot\mathbf{e}_2$. The real part δ and the imaginary part ϵ are called the *attenuation* and *phase* constants, respectively. The real part is a measure of the attenuation of a wave as it progresses from one unit cell to the next. For waves propagating without attenuation, the real part is zero and the components of the wave vector reduce to $k_1=i\epsilon_1$ and $k_2=i\epsilon_2$. The imaginary part or the phase constant is a measure of the phase change across one unit cell.

In simple terms, Bloch's theorem (or Floquet's principle in the case of one-dimensional periodic structures) states that for any structure with repetitive identical units, the change in complex wave amplitude across a unit cell, due to a propagating wave without attenuation, does not depend upon the location of the unit cell within the structure. By virtue of this theorem, one can understand wave propagation through the *entire* lattice by considering wave motion within a *single* unit cell. Bloch's theorem thus leads to enormous savings in the analysis of wave propagation in periodic structures.

It is convenient to define a reciprocal lattice in the wave vector space (k space), such that the basis vectors of the direct and reciprocal lattice satisfy

$$\mathbf{e}_i \cdot \mathbf{e}_j^* = \delta_{ij}, \tag{3}$$

where $\mathbf{e}_i$ denote the basis vectors of the direct lattice and $\mathbf{e}_j^*$ denote the basis of reciprocal lattice, δ_{ij} is the Kronecker delta function. For a two-dimensional lattice, the subscripts i and j take the integer values 1 and 2.

The wave vectors can be expressed in terms of the reciprocal lattice basis $\mathbf{e}_i^*$. Since the reciprocal lattice is also periodic, one can restrict the wave vectors to certain regions in the reciprocal lattice called Brillouin Zones (Brillouin, 1953). For computational efficiency, the wave vectors may be restricted to the edges of the irreducible part of the first Brillouin zone to explore bandgaps, since the band extrema almost always occur along the boundaries of the irreducible zone (Kittel, 1962). The first Brillouin zone is defined as a Wigner-Seitz or primitive unit cell of the reciprocal lattice, and it can be constructed as follows.

(1) Select any lattice point in the reciprocal lattice as the origin and connect it to neighboring points.
(2) Construct the perpendicular bisectors of these lines. The region of intersection is the first Brillouin zone.

The four lattices of interest in the present study are shown in Figs. 1–4 along with a choice of primitive unit cell, direct lattice translation vectors, and first Brillouin zone. The wave vectors are restricted to the edges of the irreducible first Brillouin Zone shown by the shaded region OAB. The basis vectors for the direct lattice and reciprocal lattice are tabulated in Table II and the Brillouin zone points are tabulated in Table III for each lattice.

TABLE II. Primitive translation vectors of the four lattices: L denotes the length of each beam of the lattice; $\mathbf{i}$ and $\mathbf{j}$ are the Cartesian unit vectors in the $x-y$ plane. Note that $\mathbf{e}_i\cdot\mathbf{e}_j^*=\delta_{ij}$.

Topology	Direct lattice	Reciprocal lattice
Hexagonal honeycomb	$\mathbf{e}_1=\sqrt{3}L\left(\frac{1}{2}\mathbf{i}+\frac{\sqrt{3}}{2}\mathbf{j}\right)$	$\mathbf{e}_1^*=\frac{1}{\sqrt{3}L}\left(\mathbf{i}+\frac{1}{\sqrt{3}}\mathbf{j}\right)$
	$\mathbf{e}_2=\sqrt{3}L\left(-\frac{1}{2}\mathbf{i}+\frac{\sqrt{3}}{2}\mathbf{j}\right)$	$\mathbf{e}_2^*=\frac{1}{\sqrt{3}L}\left(-\mathbf{i}+\frac{1}{\sqrt{3}}\mathbf{j}\right)$
Kagomé lattice	$\mathbf{e}_1=2L\left(\frac{1}{2}\mathbf{i}+\frac{\sqrt{3}}{2}\mathbf{j}\right)$	$\mathbf{e}_1^*=\frac{1}{2L}\left(\mathbf{i}+\frac{1}{\sqrt{3}}\mathbf{j}\right)$
	$\mathbf{e}_2=2L\left(-\frac{1}{2}\mathbf{i}+\frac{\sqrt{3}}{2}\mathbf{j}\right)$	$\mathbf{e}_2^*=\frac{1}{2L}\left(-\mathbf{i}+\frac{1}{\sqrt{3}}\mathbf{j}\right)$
Triangular honeycomb	$\mathbf{e}_1=L\mathbf{i}$	$\mathbf{e}_1^*=\frac{1}{L}\left(\mathbf{i}-\frac{1}{\sqrt{3}}\mathbf{j}\right)$
	$\mathbf{e}_2=L\left(\frac{1}{2}\mathbf{i}+\frac{\sqrt{3}}{2}\mathbf{j}\right)$	$\mathbf{e}_2^*=\frac{1}{L}\left(\frac{2}{\sqrt{3}}\mathbf{j}\right)$
Square honeycomb	$\mathbf{e}_1=L\mathbf{i}$	$\mathbf{e}_1^*=\frac{1}{L}\mathbf{i}$
	$\mathbf{e}_2=L\mathbf{j}$	$\mathbf{e}_2^*=\frac{1}{L}\mathbf{j}$

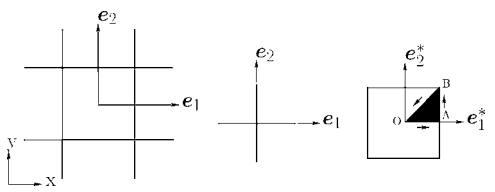


FIG. 4. Square honeycomb (left) with selected primitive unit cell (middle) and the first Brillouin zone (right) in k space. The basis vectors of the direct lattice ($\mathbf{e}_i$) and the reciprocal lattice ($\mathbf{e}_i^*$) are also shown. The points O, A, and B are as defined in Table III.

TABLE III. The irreducible first Brillouin zone points of the four lattices. The wave vectors are chosen along the locus OAB; i.e.; along the edges of the irreducible part of the first Brillouin zone.

Topology	Cartesian basis	Reciprocal basis
Hexagonal honeycomb	$O=(0,0);\quad A=\frac{1}{L}\left(0,\frac{1}{3}\right);\quad B=\frac{1}{L}\left(\frac{1}{3\sqrt{3}},\frac{1}{3}\right)$	$O=(0,0);\quad A=\left(\frac{1}{2},\frac{1}{2}\right);\quad B=\left(\frac{2}{3},\frac{1}{3}\right)$
Kagomé lattice	$O=(0,0);\ A=\frac{1}{L}\left(0,\frac{1}{3}\right);\quad B=\frac{1}{L}\left(\frac{1}{3\sqrt{3}},\frac{1}{3}\right)$	$O=(0,0);\quad A=\left(\frac{1}{2},\frac{1}{2}\right);\quad B=\left(\frac{2}{3},\frac{1}{3}\right)$
Triangular honeycomb	$O=(0,0);\quad A=\frac{1}{L}\left(0,\frac{1}{\sqrt{3}}\right);\quad B=\frac{1}{L}\left(\frac{1}{3},\frac{1}{\sqrt{3}}\right)$	$O=(0,0);\ A=\left(0,\frac{1}{2}\right);\quad B=\left(\frac{1}{3},\frac{2}{3}\right)$
Square honeycomb	$O=(0,0);\quad A=\frac{1}{L}\left(0,\frac{1}{2}\right);\quad B=\frac{1}{L}\left(\frac{1}{2},\frac{1}{2}\right)$	$O=(0,0);\quad A=\left(0,\frac{1}{2}\right);\quad B=\left(\frac{1}{2},\frac{1}{2}\right)$

III. FINITE ELEMENT MODELING OF THE UNIT CELL

Each lattice is considered to be a rigid-jointed network of beams with no prestress. The unit cell is discretized into a network of Timoshenko beams. Each beam is assumed to have three degrees of freedom at each end: two translations (u,v) in the (x,y) plane and a rotation θ_z about the z axis. The continuous variation of these displacements within a typical beam element, shown in Fig. 5, is approximated by

$$u(x,t)=\sum_{r=1}^{6} a_r(x)q_r(t),\quad v(x,t)=\sum_{r=1}^{6} b_r(x)q_r(t),$$
$$\theta_z(x,t)=\sum_{r=1}^{6} c_r(x)q_r(t), \tag{4}$$

where x is measured along the axis of the beam, and the degrees of freedom q_r consist of the six nodal degrees of freedom $(u_1,v_1,\theta_{z1},u_2,v_2,\theta_{z2})$. The shape functions a_r, b_r, and c_r, $(r=1\cdots6)$ for the six nodal displacements are given in Appendix A.

The kinetic and potential energies per unit thickness of the beam into the page are given by:

$$T=\frac{1}{2}\int_{-L/2}^{L/2}\rho d\dot{u}^2dx+\frac{1}{2}\int_{-L/2}^{L/2}\rho d\dot{v}^2dx+\frac{1}{2}\int_{-L/2}^{L/2}\rho I_z\dot{\theta}_z^2dx$$

$$U=\frac{1}{2}\int_{-L/2}^{L/2}Ed\left(\frac{du}{dx}\right)^2dx+\frac{1}{2}\int_{-L/2}^{L/2}EI_z\left(\frac{d\theta_z}{dx}\right)^2dx+\frac{1}{2}\int_{-L/2}^{L/2}\kappa dG\left(\frac{dv}{dx}-(\theta_z)\right)^2dx, \tag{5}$$

where ρ is the density of the material used to make the lattice. L and I_z denote the length and second moment of area of the beam, respectively; and κ denotes the shear correction factor used in Timoshenko beam theory (Weaver and Jonhston, 1987). Substitute Eq. (4) into Eq. (5) in order to obtain

$$T=\frac{1}{2}\sum_{r=1}^{6}\sum_{s=1}^{6}\dot{q}_r\dot{q}_s\int_{-L/2}^{L/2}(\rho da_ra_s+\rho db_rb_s+\rho I_zc_rc_s)dx,$$
$$U=\frac{1}{2}\sum_{r=1}^{6}\sum_{s=1}^{6}q_rq_s\int_{-L/2}^{L/2}[Eda_r'a_s'+EI_zb_r''b_s''+\kappa Gd(b_r'-c_r)\times(b_s'-c_s)]dx, \tag{6}$$

where the primes denote differentiation with respect to the axial coordinate x. The equations of motion are obtained by applying Hamilton's variational principle,

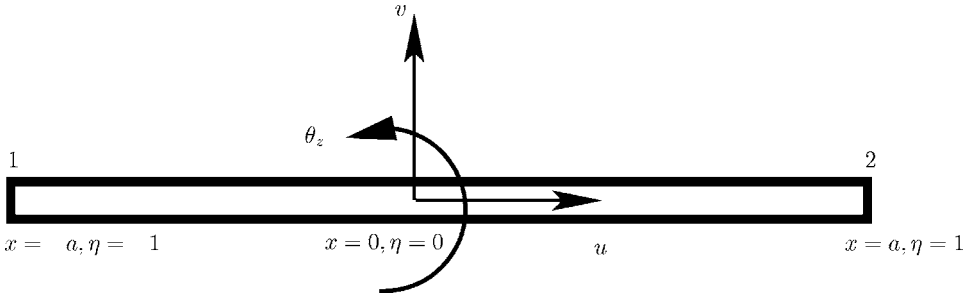


FIG. 5. Beam element with nodes numbered 1 and 2. The three nodal degrees of freedom are shown together with the local element coordinate axes with origin located at the middle of the beam. The nondimensional coordinate is $\eta=x/a=2x/L$, where L is the length of the beam.

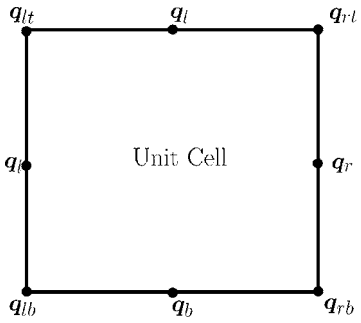


FIG. 6. A generic unit cell for a two-dimensional periodic structure, showing the degrees of freedom shared with the neighboring unit cells.

$$\delta \int \mathcal{L} dt = 0, \quad \mathcal{L} = T - U + W_e, \tag{7}$$

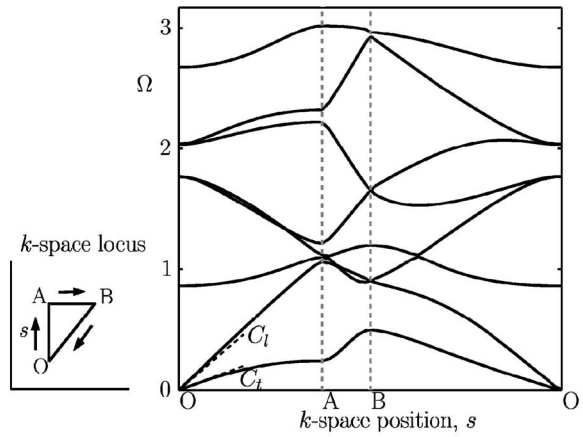
where δ denotes the first variation and W_e denotes the work done by the external forces. Upon evaluating the first variation in Eq. (7), the Euler-Lagrangian equations of motion for the dynamics of the beam element are obtained as

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_r}\right) - \frac{\partial \mathcal{L}}{\partial q_r} = f_r, \tag{8}$$

where $\mathcal{L}$ is the Lagrangian of the dynamical system as defined in Eq. (7) and f_r is the force corresponding to the degree of freedom q_r. The above equations of motion can be written for each beam element and the assembled equation of the motion for the unit cell take the form

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f}, \tag{9}$$

where the matrices $\mathbf{M}$, $\mathbf{K}$ denote the assembled global mass and stiffness matrix of the unit cell, respectively. The vectors $\mathbf{q}$, $\ddot{\mathbf{q}}$, and $\mathbf{f}$, respectively, denote the nodal displacements, accelerations, and forces. For any jth node the nodal displacement vector is given by $\mathbf{q}_j = [u_j \; v_j \; \theta_{zj}]^T$.



k-space position	Branch 1	Branch 2	Branch 3	Branch 4
O				
A				
B				

FIG. 7. Band structure of a hexagonal honeycomb with slenderness ratio equal to 10. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.

IV. ANALYSIS OF FREE WAVE MOTION

Having formulated the equations of motion of a unit cell, the propagation of planar harmonic waves at a radial frequency ω within the entire lattice can be investigated by invoking Bloch's theorem. The equations of motion in Eq. (9) follow as

$$[-\omega^2\mathbf{M} + \mathbf{K}]\mathbf{q} = \mathbf{f} \text{ or } \mathbf{D}\mathbf{q} = \mathbf{f}, \quad \mathbf{D} = [-\omega^2\mathbf{M} + \mathbf{K}], \tag{10}$$

where the *dynamic* stiffness $\mathbf{D}$ reduces to the static stiffness at zero frequency.

By virtue of Bloch's theorem, the following relationships between the displacements, $\mathbf{q}$, and forces, $\mathbf{f}$, are obtained:

$$\begin{aligned}
&\mathbf{q}_r = e^{k_1}\mathbf{q}_l, \quad \mathbf{q}_t = e^{k_2}\mathbf{q}_b, \\
&\mathbf{q}_{rb} = e^{k_1}\mathbf{q}_{lb}, \quad \mathbf{q}_{rt} = e^{k_1+k_2}\mathbf{q}_{lb}, \quad \mathbf{q}_{lt} = e^{k_2}\mathbf{q}_{lb} \\
&\mathbf{f}_r = -e^{k_1}\mathbf{f}_l, \quad \mathbf{f}_t = -e^{k_2}\mathbf{f}_b, \\
&\mathbf{f}_{rt} + e^{k_1}\mathbf{f}_{lt} + e^{k_2}\mathbf{f}_{rb} + e^{k_1+k_2}\mathbf{f}_{lb} = 0,
\end{aligned} \tag{11}$$

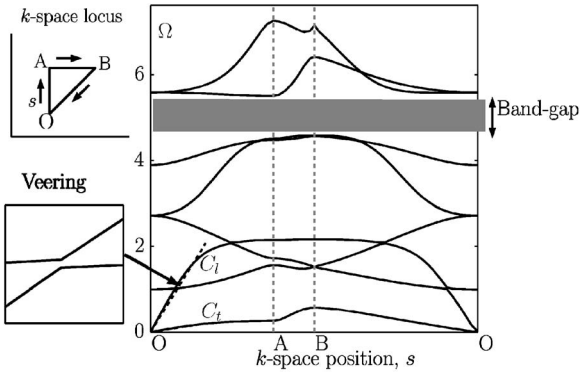
where the subscripts l, r, b, t, and i, respectively, denote the displacements corresponding to the left, right, bottom, top, and internal nodes of a generic unit cell, as shown in Fig. 6. The displacements of the corner nodes are denoted by double subscripts: for example, lb denotes the left bottom corner.

Using the above relationships one can define the following transformation:

$$\mathbf{q} = \mathbf{T}\tilde{\mathbf{q}}, \tag{12}$$

$$\mathbf{T} = \begin{bmatrix}
\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{I}e^{k_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{I}e^{k_2} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}e^{k_1} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}e^{k_2} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}e^{(k_1+k_2)} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}
\end{bmatrix}, \quad \tilde{\mathbf{q}} = \begin{bmatrix} \mathbf{q}_l \\ \mathbf{q}_b \\ \mathbf{q}_{lb} \\ \mathbf{q}_i \end{bmatrix},$$

where $\tilde{\mathbf{q}}$ denote the displacements of the nodes in the Bloch reduced coordinates. Now, substitute the transformation given by Eq. (12) into the governing equations of motion in Eq. (10) and premultiply the resulting equation with $\mathbf{T}^{\mathrm{H}}$ to enforce force equilibrium (Langley, 1993). One obtains the



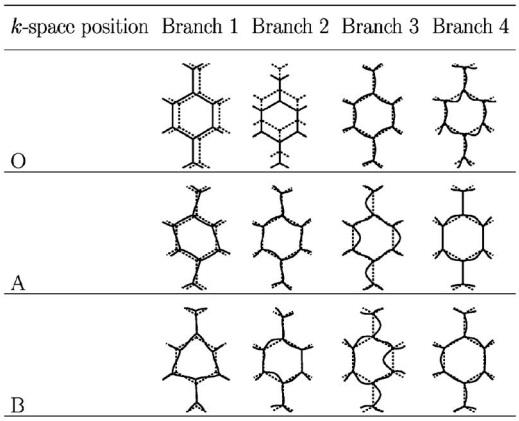
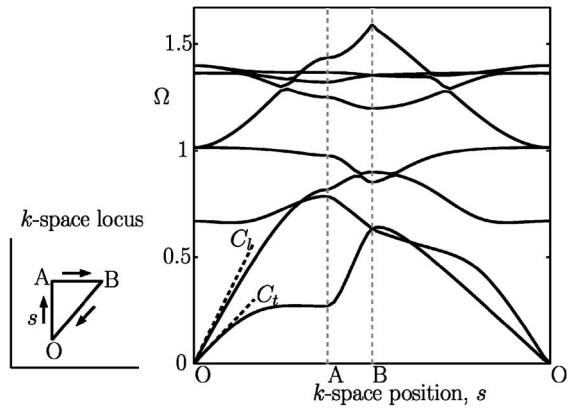


FIG. 8. Band structure of a hexagonal honeycomb with a slenderness ratio equal to 50. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.



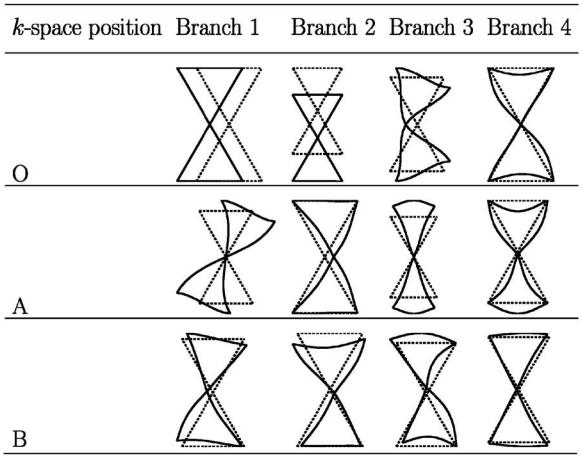


FIG. 9. Band structure of a Kagomé lattice with slenderness ratio equal to 10. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.

following governing equations in the reduced coordinates:

$$\tilde{\mathbf{D}}\tilde{\mathbf{q}} = \tilde{\mathbf{f}}, \quad \tilde{\mathbf{D}} = \mathbf{T}^{\mathrm{H}}\mathbf{D}\mathbf{T}, \quad \tilde{\mathbf{f}} = \mathbf{T}^{\mathrm{H}}\mathbf{f}, \tag{13}$$

where the superscript $^{\mathrm{H}}$ denotes the Hermitian transpose. For a plane wave propagating without attenuation in the x-y plane, the propagation constants along the x and y directions are $k_1 = i\epsilon_1$ and $k_2 = i\epsilon_2$. For free wave motion ($\mathbf{f}=0$), the above equation can be written in the frequency domain to give the following eigenvalue problem:

$$\tilde{\mathbf{D}}(k_1, k_2, \omega)\tilde{\mathbf{q}} = 0. \tag{14}$$

Any triad (k_1, k_2, ω) obtained by solving the eigenvalue problem in Eq. (14) represents a plane wave propagating at frequency ω.

In the eigenvalue problem defined by Eq. (14), there exist three unknowns: the two propagation constants k_1, k_2 that are complex, in general, and the frequency of wave propagation ω that is real, since the matrix $\tilde{\mathbf{D}}$ in the eigenvalue problem is Hermitian. At least two of the three unknowns have to be specified to obtain the third. For wave motion without attenuation the propagation constants are purely imaginary of the form $k_1 = i\epsilon_1$ and $k_2 = i\epsilon_2$. In this case one obtains the frequencies of wave propagation as a solution to the linear algebraic eigenvalue problem defined in Eq. (14) for each pair of phase constants (ϵ_1, ϵ_2). The solution is a surface, called the dispersion surface, in the $\omega - k_1 - k_2$ coordinates. There exist as many surfaces as there are eigenvalues of the problem in Eq. (14). If two surfaces do not overlap each other then there is a gap along the ω axis in which no wave motion occurs. This gap between dispersion surfaces is called the bandgap in the solid-state physics literature (Kittel, 1962) and the stop band in structural dynamics (Mead, 1973, 1996). For all frequencies on the phase constant surface, wave motion can occur and hence the frequency range occupied by these surfaces is a pass band. Furthermore, the normal to the phase constant surface at any point gives the Poynting vector or group velocity, and this indicates the speed and direction of energy flow.

V. BAND STRUCTURE OF THE FOUR LATTICE TOPOLOGIES

The computational procedure adopted to calculate the dispersion surfaces (band structure) for the four lattices is as follows.

(1) Select a primitive unit cell of the lattice.
(2) Construct the mass and stiffness matrices of the unit cell using the finite element technique described in Sec. III.
(3) Apply Bloch's principle to the equations of motion of the unit cell and form the eigenvalue problem in Eq. (14).
(4) Specify the phase constants (ϵ_1, ϵ_2) by restricting the wave vector to the edges of the irreducible part of the first Brillouin zone.

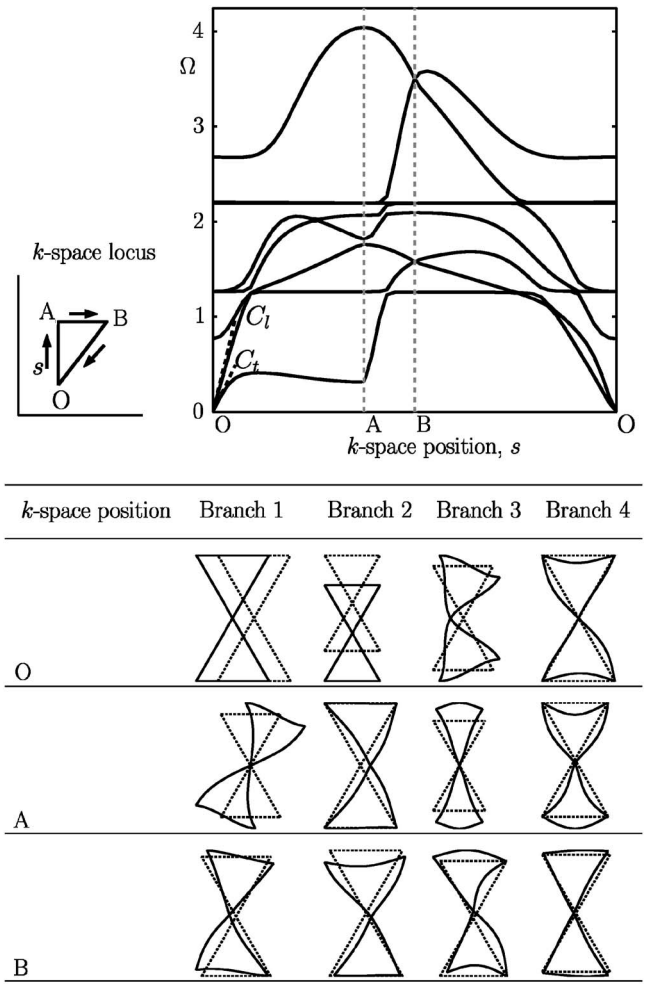


FIG. 10. Band structure of a Kagomé lattice with a slenderness ratio equal to 50. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.

(5) Solve the resulting linear algebraic eigenvalue problem in Eq. (14) for the wave propagation frequencies.

The four lattices are shown in Figs. 1–4 along with the choice of primitive unit cell, direct lattice translation vectors, and the first Brillouin zone. To explore the bandgaps it is sufficient to choose wave vectors along the edges of the first irreducible Brillouin zone (Kittel, 1962). Instead of solving the eigenvalue problem in Eq. (14) for each pair (ϵ_1, ϵ_2) over the shaded region OAB in the first Brillouin zone, one need only explore the edges of the triangle OAB. The parameter s is introduced as the arclength along the perimeter OABO of the shaded region OAB in the first Brillouin zone. It is a scalar pathlength parameter in k space and is used to denote the location of any point on the perimeter. Thus, the extremes of the frequency on a three-dimensional dispersion surface can be represented by a two-dimensional dispersion curve with the wave vector as the abscissa and the frequency as the ordinate. The bandgaps now correspond to regions along the ordinate wherein no dispersion branch is present. Within this frequency band wave motion cannot occur, and hence these are stop bands. In contrast, the frequency values for which there is at least one dispersion curve correspond to the pass bands. The long-wavelength limit corresponds to the origin, as denoted by point O in the first Brillouin zone.

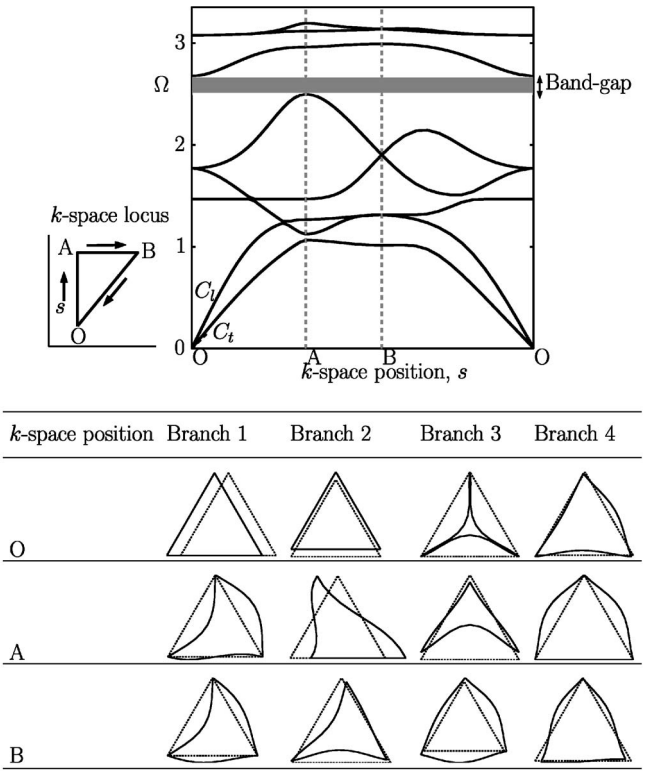


FIG. 11. Band structure of a triangular honeycomb with a slenderness ratio equal to 10. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.

The band structure of each lattice is computed by solving the eigenvalue problem in Eq. (14) for wave vectors along the closed locus O-A-B-O in k space. Results are presented for the two extremes of the slenderness ratio ($\lambda = 2\sqrt{3}L/d$) equal to 10 and 50. Recall that the relative density scales directly with the slenderness ratio, and the scaling factor for each geometry is summarized in Table I. For each topology, the band structure of the lattice with slenderness ratio equal to 20 was found to be similar to that of a lattice with slenderness ratio of 10, and so these plots are omitted in the interest of brevity. We consider in turn the hexagonal honeycomb, Kagomé lattice, triangular honeycomb, and finally the square honeycomb.

A. Hexagonal honeycomb

The dispersion curves for slenderness ratios of 10 and 50 are shown in Figs. 7 and 8. In these plots a nondimensional wave propagation frequency is plotted on the vertical axis as a function of the wave vector locus O-A-B-O along the edges of the irreducible first Brillouin zone, using the arclength parameter s in k space. A convenient nondimensional frequency Ω is defined as

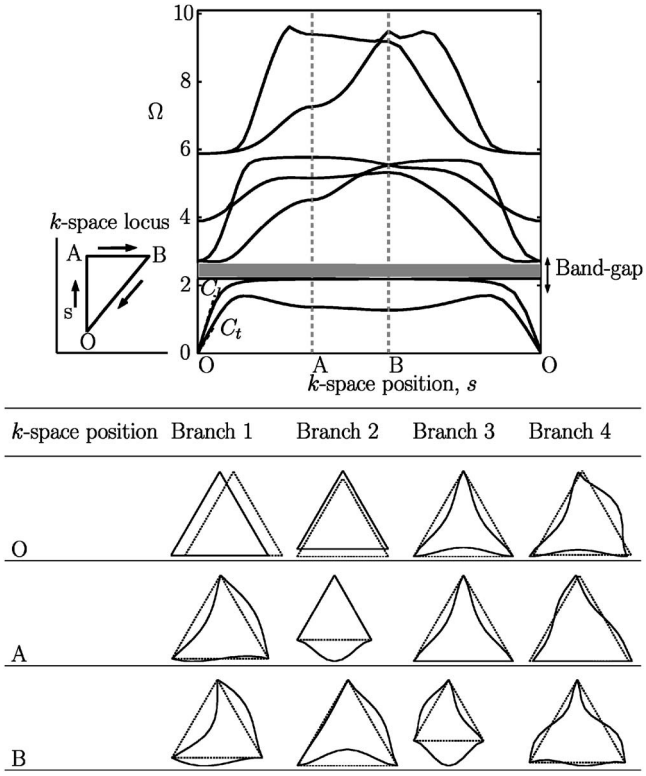


FIG. 12. Band structure of a triangular honeycomb with a slenderness ratio equal to 50. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.

$$\Omega = \frac{\omega}{\omega_1}, \tag{15}$$

where ω is the frequency of the plane wave obtained by solving the eigenvalue problem Eq. (14) and $\omega_1 = \pi^2 \sqrt{(EI/\rho\, dL^4)}$ is the first pinned-pinned flexural resonance frequency of a lattice beam. Consequently, at $\Omega = 1$ the cell deformation exhibits the first pinned-pinned flexural mode of the beam.

The point O corresponds to the long-wavelength limit, where the effective medium representation of the lattice is valid. Two branches of the dispersion curve emanate from the origin O: these are the longitudinal waves (also known as irrotational or dilatational waves) and transverse waves (also known as distortional, shear, or equivoluminal waves). Recall that the tangent to the dispersion curve at any point gives the group velocity while the secant slope of the line connecting the origin O to the point of interest on the dispersion curve gives the phase velocity. The two group velocities corresponding to the dispersion branches in the long-wavelength limit correspond to those of an effective elastic medium with modulus as given in Table I for the hexagonal honeycomb. For an isotropic medium with Young's modulus E^*, shear modulus G^*, bulk modulus K^*, and density ρ^*, the group velocities are given by

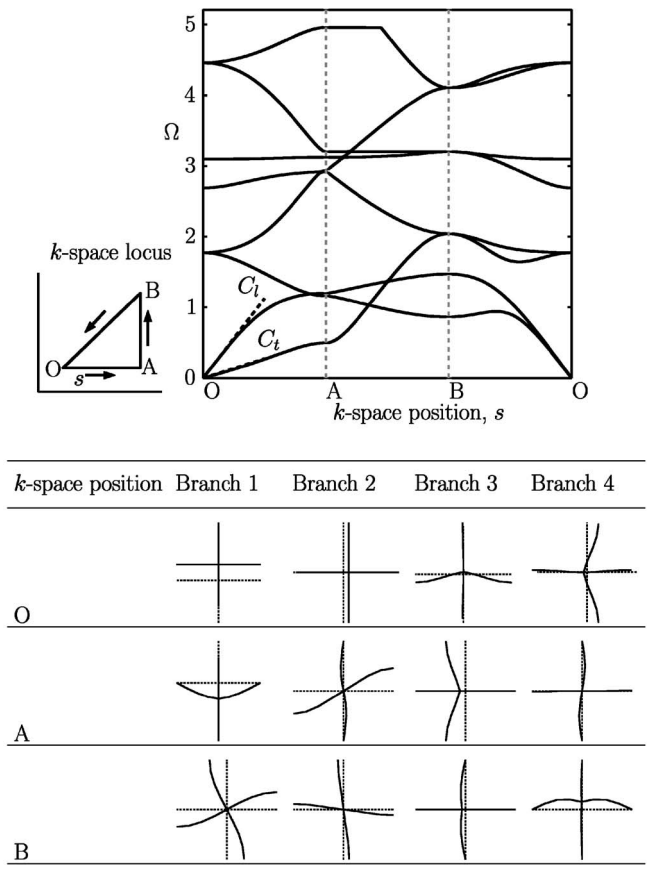


FIG. 13. Band structure of a square honeycomb with a slenderness ratio equal to 10. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.

$$C_l = \sqrt{\frac{K^* + G^*}{\rho^*}}, \quad C_t = \sqrt{\frac{G^*}{\rho^*}}. \tag{16}$$

Thus, the two lines starting at O and with slopes corresponding to the group velocities of C_l and C_t are a best approximation to the dispersion curves in the long-wavelength limit. The deviation of these long-wavelength asymptotes from the dispersion curves indicates the range of validity of effective medium theories: at higher frequencies of wave propagation the effective medium results in Table I do not apply. At these frequencies the detailed geometry of the lattice has a significant influence upon the wave-bearing properties.

A common feature of the dispersion curves for the hexagonal honeycomb (and for the other topologies) is the phenomenon of veering of frequencies (or repulsion of the dispersion branches), where the dispersion curves lay close to one another along the locus O-A-B-O in k space. This is a ubiquitous feature of eigenvalue problems of weakly coupled systems (Perkins and Mote, Jr., 1986). Such veering can be observed in Fig. 8 between the second and third branches of the dispersion curve along the locus O-A. From the magnified picture of the veering zone, it can be noted that the eigenvalues do not cross but veer away from each other.

The eigenwaves of a typical honeycomb cell for each dispersion curve are summarized in tabular form within Figs. 7 and 8 at the points O, A and B in k space. The three rows

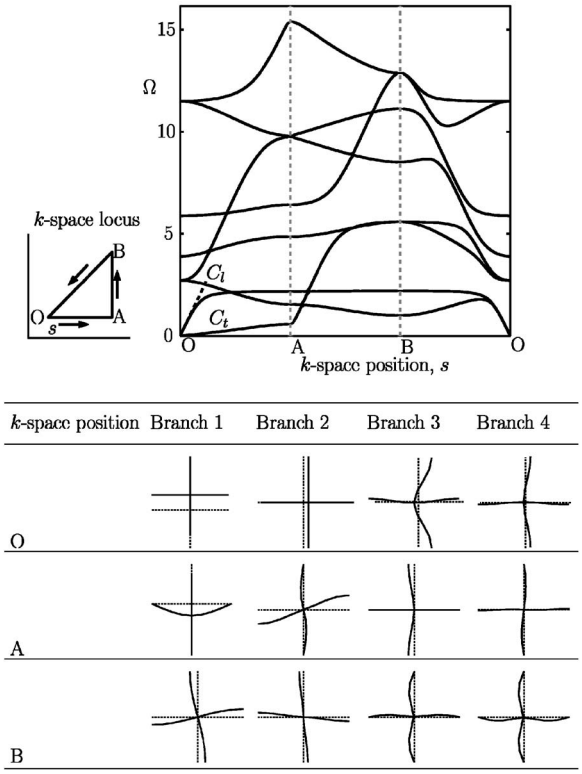


FIG. 14. Band structure of a square honeycomb with slenderness ratio equal to 50. The eigenwaves of a typical cell are shown in tabular form. The three rows correspond to the three points O, A, and B in k space, while the four columns correspond to the first four dispersion branches in ascending order.

in each table correspond to the three points O, A, and B, while the four columns correspond to the first four dispersion branches in ascending order. Consider the four eigenwaves in the first column of Fig. 7. At point O, the unit cell exhibits rigid body translation. Transverse wave motion occurs along the first branch, with increasing wave number; the transverse nature of the eigenwaves is clear from the cell deformation at point A. The second column of the table gives the second branch of the dispersion curve and this comprises a longitudinal wave. In higher branches, the eigenwave exhibits combined transverse and longitudinal motion.

The lattice with a slenderness ratio equal to 50 shows a complete bandgap (shaded space) between the sixth and the seventh branches of the dispersion curves. We note that no complete bandgaps are exhibited by the hexagonal honeycomb with a slenderness ratio of 10 (or 20). However, partial bandgaps for confined regions along the locus O-A-B-O are noted in Fig. 7.

B. Kagomé lattice

The dispersion responses of the Kagomé lattice (and of the triangular honeycomb and square honeycomb) are similarly shown in a plot of Ω versus length parameter s in k space. See Figs. 9 and 10 for slenderness ratios of 10 and 50, respectively. The long-wavelength asymptotes, as calculated from the group velocities using Eq. (16), are superimposed on the plots and agree with the dispersion curves. The Kagomé lattice does not exhibit any complete bandgaps for the slenderness ratios considered.

It is instructive to compare the long-wavelength behavior of transverse waves in the Kagomé and honeycomb lattices. For the honeycomb, the slope of the dispersion curve associated with the transverse wave decreases with an increase in slenderness ratio while the dispersion curves of the Kagomé lattice do not show this. This can be explained as follows. The transverse wave speed depends upon the ratio of effective shear modulus G^* to density according to Eq. (16). Now G^* scales as ρ^{*3} for the hexagonal honeycomb whereas G^* scales as ρ^* for the Kagomé lattice. Consequently, the transverse wave decreases with an increase in the slenderness ratio for the hexagonal honeycomb, but not for the Kagomé lattice.

The first dispersion curve constitutes transverse wave motion, while the second curve exhibits longitudinal motion. Higher branches display a combination of both transverse and longitudinal motion.

C. Triangular lattice

The dispersion results for the triangular lattice are given in Figs. 11 and 12 for slenderness ratios of 10 and 50, respectively. Again, the long-wavelength asymptotes agree with the dispersion curve. For all slenderness ratios considered (10, 20, and 50), a complete bandgap (shaded region) is present. But the location of the bandgap depends upon the slenderness ratio. For a ratio of 10, the gap exists between the fifth and sixth branches of the dispersion curve at a nondimensional frequency Ω centered on 2.6. At a slenderness ratio of 20, the gap exists between the sixth and seventh branches at a nondimensional frequency Ω centered on 4.6 (figure not shown). An increase in a slenderness ratio of 50 leads to a bandgap between the third and fourth branches at a nondimensional frequency centered on 2.4. The width of the bandgap is almost independent of the slenderness ratio, and equals about 0.2.

D. Square lattice

The dispersion results for the square lattice are given in Figs. 13 and 14. The group velocities for the longitudinal and shear waves in the long-wavelength limit are given by

$$C_l = \sqrt{\frac{2K^*}{\rho^*}}, \quad C_t = \sqrt{\frac{G^*}{\rho^*}}, \tag{17}$$

where the effective bulk modulus (K^*) and shear modulus (G^*) are given in Table I. As before, the long-wavelength asymptotes as calculated from the group velocities via Eq. (17) agree well with the dispersion curve.

The general features are similar to that already discussed for the other lattices: the first and second branches comprise shear waves and longitudinal waves, respectively. Pinned-pinned eigenwaves exist at $\Omega = 1$. Above this frequency value the dispersion curves tend to cluster. No complete bandgaps are observed over the range of slenderness ratios considered.

E. Directionality of wave propagation

The directionality of waves indicates the degree to which a medium is isotropic with respect to wave propaga-

tion. In an isotropic medium no preferred directions exist, and waves propagate equally in all directions. A common technique for displaying directionality is to construct isofrequency contours of the dispersion surface and to plot these contours in a Cartesian reference frame in k space. The physical coordinate system (x,y) has already been introduced for each lattice in Figs. 1–4. The wave vectors (k_x,k_y) are aligned with these physical orthonormal vectors, and form the axes of the isofrequency plots shown in Fig. 15 for each lattice at a slenderness ratio of 50. Contours are given for selected nondimensional frequencies Ω.

The relative shapes of the isofrequency contours are compared in Fig. 15 for the four topologies of interest. The symmetries of these contours show the rotational and reflective symmetries of the parent lattices. At low frequencies, the triangular and hexagonal honeycomb and Kagomé lattice are isotropic, whereas the square honeycomb is strongly anisotropic. At high frequencies, the lattice symmetries have a strong influence upon directionality.

VI. CONCLUSIONS

Floquet-Bloch principles have been used to explore plane wave propagation in several two-dimensional lattices. Dispersion curves are obtained by solving the eigenvalue problem for wave propagation, and the dependence of frequency upon the wave number has been determined in order to reveal the band structure. Three of the lattices are isotropic under static loading, while the fourth (the square honeycomb) is strongly anisotropic under static loading. The nodal connectivity varies from one structure to the next (from 3 to 6), and this has a major influence upon the static behavior. In this study we address whether similar differences in response occur under wave loading. The main findings are as follows.

(1) Long-wavelength asymptotes, based upon effective medium theories, agree with the dispersion curves at low frequency. However, at shorter wavelengths these asymptotes diverge from the dispersion curves, thereby indicating the range of validity of effective medium theories.
(2) Complete bandgaps are present for a wider range of relative density of lattice material for the triangular honeycomb than for the hexagonal honeycomb. They exist at frequencies above the first pin-pin resonance frequency for a single bar of the microstructure.
(3) Wave directionality plots at high frequencies show symmetries that are consistent with those of the parent lattice. Hexagonal, Kagomé, and triangular lattices all exhibit six-fold symmetry. They possess an isotropic response in the long-wavelength limit. The square lattice has four-fold symmetry and is strongly anisotropic over the full frequency range.

Imperfections in the lattice topology can lead to the localization of elastic waves, as discussed in Bendiksen (2000) and Hodges and Woodhouse (1983). While there is significant literature dealing with the localization of elastic waves in one-dimensional periodic structures, little is known about two-dimensional periodic structures. Future work will address the influence of imperfections on the band structure of

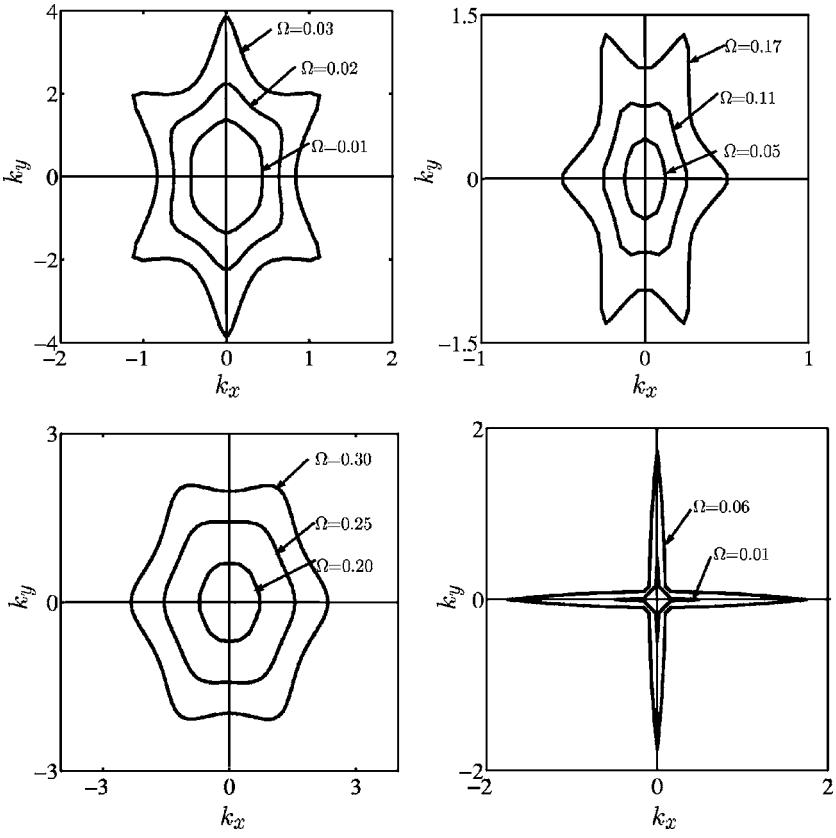


FIG. 15. Directionality of plane wave propagation in the four lattice topologies with slenderness ratio equal to 50: (a) hexagonal honeycomb; (b) Kagomé lattice; (c) triangular honeycomb; (d) square honeycomb. The nondimensional frequency (Ω) associated with each contour is labeled. At each frequency, energy flow direction is given by the normal to the contour and is in the direction of maximum rate of change of frequencies.

the topologies studied in this work. It is anticipated that a small imperfection is sufficient to alter significantly the band structure.

ACKNOWLEDGMENTS

The work carried out in this report is partially funded by the EU-HYMM project: 6th FP-NMP STREP—Contract No: NMP3-CT-2003-505206. Helpful technical discussions with R. S. Langley and Angelo Merlo are gratefully acknowledged.

APPENDIX: SHAPE FUNCTIONS OF TIMOSHENKO BEAM-COLUMN ELEMENT

The shape functions a_r, b_r, c_r, $r=1\cdots6$ for the six nodal displacements of a beam-column element shown in Fig. 5 are as follows:

$$a_1 = \frac{1}{2}(1-\xi),$$

$$a_2 = 0, \quad a_3 = 0,$$

$$a_4 = 1/2(1+\xi),$$

$$a_5 = 0, \quad a_6 = 0,$$

$$b_1 = 0,$$

$$b_2 = \frac{1}{4(1+3\beta)}[2+6\beta-3(1+2\beta)\xi+\xi^3],$$

$$b_3 = \frac{1}{4(1+3\beta)}[1+3\beta-\xi-(1+3\beta)\xi^2+\xi^3],$$

$$b_4 = 0,$$

$$b_5 = \frac{1}{4(1+3\beta)}[2+6\beta+3(1+2\beta)\xi-\xi^3],$$

$$b_6 = \frac{1}{4(1+3\beta)}[-(1+3\beta)-\xi+(1+3\beta)\beta^2+\xi^3],$$

$$c_1 = 0,$$

$$c_2 = \frac{1}{4(1+3\beta)}(-3+3\xi^2),$$

$$c_3 = \frac{1}{4(1+3\beta)}[-1+6\beta-(2+6\beta)\xi+3\xi^2],$$

$$c_4 = 0,$$

$$c_5 = \frac{1}{4(1+3\beta)}(3-3\xi^2),$$

$$c_6 = \frac{1}{4(1+3\beta)}[-1+6\beta+(2+6\beta)\xi+3\xi^2],$$

where

$$\xi = \frac{x}{L}, \quad \beta = \frac{EI_z}{\kappa GAL^2}. \tag{A1}$$

In the above equation, E and G denote the Young's modulus and rigidity modulus respectively; L, A, I_z denote the length, cross-sectional area and second moment of area of the beam, respectively; and κ denotes the shear correction factor used in Timoshenko beam theory (Weaver and Jonhston, 1987).

Bendiksen, O. O. (**2000**). "Localization phenomenon in structural dynamics," Chaos, Solitons Fractals **11**, 1621–1660.

Brillouin, L. (**1953**). *Wave Propagation in Periodic Structures*, 2nd ed. (Dover, New York).

Christensen, R. M. (**2000**). "Mechanics of cellular and other low density materials," Int. J. Solids Struct. **37**, 93–104.

Gibson, L. J., and Ashby, M. F. (**1997**). *Cellular Solids: Structure and Properties*, 2nd ed. (Cambridge University Press, Cambridge).

Hodges, C. H., and Woodhouse, J. (**1983**). "Vibration isolation from irregularity in a nearly periodic structure: theory and measurements," J. Acoust. Soc. Am. **74**, 894–905.

Hutchinson, R. G. (**2004**). "Mechanics of lattice materials," Ph.D. thesis, University of Cambridge, Department of Engineering.

Kittel, C. (**1962**). *Elementary Solid State Physics: A Short Course*, 1st ed. (Wiley, New York).

Langley, R. S. (**1993**). "A note on the forced boundary conditions for two-dimensional periodic structures with corner freedoms," J. Sound Vib. **167**, 377–381.

Langley, R. S., Bardell, N. S., and Ruivo, H. M. (**1997**). "The response of two-dimensional periodic structures to harmonic point loading: a theoretical and experimental study of a beam grillage," J. Sound Vib. **207**, 521–535.

Mead, D. (**1973**). "A general theory of harmonic wave propagation in linear periodic systems with multiple coupling," J. Sound Vib. **27**, 235–260.

Mead, D. (**1996**). "Wave propagation in continuous periodic structures: research contributions from Southampton 1964–1995," J. Sound Vib. **190**, 495–524.

Noor, A. K., Anderson, M. S., and Greene, W. H. (**1978**). "Continuum models for beam-and platelike lattice structures," AIAA J. **16**, 1219–1228.

Perkins, N. C., and Mote, C. D. Jr. (**1986**). "Comment on curve veering in eigenvalue problems," J. Sound Vib. **106**, 451–463.

Ruzzene, M., Scarpa, F., and Soranna, F. (**2003**). "Wave beaming effects in two-dimensional cellular structures," Smart Mater. Struct. **12**, 363–372.

Sigmund, O., and Jensen, J. S. (**2003**). "Systematic design of photonic bandgap materials and structures by topology optimization," Proc. R. Soc. London, Ser. A **361**, 1001–1019.

Torquato, S., Gubiansky, L. V., Silva, M. J., and Gibson, L. J. (**1998**). "Effective mechanical and transport properties of cellular solids," Int. J. Mech. Sci. **40**, 71–82.

Weaver, W., and Jonhston, P. R. (**1987**). *Structural Dynamics by Finite Elements*, 1st ed. (Prentice-Hall, Englewood Cliffs, NJ).

Wicks, N., and Hutchinson, J. W. (**2001**). "Optimal truss plates," Int. J. Solids Struct. **38**, 5165–5183.

Elastic wave propagation in sinusoidally corrugated waveguides

Sourav Banerjee[a] and Tribikram Kundu[b]
Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721

(Received 15 November 2005; revised 12 January 2006; accepted 13 January 2006)

The ultrasonic wave propagation in sinusoidally corrugated waveguides is studied in this paper. Periodically corrugated waveguides are gaining popularity in the field of vibration control and for designing structures with desired acoustic band gaps. Currently only numerical method (Boundary Element Method or Finite Element Method) based packages (e.g., PZFlex) are in principle capable of modeling ultrasonic fields in complex structures with rapid change of curvatures at the interfaces and boundaries but no analyses have been reported. However, the packages are very CPU intensive; it requires a huge amount of computation memory and time for its execution. In this paper a new semi-analytical technique called Distributed Point Source Method (DPSM) is used to model the ultrasonic field in sinusoidally corrugated waveguides immersed in water where the interface curvature changes rapidly. DPSM results are compared with analytical solutions. It is found that when a narrow ultrasonic beam hits the corrugation peaks at an angle, the wave propagates in the backward direction in waveguides with high corrugation depth. However, in waveguides with small corrugation the wave propagates in the forward direction. The forward and backward propagation phenomenon is found to be independent of the signal frequency and depends on the degree of corrugation. [DOI: 10.1121/1.2172170]

PACS number(s): 43.20.Fn, 43.20.El, 43.20.Bi [TDM] Pages: 2006–2017

I. INTRODUCTION

In recent years acoustic frequency filters are gaining popularity in the field of vibration and noise control and in acoustic bandgap analysis. The structures are being designed with periodic geometries to create acoustic bandgaps at desired frequencies. For the efficient design of such structures and correctly interpreting the experimental results with these structures, a complete understanding of elastic wave propagation in periodically corrugated structures is necessary. Another application of this study is in the nondestructive evaluation of different aerospace structures, components of integrated smart structures with nonplanar boundaries and civil structural components (rebars, pipelines, etc.).

The wave propagation analysis in structures with planar and curved boundaries has been the subject of numerous investigations for over five decades. The analytical solution of wave propagation in structures with nonplanar boundaries and interfaces has been the topic of investigation in the last three decades (Nayfeh *et al.*, 1978; Boström, 1983, 1989; Standström, 1986; Fokkemma, 1980; Glass and Maradudin, 1983; El-Bahrawy, 1994a, 1994b; Banerjee and Kundu, 2004; Declercq *et al.*, 2005). Stop bands and pass bands of the Rayleigh-Lamb symmetric modes in sinusoidally corrugated waveguides have been studied by El-Bahrawy (1994a). Only recently, generalized dispersion equations for periodically corrugated waveguides have been studied and solutions for both symmetric and antisymmetric modes in a sinusoidally corrugated waveguide have been presented (Banerjee and Kundu, 2006a).

In this paper a complete problem with a corrugated waveguide and two ultrasonic transducers is solved. The complete problem involves excitation of the corrugated plate by bounded acoustic beams that are generated by ultrasonic transducers of finite dimension [see Fig. 1(a)]. To solve this complete problem, appropriate modeling of the bounded acoustic beams in addition to the wave propagation modeling in corrugated plates is necessary. Modeling of ultrasonic and sonic fields generated by planar transducers of a finite dimension is one of the basic problems in textbooks (Rayleigh, 1965; Morse and Ingard,1968; Schmerr, 1998; Kundu, 2004). A good review of the earlier developments of the ultrasonic field modeling in front of a planar transducer can be found in Harris (1981). A list of the more recent developments in this field of research has been given by Sha *et al.* (2003). The pressure field in front of a planar transducer in homogeneous isotropic materials has been computed both in the time domain (Stepanishen, 1971; Harris, 1981; Jensen and Svendsen, 1992) and in the frequency domain (Ingenito and Cook, 1969; Lockwwod and Willette, 1973; Scarano *et al.*, 1985; Hah and Sung, 1992; Wu *et al.*, 1995; Lerch *et al.*, 1998). In addition to the ultrasonic field modeling in isotropic materials, progress has been made in the modeling of the ultrasonic radiation field in transversely isotropic and orthotropic media as well (Spies, 1994, 1995). Most of the abovementioned investigations are based on Huygen's principle, where the total field is obtained from the linear sum of point sources distributed over the transducer. The integral representation of this field is known as the Rayleigh-Sommerfield integral. Another technique based on the Gauss-Hermite beam model for ultrasonic field modeling in anisotropic ma-

[a]Electronic mail: sourav@email.arizona.edu
[b]Corresponding author: Tribikram Kundu, Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona 85721. Telephone: (520) 621 6573; electronic mail: tkundu@email.arizona.edu

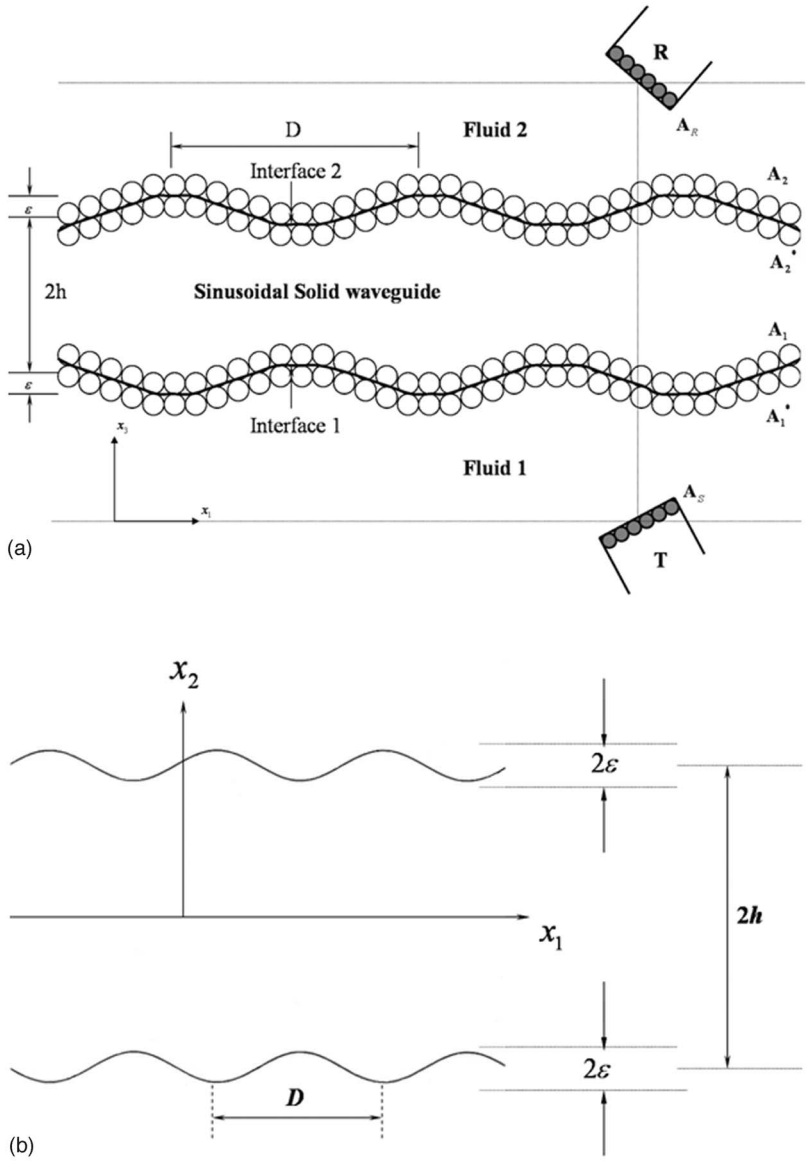


FIG. 1. (a) Sinusoidally corrugated waveguide between two transducers—geometry for the DPSM analysis. (b) Sinusoidally corrugated waveguide showing different parameters considered for the analytical solution.

terials with a paraxial approximation was proposed by Newberry and Thompson (1989). Since numerical integration is a time-consuming operation, Wen and Breazeale (1988) proposed an alternative approach. They computed the total field by superimposing a number of Gaussian beam solutions. They have shown that by superimposing only ten Gaussian solutions, the field radiated by a circular piston transducer can be modeled. Schmerr (2000) followed this approach to compute the ultrasonic field near a curved fluid-solid interface. Later Spies (1999) and Schmerr *et al.* (2003) extended this technique to a homogeneous anisotropic solid and water immersed anisotropic solid, respectively. Although a significant progress has been made in the ultrasonic field modeling in a homogeneous medium, the effect of curved interface with gradually varying curvature near an ultrasonic transducer of finite dimension has not been studied extensively yet. Recently Schmerr (2000) and Schmerr *et al.* (2003) studied the ultrasonic field near a fluid-solid curved interface. Spies (2004) studied the effect of the interface on the ultrasonic wave propagation in an inhomogeneous anisotropic medium with the farfield approximation. These investigators followed multi-Gaussian beam modeling approach. Although this technique has some computational advantage it also has a number of limitations similar to those of other paraxial models. For example, it cannot correctly model the critical reflection phenomenon; it cannot model a transmitted beam at an interface near grazing incidence. This technique also fails if the interface has different curvatures (gradually varying curvature), or when the radius of curvature of the transducer is small, as observed in acoustic microscopy experiments with its tightly focused lens. A detail description of the limitations of the multi-Gaussian paraxial models can be found in Schmerr *et al.* (2003).

The technique based on the DPSM (Distributed Point Source Method), proposed by Placko and Kundu (2001, 2004) avoids the above-mentioned limitations and does not require any farfield approximation. In this technique, one layer of point sources are distributed near the transducer face and two layers are placed near the interface. The advantage of the DPSM technique is that it not only avoids the paraxial

approximation it also does not require any ray tracing. The DPSM technique can handle complex geometries of the interface and the transducer. All methods developed before DPSM for the ultrasonic field radiation modeling near an interface requires ray tracing. The ray tracing technique becomes cumbersome in the presence of multiple interfaces while such geometries can be easily modeled by the DPSM technique (Banerjee, Kundu, and Placko, 2006).

The DPSM technique for ultrasonic field modeling was first developed by Placko and Kundu (2001). They successfully used this technique to model ultrasonic fields in a homogeneous fluid, and in a nonhomogeneous fluid with one interface (Lee *et al.*, 2002; Placko *et al.*, 2002) as well as multiple interfaces (Banerjee, 2005). The interaction between two transducers for different transducer arrangements and source strengths, placed in a homogeneous fluid, has been studied by Ahmad *et al.* (2003). The scattered ultrasonic field generated by a solid scatterer of finite dimension placed in a homogeneous fluid has also been modeled by the DPSM technique (Placko *et al.*, 2003). Recently the method has been extended to model the phased array transducers (Ahmad *et al.*, 2005). All these works modeled the ultrasonic field in a fluid medium. Only recently, the method has been extended to model the ultrasonic fields inside solid structures with planar boundaries (Banerjee and Kundu, 2006b). In the current paper the ultrasonic field in a sinusoidally corrugated waveguide has been modeled by the DPSM technique. The details of this modeling, as described in the subsequent sections, are quite challenging because of the continuous variations of the curvature of the fluid-solid interface. Numerical results for corrugated waveguides showing forward and backward propagations of guided waves depending on the degree of corrugation are reported here for the first time in the literature.

II. THEORY

A. Problem geometry

A symmetrically corrugated sinusoidal waveguide is considered. On two sides of the waveguide Fluid 1 and Fluid 2 are used as the coupling fluids that transmit ultrasonic waves from the ultrasonic transducers to the waveguide [see Fig. 1(a)]. To model the ultrasonic field inside the waveguide and the fluid, the DPSM technique (Placko *et al.*, 2001; Lee *et al.*, 2002; Ahmad *et al.*, 2005) is employed. Following the basics of the DPSM technique, four sets of point sources are distributed on both sides of the waveguide, as shown in Fig. 1(a). Point sources are also distributed behind the transducer faces. Transducer sources are denoted as $\mathbf{A}_S$ and $\mathbf{A}_R$ in Fig. 1(a). $\mathbf{A}_S, \mathbf{A}_R, \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_1^*$, and $\mathbf{A}_2^*$ are the source strength vectors for the sources distributed near the transducer surfaces and two interfaces [see Fig. 1(a)]. The period of corrugation of the sinusoidal waveguide is D and the depth of corrugation is equal to ε [see Fig. 1(b)].

B. Matrix formulation

The particle velocity and pressure in fluids at the interfaces can be expressed in matrix form (Kundu, 2004). Let $T1$ and $T2$ be two different sets of target points in the fluid below and above the Interfaces 1 and 2, respectively. The velocity at the target points can be written as

$$\mathbf{V}_{T1} = \mathbf{M}_{(T1)S}\mathbf{A}_S + \mathbf{M}_{(T1)1}\mathbf{A}_1, \tag{1}$$

$$\mathbf{V}_{T2} = \mathbf{M}_{(T2)R}\mathbf{A}_R + \mathbf{M}_{(T2)2^*}\mathbf{A}_2^*. \tag{2}$$

Similarly, the pressure fields at the target points are

$$\mathbf{PR}_{T1} = \mathbf{PR}_{T1}^s + \mathbf{PR}_{T1}^1 = \mathbf{Q}_{(T1)S}\mathbf{A}_S + \mathbf{Q}_{(T1)1}\mathbf{A}_1, \tag{3}$$

$$\mathbf{PR}_{T2} = \mathbf{PR}_{T2}^s + \mathbf{PR}_{T2}^{2^*} = \mathbf{Q}_{(T2)R}\mathbf{A}_R + \mathbf{Q}_{(T2)2^*}\mathbf{A}_2^*. \tag{4}$$

Elements of the matrices written in Eqs. (1)–(4) are given in Kundu (2004).

Boundary surfaces of the sinusoidal waveguide are nonplanar. At every point of the interface, normal stress and normal displacement are to be defined to satisfy the continuity conditions across the interface. The direction cosine of the sinusoidal waveguide at any point on the surface can be defined as $n=(n_1e_1+n_2e_2)$. Projections of unit normal (n) on x_1 and x_2 axes are given in Eqs. (5) and (6), respectively,

$$n_1 = \frac{\frac{2\pi\varepsilon}{D}\sin\left(\frac{2\pi x_1}{D}\right)}{\left[\left(\frac{2\pi\varepsilon}{D}\right)^2 \sin^2\left(\frac{2\pi x_1}{D}\right) + 1\right]^{1/2}}, \tag{5}$$

$$n_2 = \frac{1}{\left[\left(\frac{2\pi\varepsilon}{D}\right)^2 \sin^2\left(\frac{2\pi x_1}{D}\right) + 1\right]^{1/2}}. \tag{6}$$

Point sources needed for modeling isotropic solids are different from those used for fluid modeling. Every point source for the solid modeling has three different force components in three mutually perpendicular directions. For a point source acting at **y** in an isotropic solid, the stresses developed at point **x** have been expressed by Banerjee (2005) and Banerjee and Kundu (2006b). Assuming a point force acting along the x_j direction, stresses at point **x** on the boundary of the sinusoidal waveguide can be written as

$$\boldsymbol{\sigma}^j = \begin{bmatrix} \sigma_{11}^j & \sigma_{12}^j & \sigma_{13}^j \\ \sigma_{21}^j & \sigma_{22}^j & \sigma_{23}^j \\ \sigma_{31}^j & \sigma_{32}^j & \sigma_{33}^j \end{bmatrix}. \tag{7}$$

The transformation matrix at point **x** is

$$\mathbf{T} = \begin{bmatrix} n_2 & -n_1 & 0 \\ n_1 & n_2 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{8}$$

Therefore, transformed stresses at point **x** is

$$\underline{\boldsymbol{\sigma}'^{j}} = \mathbf{T}\boldsymbol{\sigma}^{j}\mathbf{T}^{T} = \begin{bmatrix} n_2 & -n_1 & 0 \\ n_1 & n_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11}^{j} & \sigma_{12}^{j} & \sigma_{13}^{j} \\ \sigma_{21}^{j} & \sigma_{22}^{j} & \sigma_{23}^{j} \\ \sigma_{31}^{j} & \sigma_{32}^{j} & \sigma_{33}^{j} \end{bmatrix} \times \begin{bmatrix} n_2 & n_1 & 0 \\ -n_1 & n_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \sigma_{11}'^{j} & \sigma_{12}'^{j} & \sigma_{13}'^{j} \\ \sigma_{21}'^{j} & \sigma_{22}'^{j} & \sigma_{23}'^{j} \\ \sigma_{31}'^{j} & \sigma_{32}'^{j} & \sigma_{33}'^{j} \end{bmatrix}. \quad (9)$$

To define the boundary conditions at point **x**, one normal stress, perpendicular to the sinusoidal boundary surface and two shear stresses, parallel to the boundary surface are needed. Considering a set of M point sources distributed on the sinusoidal surface, the normal stress and the shear stress components can be defined as

$$S_{22}' = \sum_{m=1}^{M} [(\sigma_{22}'^{1})^m P_1^m + (\sigma_{22}'^{2})^m P_2^m + (\sigma_{22}'^{3})^m P_3^m] = \sum_{m=1}^{M} \underline{\mathbf{s}_{22}'}^m \left(\frac{\mathbf{P}}{4\pi}\right)^m, \quad (10)$$

$$S_{21}' = \sum_{m=1}^{M} [(\sigma_{21}'^{1})^m P_1^m + (\sigma_{21}'^{2})^m P_2^m + (\sigma_{21}'^{3})^m P_3^m] = \sum_{m=1}^{M} \underline{\mathbf{s}_{21}'}^m \left(\frac{\mathbf{P}}{4\pi}\right)^m, \quad (11)$$

$$S_{23}' = \sum_{m=1}^{M} [(\sigma_{23}'^{1})^m P_1^m + (\sigma_{23}'^{2})^m P_2^m + (\sigma_{23}'^{3})^m P_3^m] = \sum_{m=1}^{M} \underline{\mathbf{s}_{23}'}^m \left(\frac{\mathbf{P}}{4\pi}\right)^m. \quad (12)$$

Displacements at point **x** generated by a point source acting at point **y** in an isotropic solid can be obtained from Mal and Singh (1991). The displacements at **x** due to the point force acting along the x_j direction are denoted as G_{1j}, G_{2j}, and G_{3j}. Considering the same point force along the x_j direction, the normal displacement of the sinusoidal solid surface at **x** can be written as

$$u_n^j = G_{1j}n_1 + G_{2j}n_2. \quad (13)$$

Considering a set of M point sources distributed on the interface, the normal displacement at point **x** on the sinusoidal surface can be written as

$$\underline{un} = \sum_{m=1}^{M} [(G_{11}n_1 + G_{21}n_2)^m P_1^m + (G_{12}n_1 + G_{22}n_2)^m P_2^m + (G_{13}n_1 + G_{23}n_2)^m P_3^m] = \sum_{m=1}^{M} \underline{\mathbf{Gn}}^m \mathbf{P}^m. \quad (14)$$

Let T be a set of target points in the solid. Normal displacements at these points (T) on the sinusoidal surface can be written in the following form:

$$\underline{\mathbf{un}}_T = \mathbf{DSn}_{T1^*}\mathbf{A}_1^* + \mathbf{DSn}_{T2}\mathbf{A}_2. \quad (15)$$

Similarly transformed normal stress and shear stresses at the target points (T) on the sinusoidal surface can be written as

$$\mathbf{s22}_T' = \mathbf{S22}_{T1^*}'\mathbf{A}_1^* + \mathbf{S22}_{T2}'\mathbf{A}_2, \quad (16a)$$

$$\mathbf{s21}_T' = \mathbf{S21}_{T1^*}'\mathbf{A}_1^* + \mathbf{S21}_{T2}'\mathbf{A}_2, \quad (16b)$$

$$\mathbf{s23}_T' = \mathbf{S23}_{T1^*}'\mathbf{A}_1^* + \mathbf{S23}_{T2}'\mathbf{A}_2. \quad (16c)$$

Matrices $\mathbf{DSn}_{TS}$ and $\mathbf{S22}_{TS}'$ are given in the Appendix [see Eqs. (A1) and (A2)]. Similarly $\mathbf{S21}_{TS}'$ and $\mathbf{S23}_{TS}'$ can be expressed. Subscripts T and S denote sets of target and source points, respectively.

In a fluid medium, the displacement components at point **x** generated by a point source at **y** are expressed as follows (Banerjee, 2005):

$$u_1 = \frac{1}{4\pi\rho\omega^2}\left(\frac{1}{r}ik_f R_1 e^{ik_f r} - \frac{e^{ik_f r}}{r^2}R_1\right), \quad (17)$$

$$u_2 = \frac{1}{4\pi\rho\omega^2}\left(\frac{1}{r}ik_f R_2 e^{ik_f r} - \frac{e^{ik_f r}}{r^2}R_2\right), \quad (18)$$

$$u_3 = \frac{1}{4\pi\rho\omega^2}\left(\frac{1}{r}ik_f R_3 e^{ik_f r} - \frac{e^{ik_f r}}{r^2}R_3\right), \quad (19)$$

where $R_j = (x_j - y_j)/r$, j takes values 1, 2, and 3.

Using the direction cosines (n_i) of the normal vector to the corrugated surface, the displacement component normal to the corrugated interface at point **x** can be written as

$$u_{fn} = u_1 n_1 + u_2 n_2. \quad (20)$$

Following the same rule in presence of transducers [see Fig. 1(a)], the displacement of the fluid at Interfaces 1 and 2 can be written as

$$\underline{\mathbf{Un}}_{I1} = [(\mathbf{DF2}_{(I1)S})n_2 + (\mathbf{DF1}_{(I1)S})n_1]\mathbf{A}_S + [(\mathbf{DF2}_{(I1)1})n_2 + (\mathbf{DF1}_{(I1)1})n_1]\mathbf{A}_1, \quad (21)$$

$$\underline{\mathbf{Un}}_{I2} = [(\mathbf{DF2}_{(I2)R})n_2 + (\mathbf{DF1}_{(I2)R})n_1]\mathbf{A}_R + [(\mathbf{DF2}_{(I2)2^*})n_2 + (\mathbf{DF1}_{(I2)2^*})n_1]\mathbf{A}_2^*, \quad (22)$$

or

$$\underline{\mathbf{Un}}_{I1} = \mathbf{DFn}_{(I1)S}\mathbf{A}_S + \mathbf{DFn}_{(I1)1}\mathbf{A}_1, \quad (23)$$

$$\underline{\mathbf{Un}}_{I2} = \mathbf{DFn}_{(I2)R}\mathbf{A}_R + \mathbf{DFn}_{(I2)2^*}\mathbf{A}_2^*. \quad (24)$$

Matrix $\mathbf{DFn}_{TS}$ is given in the Appendix [Eq. (A3)], where T and S denote sets of target and source points, respectively.

Let us consider a set of target points on "Interface 1" (then the set of target points will be denoted as $I1$) and the transformed normal stress and shear stress matrices for the referenced target points can be written as

$$\mathbf{s22}_{I1}' = \mathbf{S22}_{I11^*}'\mathbf{A}_1^* + \mathbf{S22}_{I12}'\mathbf{A}_2, \quad (25a)$$

$$\mathbf{s21}_{I1}' = \mathbf{S21}_{I11^*}'\mathbf{A}_1^* + \mathbf{S21}_{I12}'\mathbf{A}_2, \quad (25b)$$

$$\mathbf{s23}'_{I1} = \mathbf{S23}'_{I11^*}\mathbf{A}_1^* + \mathbf{S23}'_{I12}\mathbf{A}_2. \tag{25c}$$

Similarly, on Interface 2, the set of target points are denoted as $I2$ and the transformed normal and shear stresses on the sinusoidal surface can be written as

$$\mathbf{s22}'_{I2} = \mathbf{S22}'_{I21^*}\mathbf{A}_1^* + \mathbf{S22}'_{I22}\mathbf{A}_2, \tag{26a}$$

$$\mathbf{s21}'_{I2} = \mathbf{S21}'_{I21^*}\mathbf{A}_1^* + \mathbf{S21}'_{I22}\mathbf{A}_2, \tag{26b}$$

$$\mathbf{s23}'_{I2} = \mathbf{S23}'_{I21^*}\mathbf{A}_1^* + \mathbf{S23}'_{I22}\mathbf{A}_2. \tag{26c}$$

Inside the solid at interfaces $I1$ and $I2$, the normal displacements can be written as

$$\underline{\mathbf{un}}_{I1} = \mathbf{DSn}_{(I1)1^*}\mathbf{A}_1^* + \mathbf{DSn}_{(I1)2}\mathbf{A}_2, \tag{27}$$

$$\underline{\mathbf{un}}_{I2} = \mathbf{DSn}_{(I2)1^*}\mathbf{A}_1^* + \mathbf{DSn}_{(I2)2}\mathbf{A}_2. \tag{28}$$

C. Boundary and continuity conditions

Across the fluid-solid interface the displacement component normal to the interface should be continuous. Also, at the interface, the transformed normal stress ($\mathbf{s22}'$) in the solid and pressure in the fluid should be continuous. Whereas, the shear stresses at the interface must vanish. Let the normal velocities at the transducer faces be $\mathbf{V}_{S0}$ and $\mathbf{V}_{R0}$, for the lower and upper transducers, respectively. The boundary conditions at the transducer faces are

$$\mathbf{M}_{SS}\mathbf{A}_S + \mathbf{M}_{S1}\mathbf{A}_1 = \mathbf{V}_{S0}, \tag{29}$$

$$\mathbf{M}_{R2^*}\mathbf{A}_2^* + \mathbf{M}_{RR}\mathbf{A}_R = \mathbf{V}_{R0}. \tag{30}$$

At the interfaces, from the continuity of the normal stress,

$$\mathbf{Q}_{1S}\mathbf{A}_S + \mathbf{Q}_{11}\mathbf{A}_1 = -\mathbf{S22}'_{11^*}\mathbf{A}_1^* - \mathbf{S22}'_{12}\mathbf{A}_2, \tag{31}$$

$$\mathbf{Q}_{22^*}\mathbf{A}_2^* + \mathbf{Q}_{2R}\mathbf{A}_R = -\mathbf{S22}'_{21^*}\mathbf{A}_1^* - \mathbf{S22}'_{22}\mathbf{A}_2. \tag{32}$$

Continuity of the normal displacement gives

$$\mathbf{DFn}_{1S}\mathbf{A}_S + \mathbf{DFn}_{11}\mathbf{A}_1 = \mathbf{DSn}_{11^*}\mathbf{A}_1^* + \mathbf{DSn}_{12}\mathbf{A}_2, \tag{33}$$

$$\mathbf{DFn}_{22^*}\mathbf{A}_2^* + \mathbf{DFn}_{2R}\mathbf{A}_R = \mathbf{DSn}_{21^*}\mathbf{A}_1^* + \mathbf{DSn}_{22}\mathbf{A}_2, \tag{34}$$

and from the vanishing shear stress condition at the fluid-solid interface,

$$\mathbf{S21}'_{11^*}\mathbf{A}_1^* + \mathbf{S21}'_{12}\mathbf{A}_2 = 0, \tag{35}$$

$$\mathbf{S23}'_{11^*}\mathbf{A}_1^* + \mathbf{S23}'_{12}\mathbf{A}_2 = 0. \tag{36}$$

Equations (29)–(36) can be written in matrix form,

$$\begin{bmatrix}
\mathbf{M}_{SS} & \mathbf{M}_{S1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{Q}_{1S} & \mathbf{Q}_{11} & \mathbf{S22}'_{11^*} & \mathbf{S22}'_{12} & \mathbf{0} & \mathbf{0} \\
\mathbf{DFn}_{1S} & \mathbf{DFn}_{11} & -\mathbf{DSn}_{11^*} & -\mathbf{DSn}_{12} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{S21}'_{11^*} & \mathbf{S21}'_{12} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{S23}'_{11^*} & \mathbf{S23}'_{12} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{S23}'_{21^*} & \mathbf{S23}'_{22} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{S21}'_{21^*} & \mathbf{S21}'_{22} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{S22}'_{21^*} & \mathbf{S22}'_{22} & \mathbf{Q}_{22^*} & \mathbf{Q}_{2R} \\
\mathbf{0} & \mathbf{0} & -\mathbf{DSn}_{21^*} & -\mathbf{DSn}_{22} & \mathbf{DFn}_{22^*} & \mathbf{DFn}_{2R} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{M}_{R2^*} & \mathbf{M}_{RR}
\end{bmatrix}_{(2N+8M)x(2N+8M)}
\begin{Bmatrix}
\mathbf{A}_S \\ \mathbf{A}_1 \\ \mathbf{A}_1^* \\ \mathbf{A}_2 \\ \mathbf{A}_2^* \\ \mathbf{A}_R
\end{Bmatrix}_{(2N+8M)}
=
\begin{Bmatrix}
\mathbf{V}_{S0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{V}_{R0}
\end{Bmatrix}_{(2N+8M)}, \tag{37}$$

or

$$[\underline{\mathbf{MT}}]\{\mathbf{\Omega}\} = \{\mathbf{V}\}. \tag{38}$$

D. Solution

The vector of source strengths of the complete system can be obtained from Eq. (38) by taking inverse of $[\underline{\mathbf{MT}}]$ and multiplying it with the vector $\{\mathbf{V}\}$,

$$\{\mathbf{\Omega}\} = [\underline{\mathbf{MT}}]^{-1}\{\mathbf{V}\}. \tag{39}$$

After calculating the source strengths, the pressure, velocity, stress, and displacement values at any point can be obtained.

E. Analytical solution of wave propagation in sinusoidally corrugated waveguide

The analytical solution for the complete problem geometry including the waveguide and two transducers as shown in Fig. 1(a) is not available. However, the problem of guided wave propagation in a corrugated plate as shown in Fig. 1(b)

can be solved analytically. Wave propagation in corrugated waveguides with small corrugation, where the perturbation method can be applied, was first studied by Nayfeh *et al.* (1978). This solution cannot be used in many practical applications when the corrugation height is not necessarily small in comparison to the plate thickness. The analysis of wave propagation in electromagnetic waveguides with a high degree of corrugation was studied by Boström (1983). Later, Standström (1986) discussed stop bands in sinusoidally corrugated waveguides by applying the null-field approach, developed by Waterman (1975). Standström (1987) compared different techniques for the corrugated plate analysis.

The elastic wave propagation analysis near sinusoidally corrugated fluid-solid interface by the modal superposition technique has been discussed by Fokkemma (1980). Although a number of researchers have studied the electromagnetic wave propagation near surface grating and in corrugated waveguides, not many investigators have studied the problem of elastic wave propagation in corrugated plates. The problem of elastic wave propagation in a sinusoidally corrugated waveguide has been considered by El-Bahrawy (1994a) for only symmetric Rayleigh-Lamb modes. A classical modal technique was adopted for this analysis. In El-Bahrawy's study the dispersion equation was developed for only symmetric modes. Stop bands and pass bands of the symmetric modes were studied extensively by El-Bahrawy.

In this paper the analytical solution is adopted from El-Bahrawy's (1994a) work. The dispersion relation for the symmetric modes in a sinusoidally corrugated waveguide is presented in Eq. (40). The parameters (ε, D, h) used in the following equations are defined in Fig. 1(b):

$$T_{ij}\alpha_j = 0. \tag{40a}$$

Therefore, for nontrivial solutions of α_j,

$$\mathrm{Det}[T] = 0, \tag{40b}$$

where

$$\begin{aligned}
T_{1n1m} &= -i\frac{D}{2}\left(\frac{k_n}{\beta_n}\right)(i^{n-m}e^{ih\beta_n} - i^{m-n}e^{-ih\beta_n}) \\
&\quad\times\left[\left(\frac{4(n-m)\pi}{D}\right)k_n + 2\beta_n^2 - k_s^2\right]J_{n-m}(\varepsilon\beta_n), \\
T_{1n2m} &= -(i^{n-m}e^{ih\eta_n} - i^{m-n}e^{-ih\eta_n}) \\
&\quad\times\left[\left(\frac{(n-m)\pi}{\eta_n}\right)(k_s^2 - 2\eta_n^2) + Dk_n\eta_n\right]J_{n-m}(\varepsilon\beta_n), \\
T_{2n1m} &= -k_n(i^{n-m}e^{ih\beta_n} + i^{m-n}e^{-ih\beta_n}) \\
&\quad\times\left[\left(\frac{(n-m)\pi}{\beta_n^2}\right)(k_s^2 - 2\beta_n^2) + Dk_n\right]J_{n-m}(\varepsilon\beta_n), \\
T_{2n2m} &= -(i^{n-m}e^{ih\eta_n} + i^{m-n}e^{-ih\eta_n}) \\
&\quad\times\left(-\frac{D}{2}(k_s^2 - 2\beta_n^2) + 2\pi(n-m)k_n\right)J_{n-m}(\varepsilon\beta_n).
\end{aligned} \tag{40c}$$

In the above equations, if n and m take values 1, 2, 3, …, p, then i and j take values 1, 2, 3, …, 2IpI, (2IpI+1).

The displacement function can be written as

$$u_k = w_{kj}\alpha_j, \tag{41}$$

where k takes values 1, 2, and 3.

The displacement functions have been given by El-Bahrawy (1994a). Equation (40) is solved for a particular frequency and the eigenvectors corresponding to the wave number solutions are calculated. The eigenvector solutions are substituted in Eq. (41) to get displacement mode shape in the waveguide for a specific mode.

The above analytical solution is for the plane wave propagation in the waveguide. However, the wave field in the waveguide for the DPSM modeling is generated by two bounded acoustic beams. Therefore, perfect matching between the DPSM generated results and the analytical mode shapes is not expected. Only a qualitative comparison between these two results is presented in the following section. The symmetric transducer placement in the DPSM formulation generates only the symmetric modes in the waveguide. Hence, only the symmetric mode solutions of the analytical formulation are compared with the DPSM results.

TABLE I. Waveguide geometry [see Fig. 1(b)].

	$2h$	ε	D	ε/D
Waveguide 1	10	0.5	10	0.05
Waveguide 2	10	1	10	0.1
Waveguide 3	10	1.5	10	0.15
Waveguide 4	10	2	10	0.2

III. NUMERICAL IMPLEMENTATION

MATLAB 7.1 R-14 and Lapack library functions are used to generate the numerical results based on the formulation presented above. The numerical results are presented for the corrugated aluminum waveguides with Lamé constants λ and μ equal to 54.55 and 24.95 GPa, respectively, and density equal to 2.7 gm/cm^3. P-wave and S-wave speeds (c_p =6220 m/s and c_s=3040 m/s) in the material are obtained from the above elastic constants. Four different waveguides are considered in the analysis. Dimensions of the waveguides are presented in Table I. Comparisons between DPSM and analytical solutions are presented for Waveguide 2.

Equation (40) is solved numerically for two different frequencies from the pass band frequencies [El-Bahrway (1994a)]. The ultrasonic fields for these frequencies are also generated by the DPSM technique. The absolute values of the horizontal and vertical displacement components computed by these two methods are presented in Figs. 2(a) and 2(b), respectively. The plots show the displacement variations along the plate thickness. The displacement fields are normalized with respect to the horizontal displacement at x_2=0 [see Fig. 1(b)]. For comparison purposes the displacement field from the DPSM formulation is generated away from the transducers to capture the propagating guided wave modes away from the zone affected by the striking ultrasonic beams. The displacement fields corresponding to the first two symmetric modes generated from Eq. (41) are multiplied by

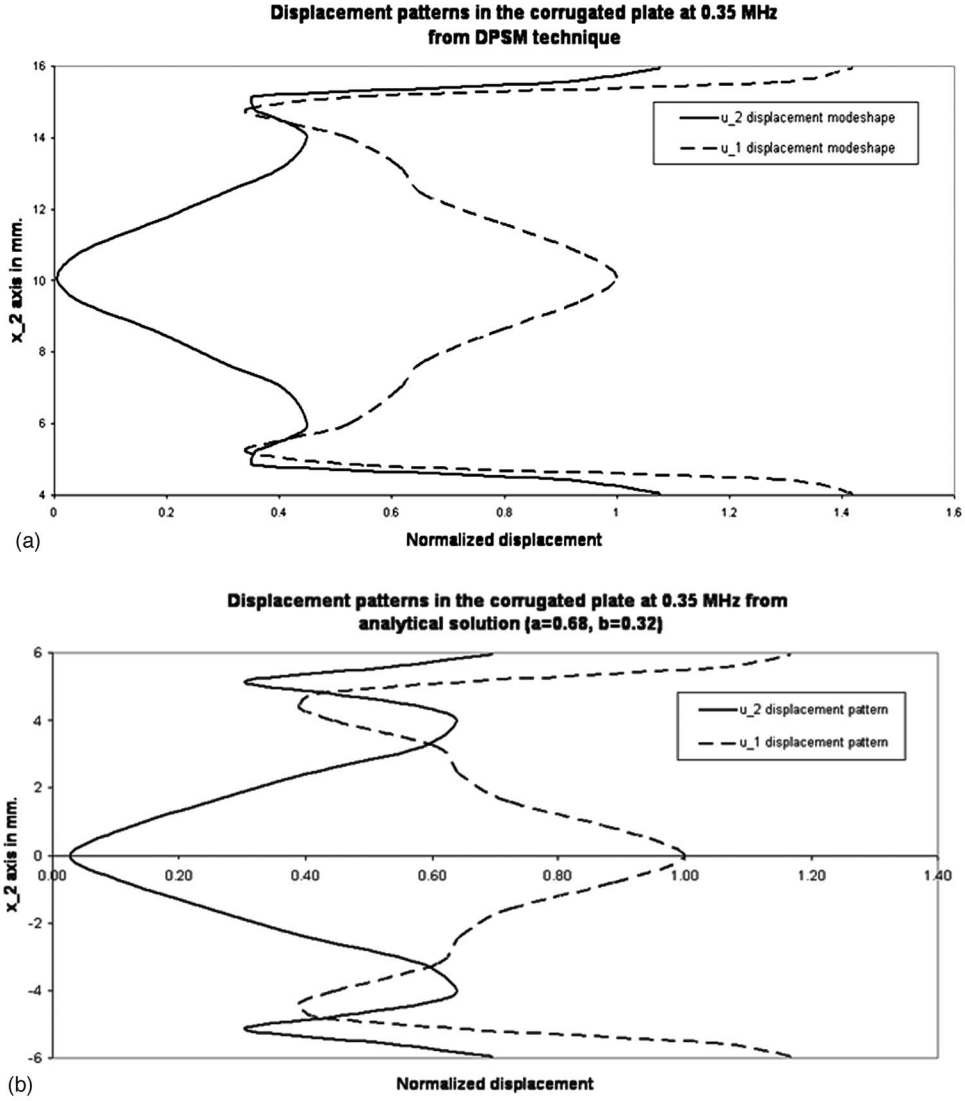


FIG. 2. Horizontal and vertical displacement variations at 0.35 MHz along the plate thickness obtained by (a) DPSM and (b) analytical solution techniques.

two weight factors and added to approximately match the DPSM results. The weighted displacement field is calculated as

$$u_i = w_1 u_i^1 + w_2 u_i^2, \tag{42}$$

where u_i^1 and u_i^2 are displacement components along the x_i direction generated by the fundamental and first higher symmetric modes, respectively. Results presented in Fig. 2(b) are generated with $w_1=0.68$ and $w_2=0.32$. Clearly, the DPSM results are qualitatively in good agreement with the analytical solution.

Ultrasonic fields in four different waveguides (see Table I) are generated by the DPSM technique. A normal incidence of the ultrasonic beam on a corrugation peak of the waveguide is considered first and then the transducers are inclined at two different angles. Results for three different orientations of the transducers are presented. Figure 3 shows different transducer orientations. The transducer frequency is set at 1 MHz. Figures 4 and 5 show the horizontal (u_1) and vertical (u_2) displacement fields, respectively, inside the waveguides. In these two figures the displacement fields are presented for Waveguides 2, 3, and 4 (see Table I for their dimensions). Figures 4(a), 4(b), and 4(c) show the u_1 displacement for normal incidence [transducer orientation is shown in Fig. 5(a)] in Waveguides 2, 3, and 4, respectively. Figures 4(d),

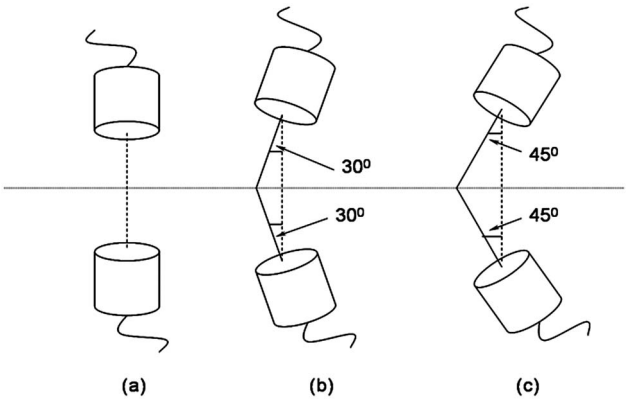


FIG. 3. Transducer orientations (a) Orientation—I: Normal Incidence. (b) Orientation—II: 30° inclination. (c) Orientation—III: 45° inclination.

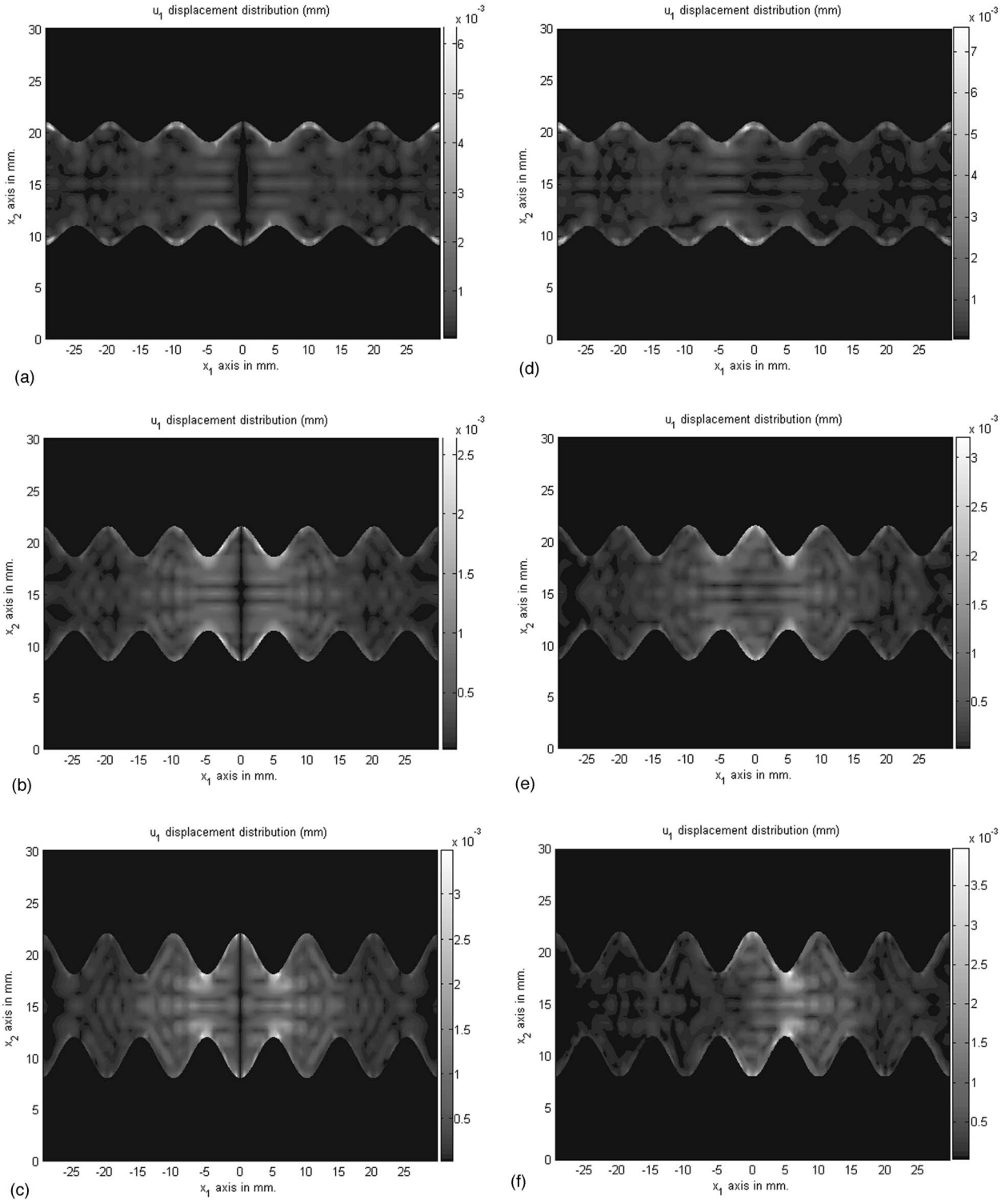


FIG. 4. Horizontal displacement fields in three different corrugated waveguides (2, 3, and 4) for two different angles of strike (0° and 30°). (a) in Waveguide 2 for normal incidence, (b) in Waveguide 3 for normal incidence, (c) in Waveguide 4 for normal incidence, (d) in Waveguide 2 for a 30° inclination angle, (e) in Waveguide 3 for a 30° inclination angle, (f) in Waveguide 4 for a 30° inclination angle. However, for inclined incidence, more energy is observed in the backward direction ($x>0$) in Fig. 4(f) (large corrugation) while the opposite trend is noticed in Fig. 4(e) (small corrugation). Table I gives waveguide dimensions.

4(e), and 4(f) show the u_1 displacement for 30° striking angle (transducer orientation is shown in Fig. 3(b)) in Waveguides 2, 3, and 4, respectively. Similarly, Figs. 5(a), 5(b), and 5(c) show the u_2 displacement for normal incidence in Waveguides 2, 3, and 4, respectively, and Figs. 5(d), 5(e), and 5(f) show the u_2 displacement for a 30° striking angle in Waveguides 2, 3, and 4, respectively. It can be seen from Figs. 4 and 5 that the ultrasonic waves in Waveguide 2 [Figs. 4(d) and 5(d)] propagate in the forward direction, or in other words, in the same direction as the horizontal component of

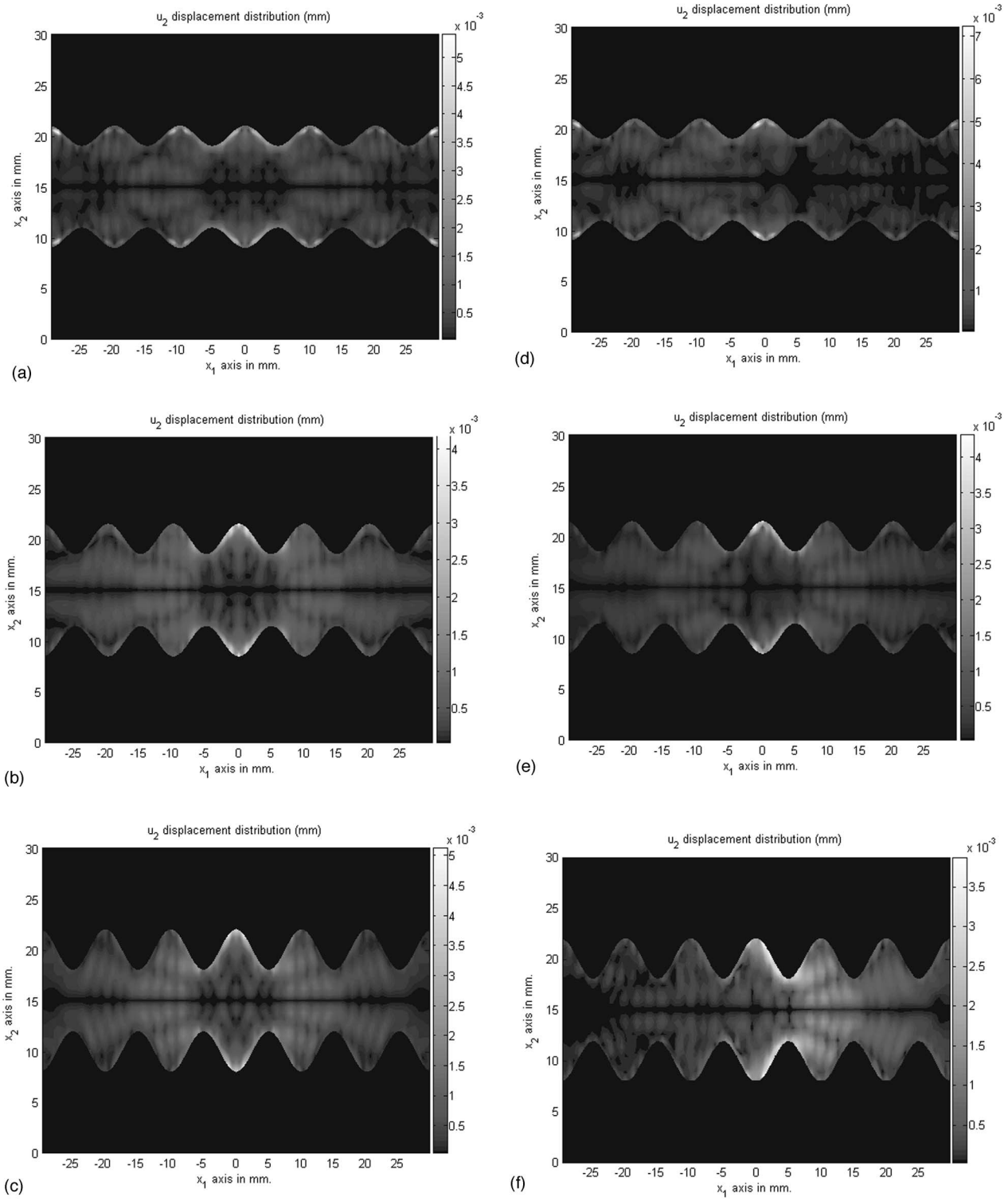


FIG. 5. Vertical displacement fields in three different corrugated waveguides (2, 3, and 4) for two different angles of strike (0° and 30°) (a) in Waveguide 2 for normal incidence, (b) in Waveguide 3 for normal incidence, (c) in Waveguide 4 for normal incidence, (d) in Waveguide 2 for a 30° inclination angle, (e) in Waveguide 3 for a 30° inclination angle, (f) in Waveguide 4 for a 30° inclination angle. As expected, symmetric displacement fields are observed for normal incidence [(a), (b), (c)]. However, for inclined incidence, more energy is observed in the backward direction ($x>0$) in (f) (large corrugation) while the opposite trend is noticed in (e) (small corrugation). Table I gives waveguide dimensions.

the striking beams. In Waveguide 4 [Figs. 4(f) and 5(f)] ultrasonic waves in the waveguide propagate in the backward direction, or, in other words, opposite to the direction of the striking beams. In Waveguide 3 [Figs. 4(e) and 5(e)] the wave propagates in both directions. The phenomenon of the wave propagation in the backward direction in Waveguides 4 and 3 is called "back-propagation." The back-propagation phenomenon can be more clearly seen in Fig. 6. Figure 6 shows amplitudes of u_1 displacement along the central plane of the waveguides. In this figure the displacement variations

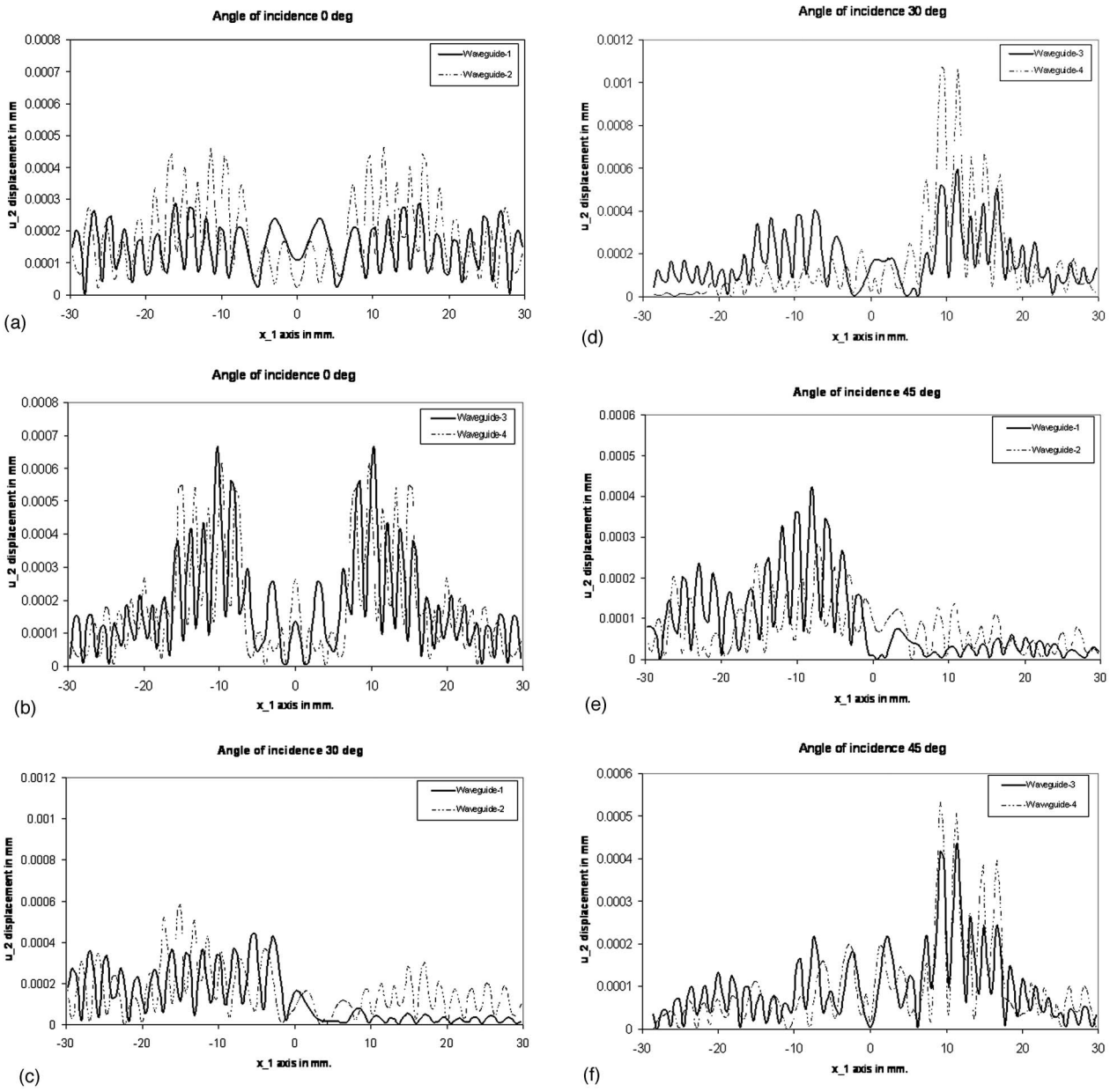


FIG. 6. Vertical displacements at the horizontal central planes of four different corrugated waveguides (dimensions are given in Table I) for three different striking angles (shown in Fig. 3) (a) normal incidence in Waveguides 1 and 2, (b) normal incidence in Waveguides 3 and 4, (c) 30° incidence in Waveguides 1 and 2, (d) 30° incidence in Waveguides 3 and 4, (e) 45° incidence in Waveguides 1 and 2, (f) 45° incidence in waveguides 3 and 4. As expected, (a) and (b) show the symmetric response for normal incidence. For inclined incidence, (c) and (e) show strong wave propagation in the forward direction for small corrugation (Waveguides 1 and 2), while (d) and (f) show strong backward direction wave propagation for large corrugation (Waveguides 3 and 4).

in all four waveguides listed in Table I are shown. These displacement fields are generated for three different transducer orientations, as shown in Fig. 3. Figure 6 clearly shows the back-propagation of ultrasonic waves [Figs. 6(d) and 6(f)] for large corrugation ($\varepsilon/D=0.2$ and 0.15) and forward propagation [Figs. 6(c) and 6(e)] for small corrugation ($\varepsilon/D=0.05$ and 0.1) when the ultrasonic beam strikes the plate at an angle. The ε/D ratio was carefully changed between 0.1 and 0.15 to find out for what value of this ratio the back-propagation starts to dominate. It is found that for the inclined incidence of the ultrasonic bounded beam on a corrugation peak when $\varepsilon/D\approx0.11$ the ultrasonic waves propagate in both directions with almost equal strength. For $\varepsilon/D>0.11$ the back-propagation dominates and for $\varepsilon/D<0.11$ the forward propagation dominates. When the signal frequency in Figs. 4–6 was changed from 1 to 2 MHz, the details of the figures changed to some extent, however, the general conclusion about the forward and backward propagation phenomenon did not change. For 2 MHz plots also (not shown here) it was observed that for $\varepsilon/D>0.11$ the back-propagation dominates and for $\varepsilon/D<0.11$ the forward propagation dominates.

IV. CONCLUSION

Elastic wave propagation in corrugated plates is modeled by the DPSM technique. Displacement mode shapes

generated by DPSM are compared with those obtained analytically. Good qualitative matching between the two sets of mode shapes is obtained. This analysis shows that when bounded acoustic beams strike a corrugated plate at an angle, the elastic waves can propagate in both forward and backward directions in the waveguide depending on the degree of corrugation. The back propagation of ultrasonic waves in corrugated waveguides for large corrugation depth is reported for the first time in this paper.

ACKNOWLEDGMENTS

The authors thank Dr. A. Boström and Dr. A. El-Bahrawy, Chalmers University of Technology, Division of Mechanics, Göteborg, Sweden and Dr. A. K. Mal, University of California, Los Angeles, for sending valuable research material on this subject. This research was partially funded by a grant from the National Science Foundation under Contract No. CMS-9901221.

APPENDIX:

Matrices expressions:

$$\mathbf{DSn}_{TS}=\begin{bmatrix} \mathbf{Gn}_1^1 & \mathbf{Gn}_1^2 & \mathbf{Gn}_1^3 & \cdots & \mathbf{Gn}_1^{M-1} & \mathbf{Gn}_1^M \\ \mathbf{Gn}_2^1 & \mathbf{Gn}_2^2 & \mathbf{Gn}_2^3 & \cdots & \mathbf{Gn}_2^{M-1} & \mathbf{Gn}_2^M \\ \mathbf{Gn}_3^1 & \mathbf{Gn}_3^2 & \mathbf{Gn}_3^3 & \cdots & \mathbf{Gn}_3^{M-1} & \mathbf{Gn}_3^M \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{Gn}_{N-1}^1 & \mathbf{Gn}_{N-1}^2 & \mathbf{Gn}_{N-1}^3 & \cdots & \mathbf{Gn}_{N-1}^{M-1} & \mathbf{Gn}_{N-1}^M \\ \mathbf{Gn}_N^1 & \mathbf{Gn}_N^2 & \mathbf{Gn}_N^3 & \cdots & \mathbf{Gn}_N^{M-1} & \mathbf{Gn}_N^M \end{bmatrix}_{(Nx3M)}, \tag{A1}$$

$$\mathbf{S22}'_{TS}=\begin{bmatrix} \mathbf{s}_{22^1}^1 & \mathbf{s}_{22^1}^2 & \mathbf{s}_{22^1}^3 & \mathbf{s}_{22^1}^4 & \mathbf{s}_{22^1}^5 & \cdots & \cdots & \mathbf{s}_{22^1}^{M-2} & \mathbf{s}_{22^1}^{M-1} & \mathbf{s}_{22^1}^M \\ \mathbf{s}_{22^2}^1 & \mathbf{s}_{22^2}^2 & \mathbf{s}_{22^2}^3 & \mathbf{s}_{22^2}^4 & \mathbf{s}_{22^2}^5 & \cdots & \cdots & \mathbf{s}_{22^2}^{M-2} & \mathbf{s}_{22^2}^{M-1} & \mathbf{s}_{22^2}^M \\ \mathbf{s}_{22^3}^1 & \mathbf{s}_{22^3}^2 & \mathbf{s}_{22^3}^3 & \mathbf{s}_{22^3}^4 & \mathbf{s}_{22^3}^5 & \cdots & \cdots & \mathbf{s}_{22^3}^{M-2} & \mathbf{s}_{22^3}^{M-1} & \mathbf{s}_{22^3}^M \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathbf{s}_{22^{N-2}}^1 & \mathbf{s}_{22^{N-2}}^2 & \mathbf{s}_{22^{N-2}}^3 & \mathbf{s}_{22^{N-2}}^4 & \mathbf{s}_{22^{N-2}}^5 & \cdots & \cdots & \mathbf{s}_{22^{N-2}}^{M-2} & \mathbf{s}_{22^{N-2}}^{M-1} & \mathbf{s}_{22^{N-2}}^M \\ \mathbf{s}_{22^{N-1}}^1 & \mathbf{s}_{22^{N-1}}^2 & \mathbf{s}_{22^{N-1}}^3 & \mathbf{s}_{22^{N-1}}^4 & \mathbf{s}_{22^{N-1}}^5 & \cdots & \cdots & \mathbf{s}_{22^{N-1}}^{M-2} & \mathbf{s}_{22^{N-1}}^{M-1} & \mathbf{s}_{22^{N-1}}^M \\ \mathbf{s}_{22^N}^1 & \mathbf{s}_{22^N}^2 & \mathbf{s}_{22^N}^3 & \mathbf{s}_{22^N}^4 & \mathbf{s}_{22^N}^5 & \cdots & \cdots & \mathbf{s}_{22^N}^{M-2} & \mathbf{s}_{22^N}^{M-1} & \mathbf{s}_{22^N}^M \end{bmatrix}_{(Nx3M)}, \tag{A2}$$

$$\mathbf{DFn}_{TS}=\begin{bmatrix} g(R_{i1}^1,r_1^1) & g(R_{i1}^2,r_1^2) & g(R_{i1}^3,r_1^3) & \cdots & g(R_{i1}^{M-1},r_1^{M-1}) & g(R_{i1}^M,r_1^M) \\ g(R_{i2}^1,r_2^1) & g(R_{i2}^2,r_2^2) & g(R_{i2}^3,r_2^3) & \cdots & g(R_{i2}^{M-1},r_2^{M-1}) & g(R_{i2}^M,r_2^M) \\ g(R_{i3}^1,r_3^1) & g(R_{i3}^2,r_3^2) & g(R_{i3}^3,r_3^3) & \cdots & g(R_{i3}^{M-1},r_3^{M-1}) & g(R_{i3}^M,r_3^M) \\ g(R_{i4}^1,r_4^1) & g(R_{i4}^2,r_4^2) & g(R_{i4}^3,r_4^3) & \cdots & g(R_{i4}^{M-1},r_4^{M-1}) & g(R_{i4}^M,r_4^M) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ g(R_{iN}^1,r_N^1) & g(R_{iN}^2,r_N^2) & g(R_{iN}^3,r_N^3) & \cdots & g(R_{iN}^{M-1},r_N^{M-1}) & g(R_{iN}^M,r_N^M) \end{bmatrix}_{(NxM)}, \tag{A3}$$

where

$$g(R_{in}^m,r_n^m)=\frac{1}{\rho\omega^2}\left[\left(\frac{1}{r_n^m}ik_fR_{2n}^me^{ik_fr_n^m}-\frac{e^{ik_fr_n^m}}{(r_n^m)^2}R_{2n}^m\right)n_2+\left(\frac{1}{r_n^m}ik_fR_{1n}^me^{ik_fr_n^m}-\frac{e^{ik_fr_n^m}}{(r_n^m)^2}R_{1n}^m\right)n_1\right],$$

$R_{in}^m=(x_{in}^m-y_{in}^m)/r_n^m$ and i take values 1, 2, and 3, except an imaginary quantity.

Ahmad, R., Kundu, T., and Placko, D. (**2003**). "Modeling of the ultrasonic field of two transducers immersed in a homogeneous fluid using distributed point source method," I2M (Instrumentation, Measurement and Metrology) Journal; Vol. **3**, pp. 87–116.

Ahmad, R., Kundu, T., and Placko, D. (**2005**). "Modeling of phased array transducers," J. Acoust. Soc. Am. **117**, 1762–1776.

Banerjee, S. (**2005**). "Elastic wave propagation in corrugated wave guides," PhD dissertation, University of Arizona, Tucson, AZ.

Banerjee, S., and Kundu, T. (**2004**). "Analysis of wave propagation in symmetrically periodic sinusoidal wave-guide," *Health Monitoring and Smart Nondestructive Evaluation of Structural and Biological Systems*, SPIE's 9th Annual International Symposium on NDE for Health Monitoring and Diagnostics, March 15–17, 2004, edited by T. Kundu, San Diego, CA, Vol. **5394**, pp. 89–98.

Banerjee, S., and Kundu, T. (2006a). "Symmetric and anti-symmetric Rayleigh-Lamb modes in sinusoidally corrugated waveguides: An analytical approach," Int. J. Solids Struct. (in press).
Banerjee, S., and Kundu, T. (2006b). "Ultrasonic field modelling in plates immersed in fluids," IEEE Trans. Ultrason. Ferroelectr. Freq. Control (submitted).
Banerjee, S., Kundu, T., and Placko, D. (2006). "Ultrasonic field modelling in multilayered fluid structures using DPSM technique," ASME J. Appl. Mech. (to be published).
Boström, A. (**1983**). "Passbands and stopbands for an electromagnetic waveguide with a periodically varying cross section," IEEE Trans. Microwave Theory Tech. **31**, 752–756.
Boström, A. (**1989**). "Propagating, damped, and leaky surface waves on the corrugated traction-free boundary of an elastic half-space," J. Acoust. Soc. Am. **85**, 1549–1555.
Declercq, N. F., Degrieck, J., Briers, R. and Leory, O. (**2005**). "Diffraction of homogeneous and inhomogeneous plane waves on a doubly corrugated liquid/solid interface," Ultrasonics **43**, 605–618.
El-Bahrawy, A. (**1994a**). "Stopbands and passbands for symmetric Rayleigh-Lamb modes in a plate with corrugated surfaces," J. Sound Vib. **170**(2), 145–160.
El-Bahrawy, A. (**1994b**). "Point force excitation of surface waves along the doubly corrugated traction-free boundary of an elastic half-space," Comm. Div. Mech. **2**.
Fokkemma, J. H. (**1980**). "Reflection and transmission of elastic waves by the spatially periodic interface between two solids (Theory of integral-equation method)," Wave Motion **2**, 375–393.
Glass, N. E., and Maradudin, A. A. (**1983**). "Leaky surface-elastic waves on both flat and strongly corrugated surfaces for isotropic, nondissipative media," J. Appl. Phys. **54**, 796–805.
Hah, Z. G., and Sung, K. M. (**1992**). "Effect of spatial sampling in the calculation of ultrasonic fields generated by piston transducers," J. Acoust. Soc. Am. **92**, 3403–3408.
Harris, G. R. (**1981**). "Review of transient field theory for a baffled planar piston," J. Acoust. Soc. Am. **70**, 10–20.
Ingenito, F., and Cook, B. D. (**1969**). "Theoretical investigation of the integrated optical effort produced by sound field radiated from plane piston transducers," J. Acoust. Soc. Am. **45**, 572–577.
Jensen, J. A., and Svendsen, N. B. (**1992**). "Calculation of pressure fields from arbitrary shaped, apodized, and excited ultrasound transducers," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 262–267.
Kundu, T. (**2004**). *Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization* (CRC Press, Boca Raton, FL), Chap. 2.
Lee, J. P., Placko, D., Alnuamaini, N., and Kundu, T. (**2002**). "Distributed point source method (DPSM) for modeling ultrasonic fields in homogeneous and non-homogeneous fluid media in presence of an interface," Ecole Normale Superieure de Cachan, France, *1st European Workshop on Structural Health Monitoring*, edited by D. L. Balageas (Pub. DEStech, PA), pp. 414–421.
Lerch, T. P., Schmerr, L. W., and Sedov, A. (**1998**). "Ultrasonic beam models: An edge element approach," J. Acoust. Soc. Am. **104**, 1256–1265.
Lockwood, J. C., and Willette, J. G. (**1973**). "High-speed method for computing the exact solution for the pressure variations in the near field of a baffled piston," J. Acoust. Soc. Am. **53**, 735–741.
Mal, A. K., and Singh, S. J. (**1991**). *Deformation of Elastic Solids* (Prentice–Hall, Englewood Cliffs, NJ).
Morse, P. M., and Ingard, U. K. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York).
Nayfeh, A. H., and Kandil, O. A. (**1978**). "Propagation waves in cylindrical hard-walled ducts with generally weak undulations," AIAA J. **16**, 1041–1045.
Newberry, B. P., and Thompson, R. B. (**1989**). "A paraxial theory for the propagation of ultrasonic beams in anisotropic solids," J. Acoust. Soc. Am. **85**, 2290–2300.
Placko, D., and Kundu, T. (**2001**). "A theoretical study of magnetic and ultrasonic sensors: Dependence of magnetic potential and acoustic pressure on the sensor geometry," *Advanced NDE for Structural and Biological Health Monitoring, Proceedings of SPIE*, SPIE's 6th Annual International Symposium on NDE for Health Monitoring and Diagnostics, edited by T. Kundu, 4–8 March, Newport Beach, California, Vol. **4335**, pp. 52–62.
Placko, D., and Kundu, T. (**2004**). "Modeling of ultrasonic field by distributed point source method," *Ultrasonic Nondestructive Evaluation: Engineering and Biological Material Characterization*, edited by T. Kundu (CRC Press, Boca Raton, FL), Chap. 2, pp. 144–201.
Placko, D., Kundu, T., and Ahmad, R. (**2002**). "Theoretical computation of acoustic pressure generated by ultrasonic sensors in presence of an interface," *Smart NDE and Health Monitoring of Structural and Biological Systems*, SPIE's 7th Annual International Symposium on NDE and Health Monitoring and Diagnostics, San Diego, CA, Vol. **4702**, pp. 157–168.
Placko, D., Kundu, T., and Ahmad, R. (**2003**). "Ultrasonic field computation in presence of a scatterer of finite dimension," *Smart NDE and Health Monitoring of Structural and Biological Systems*, SPIE's 8th Annual International Symposium on NDE and Health Monitoring and Diagnostics, San Diego, CA, Vol. **5047**, pp. 169–179.
Placko, D., Liebeaux, N., and Kundu, T. (2001). "Presentation d'une method generique pour la modelisation des capteurs de type ultrasons," Magnetiques at Electrostatiques, Instrumentation, Mesure, Metrologie (I2M Journal): Evaluation Nondestructive, Vol. **1**, pp. 101–125.
Rayleigh, L. (**1965**). *Theory of Sound* (Dover, New York), Vol. **II**, pp. 162–169.
Scarano, G., Denisenko, N., Matteucci, M., and Pappalardo, M. (**1985**). "A new approach to the derivation of the impulse response of a rectangular piston," J. Acoust. Soc. Am. **78**, 1109–1113.
Schmerr, L. W. (**1998**). *Fundamental of Ultrasonic Nondestructive Evaluation-A Modeling Approach* (Plenum, New York).
Schmerr, L. W. (**2000**). "A multi-Gaussian ultrasonic beam model for high performance simulations on a personal computer," Mater. Eval. 882–888.
Schmerr, L. W., Kim, H.-J., Huang, R., and Sedov, A. (**2003**). "Multi-Gaussian ultrasonic beam modeling," *Proceedings of the World Congress of Ultrasonics*, WCU 2003, Paris, 7–10 September, **2003**, pp. 93–99.
Sha, K., Yang, J., and Gan, W.-S. (**2003**). "A complex virtual source approach for calculating the diffraction beam field generated by a rectangular planar source," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 890–895.
Spies, M. (**1994**). "Transducer-modeling in general transversely isotropic media via point-source-synthesis theory," J. Nondestruct. Eval. **13**, 85–99.
Spies, M. (**1995**). "Elastic wave propagation in transversely isotropic media II: the generalized Rayleigh-function and an integral representation for the transducer field theory," J. Acoust. Soc. Am. **97**, 1–13.
Spies, M. (**1999**). 'transducer field modeling in anisotropic media by superposition of Gaussian base functions," J. Acoust. Soc. Am. **105**, 633–638.
Spies, M. (**2004**). "Analytical methods for modeling of ultrasonic nondestructive testing of anisotropic media," Ultrasonics **42**, 213–219.
Standström, S. E. (**1986**). "Stopbanda in a corrugated parallel plate waveguide," J. Acoust. Soc. Am. **79**, 1293–1298.
Standström, S. E. (**1987**). "A comparison of some techniques for corrugated parallel plate wave guides," J. Acoust. Soc. Am. **82**, 1797–1803.
Stepanishen, P. R. (**1971**). "Transient radiation from piston in an infinite planar baffle," J. Acoust. Soc. Am. **49**, 1627–1638.
Waterman, P. C. (**1975**). "Scattering by periodic surfaces," J. Acoust. Soc. Am. **57**, 791–802.
Wen, J. J., and Breazeale, M. A. (**1988**). "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756.
Wu, P., Kazys, R., and Stepinski, T. (**1995**). "Analysis of the numerically implemented angular spectrum approach based on the evaluation of two-dimensional acoustic fields. Part I. Errors due to the discrete Fourier transform and discretization," J. Acoust. Soc. Am. **99**, 1139–1148.

Curle's equation and acoustic scattering by a sphere

Anthony M. J. Davis
Mathematics Department, University of Alabama, Tuscaloosa, Alabama 35487-0350

Raymond J. Nagem[a)]
Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215

(Received 21 July 2005; revised 27 December 2005; accepted 29 December 2005)

Recent papers have initiated interesting comparisons between aeroacoustic theory and the results of acoustic scattering problems. In this paper, we consider some aspects of these comparisons for acoustic scattering by a sphere. We give a derivation of Curle's equation for a specific class of linear acoustic scattering problems, and, in response to previous claims to the contrary, give an explicit confirmation of Curle's equation for plane wave scattering by a stationary rigid sphere of arbitrary size in an inviscid fluid. We construct the complete solution for scattering by a rigid sphere in a viscous fluid, and show that the neglect of viscous terms in Curle's equation yields an incomplete prediction of the far field dipole pressure. We also consider the null field solution of the sphere scattering problem, and give its extension to the vorticity modes associated with viscosity. Finally, we construct a solution for an elastic sphere in a viscous fluid, and show that the rigid sphere/null field solution is recovered from the limit of infinite longitudinal and shear wave speeds in the elastic solid. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2167611]

PACS number(s): 43.20.Fn, 43.28.Ra [TDM] Pages: 2018–2026

I. INTRODUCTION

In a recent paper, titled "On acoustic radiation by a rigid object in a fluid flow," Zinoviev and Bies[1] examine the validity of Curle's equation[2] and its generalization to the Ffowcs Williams and Hawkings equation.[3] Acoustic scattering by a sphere is considered as part of the discussion of Curle's equation in Ref. 1, and also in a subsequent paper.[4] In this paper, we present a more detailed and exhaustive discussion of acoustic scattering by a sphere, along the lines of our previous work on the effects of viscosity in acoustic scattering problems.[5–7]

Curle's equation is based on Lighthill's concise formula,[8] in the absence of rigid bodies, for acoustic radiation. Lighthill's result, Eq. (3) of Ref. 1, assumes that the density perturbation ρ' is generated by a compact disturbance and therefore has only outgoing waves at infinity. Subject to this same radiation condition, Eq. (4) of Ref. 1, namely,

$$\rho'(\mathbf{x},t) = \frac{1}{4\pi c_0^2} \times \int\int\int_{V_{tot}} \frac{\partial^2[T_{ij}]}{\partial y_i \partial y_j}\frac{d\mathbf{y}}{r} + \frac{1}{4\pi}\int\int_S \left(\frac{1}{r}\frac{\partial[\rho']}{\partial n} + \frac{1}{r^2}\frac{\partial r}{\partial n}[\rho'] + \frac{1}{c_0 r}\frac{\partial r}{\partial n}\frac{\partial[\rho']}{\partial t}\right) dS(\mathbf{y}), \quad (1)$$

where the square brackets denote evaluation at the retarded time $t - r/c_0$, $r = |\mathbf{x}-\mathbf{y}|$ and $\mathbf{n}$ is the normal directed out of the fluid into the solid boundary S, can be obtained by Fourier transform and Green's function techniques, with $\partial^2[T_{ij}]/\partial y_i \partial y_j$ regarded as a given function.

The derivation of Curle's equation from Eq. (1) is not as elementary as claimed by Farassat[9] because the presence of S and its associated boundary conditions precludes the transfer of the integrand's y_i derivatives to x_i derivatives of the volume integral. It is indeed necessary to repeatedly apply the divergence theorem, a major issue raised in Ref. 1, to transform the right-hand side of Eq. (1). To make this issue definite in the context of the scattering problems that we consider here, we derive, in Sec. II, Curle's equation for linear time-harmonic acoustics, in which the time factor $e^{-i\omega t}$ is assumed and suppressed. This derivation is used to show, in detail, the difference between our treatment of the divergence theorem and that presented in Ref. 1. A more general derivation of Curle's equation can be based on the extension to moving surfaces by Ffowcs Williams and Hawkings in Ref. 3. Curle's equation may also be derived by using the theory of generalized functions, as in Ref. 10; we use more traditional methods here so that our work can be compared directly to that of Ref. 1. In Sec. III, we consider the example of plane wave scattering by a rigid sphere in an inviscid fluid and show, without assuming a small scatterer, that Curle's equation is satisfied. We also consider a second example, suggested in Ref. 4, of spherical wave scattering by a rigid sphere in an inviscid fluid.

In Sec. IV, we construct the field scattered by a rigid sphere in a viscous fluid. A complete verification of Curle's equation in this case is quite complicated. Indeed, Curle's equation is designed to show structure and provide order of magnitude estimates in cases where exact solutions are not easily obtained. Comparison with a previous result in Ref. 11 for an acoustically small sphere shows that the neglect of viscous terms in Curle's equation yields far field amplitudes that lack the viscous decay factor.

[a)]Corresponding author. Electronic mail: nagem@bu.edu

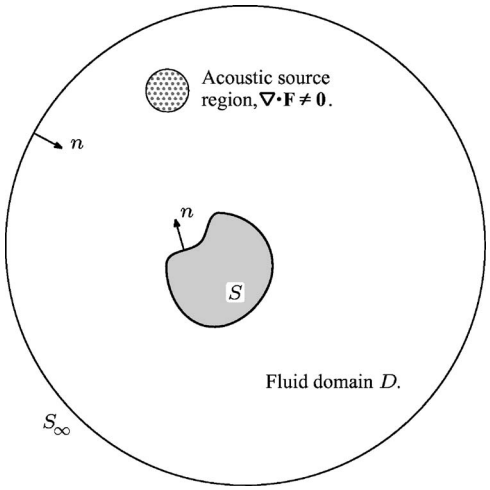


FIG. 1. Geometric configuration for the derivation of the linearized Curle's equation.

In Sec. V, we describe the null field ($T_{ij}=0$ within S) approach, suggested in Sec. 3.5 of Ref. 1, and indicate its extension to the vorticity modes associated with viscosity. In Sec. VI, we use the techniques of Sec. IV to construct a solution for an elastic sphere and show that the rigid sphere/null field solution is recovered in the limit of infinite longitudinal and shear wave speeds.

We do not consider here the entire range of issues, discussed in Refs. 4 and 9, associated with the use of Curle's equation for aeroacoustic applications. In particular, we limit ourselves here to the case of a stationary scattter. We believe, however, that our results in Secs. IV and VI are of value as a continuation of the work on viscous scattering presented in Refs. 5–7, and that our paper helps to clarify the relation between aeroacoustic theory and linear scattering problems in some specific cases.

II. CURLE'S EQUATION FOR TIME-HARMONIC ACOUSTICS

We consider the configuration shown in Fig. 1. A viscous fluid occupies the domain D, which is bounded by an arbitrary closed surface S and a large sphere S_∞. As in our earlier work,[5–7] the governing acoustic equations for the fluid are

$$-i\omega\rho + \rho_0 \nabla \cdot \mathbf{v} = 0, \tag{2}$$

$$-i\omega\rho_0 v_i = -\frac{\partial p}{\partial x_i} + \frac{\partial \sigma_{ij}}{\partial x_j} + F_i \tag{3}$$

and

$$p = c_0^2 \rho, \tag{4}$$

where ρ is the acoustic density perturbation, ρ_0 is the constant background density, p is the acoustic pressure perturbation, v_i is the fluid velocity vector, and c_0 is the sound speed. The viscous stress tensor σ_{ij} is given, for a Stokesian fluid, by[11]

$$\sigma_{ij} = \rho_0 \nu \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} - \frac{2}{3}\delta_{ij}\frac{\partial v_k}{\partial x_k} \right), \tag{5}$$

where ν is the kinematic viscosity. The source term F_i in Eq. (3) is an external body force per unit volume applied to the fluid.

Equations (2)–(5) imply the Helmholtz equations

$$(\nabla^2 + k_a^2)(\rho, \nabla \cdot \mathbf{v}) = \left(\frac{k_a^2}{\omega^2}, \frac{ik_a^2}{\omega\rho_0} \right) \nabla \cdot \mathbf{F}, \tag{6}$$

and

$$\left(\nabla^2 + \frac{i\omega}{\nu} \right) \nabla \times \mathbf{v} = -\frac{1}{\rho_0 \nu} \nabla \times \mathbf{F}, \tag{7}$$

where

$$k_a^{-2} = k_0^{-2} - i\frac{4\nu}{3\omega}, \quad k_0 = \omega/c_0. \tag{8}$$

But the equation for ρ may also be written in the form

$$(\nabla^2 + k_0^2)\rho = \frac{1}{c_0^2}\frac{\partial}{\partial x_i}\left(F_i + \frac{\partial \sigma_{ij}}{\partial x_j} \right), \tag{9}$$

where, following the Lighthill approach, the viscous stresses are grouped with the external body force. When Green's theorem is applied in the domain D to the density ρ and the Green's function $G(\mathbf{x}-\mathbf{y})$, given by

$$G(\mathbf{x}-\mathbf{y}) = \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|},$$

it follows that

$$\rho(\mathbf{x}) = -\int_D G(\mathbf{x}-\mathbf{y})\frac{1}{c_0^2}\left(\frac{\partial F_i(\mathbf{y})}{\partial y_i} + \frac{\partial^2 \sigma_{ij}(\mathbf{y})}{\partial y_i \partial y_j} \right) d^3\mathbf{y} - \int_{S+S_\infty} \left(G(\mathbf{x}-\mathbf{y})\frac{\partial \rho(\mathbf{y})}{\partial y_i} - \rho(\mathbf{y})\frac{\partial G(\mathbf{x}-\mathbf{y})}{\partial y_i} \right) dA_i(\mathbf{y}), \tag{10}$$

where $\mathbf{x}$ is a point in the fluid domain D.

As in Ref. 11, but in contrast to Eq. (1), the directed area element $dA_i = n_i dA$ points into the fluid region D. Although this expresses the divergence theorem in the less conventional form

$$\text{volume integral of } \nabla \cdot \mathbf{f} = -\text{ flux of } \mathbf{f} \text{ into the volume},$$

it does ensure that the vector field in the surface integral is the value of $\mathbf{f}$ obtained by taking a limit from the fluid domain D. This is a fundamental difference between our treatment and that of Ref. 1, made plain by the irrelevance here of any discontinuites in $\mathbf{f}$ across the bounding surface or nonexistence of $\mathbf{f}$ outside D.

To derive Curle's equation from Eq. (10), we first multiply by c_0^2, transfer the y_i derivative in the last integral to x_i, and substitute Eqs. (3) and (4). This gives

$$c_0^2\rho(\mathbf{x}) = -\int_D G(\mathbf{x}-\mathbf{y})\frac{\partial F_i(\mathbf{y})}{\partial y_i}d^3\mathbf{y} - \int_D G(\mathbf{x}-\mathbf{y})\frac{\partial^2\sigma_{ij}(\mathbf{y})}{\partial y_i\partial y_j}d^3\mathbf{y}$$
$$- \int_{S+S_\infty} G(\mathbf{x}-\mathbf{y})\frac{\partial\sigma_{ij}(\mathbf{y})}{\partial y_j}dA_i(\mathbf{y}) - \int_{S+S_\infty} G(\mathbf{x}-\mathbf{y})$$
$$\times i\omega\rho_0 v_i(\mathbf{y})dA_i(\mathbf{y}) - \int_{S+S_\infty} G(\mathbf{x}-\mathbf{y})F_i dA_i(\mathbf{y})$$
$$- \frac{\partial}{\partial x_i}\int_{S+S_\infty} G(\mathbf{x}-\mathbf{y})p(\mathbf{y})dA_i(\mathbf{y}). \quad (11)$$

The second and third integrals in Eq. (11) can be rewritten as

$$-\int_D \frac{\partial}{\partial y_i}\left[G(\mathbf{x}-\mathbf{y})\frac{\partial\sigma_{ij}(\mathbf{y})}{\partial y_j}\right]d^3\mathbf{y} - \frac{\partial}{\partial x_i}\int_D G(\mathbf{x}-\mathbf{y})\frac{\partial\sigma_{ij}(\mathbf{y})}{\partial y_j}d^3\mathbf{y}$$
$$-\int_{S+S_\infty} G(\mathbf{x}-\mathbf{y})\frac{\partial\sigma_{ij}(\mathbf{y})}{\partial y_j}dA_i(\mathbf{y}), \quad (12)$$

in which the first and last terms are identified as − (volume integral of a divergence + flux of the field into D) and therefore their net contribution is zero. A similar rearrangement, analogous to integration by parts, of the middle term of Eq. (12) yields

$$= -\frac{\partial}{\partial x_i}\left\{\int_D \frac{\partial}{\partial y_j}[G(\mathbf{x}-\mathbf{y})\sigma_{ij}(\mathbf{y})]d^3\mathbf{y} + \frac{\partial}{\partial x_j}\int_D G(\mathbf{x}-\mathbf{y})\sigma_{ij}(\mathbf{y})d^3\mathbf{y}\right\}$$
$$= \frac{\partial}{\partial x_i}\int_{S+S_\infty} G(\mathbf{x}-\mathbf{y})\sigma_{ij}(\mathbf{y})dA_j(\mathbf{y}) - \frac{\partial^2}{\partial x_i\partial x_j}\int_D G(\mathbf{x}-\mathbf{y})\sigma_{ij}(\mathbf{y})d^3\mathbf{y}.$$

The surface integrals on the large sphere S_∞ vanish due to the exponential decay associated with propagation in the viscous fluid. In the inviscid limit they vanish because G and the field variables satisfy the same radiation condition. Moreover, we assume, as illustrated in Fig. 1, that the acoustic source $\nabla\cdot\mathbf{F}$ for ρ in Eq. (9) is nonzero only in a region away from S, and hence Eq. (11) becomes

$$c_0^2\rho(\mathbf{x}) = -\int_D \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}\frac{\partial F_i(\mathbf{y})}{\partial y_i}d^3\mathbf{y} - \frac{\partial^2}{\partial x_i\partial x_j}$$
$$\times\int_D \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}\sigma_{ij}(\mathbf{y})d^3\mathbf{y} - i\omega\rho_0$$
$$\times\int_S v_i(\mathbf{y})\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}dA_i(\mathbf{y}) + \frac{\partial}{\partial x_i}$$
$$\times\int_S \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}[\sigma_{ij}(\mathbf{y}) - \delta_{ij}p(\mathbf{y})]dA_j(\mathbf{y}). \quad (13)$$

Equation (13) is Curle's equation [Eq. (2.3.9) of Ref. 11] for time-harmonic acoustic fields, with the linearization of the integrands of the volume and surface integrals.

The necessity of the above manipulations involving the divergence theorem is illustrated by the example

$$\nabla^2 V = (x_1^2 + x_2^2 + x_3^2 + a^2)^{-2} \quad (x_3 > 0),$$

subject to $V=0$ at $x_3=0$. The solution is

$$V = \frac{1}{4\pi}\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty \{[(x_1-y_1)^2 + (x_2-y_2)^2 + (x_3+y_3)^2]^{-1/2} - [(x_1-y_1)^2 + (x_2-y_2)^2 + (x_3-y_3)^2]^{-1/2}\}\frac{dy_1dy_2dy_3}{(y_1^2+y_2^2+y_3^2+a^2)^2}.$$

Evidently, x_1 and x_2 derivatives of V have a similar form with y_1 and y_2 derivatives, respectively, of $(y_1^2+y_2^2+y_3^2+a^2)^{-2}$ but the boundary at $x_3=0$ precludes the corresponding interchange of x_3- and y_3-derivatives.

In the solution of scattering problems, it is customary to decompose the velocity, density, and pressure into incident and scattered fields. The incident fields are the solutions of Eqs. (2)–(5) due to the external source F_i in the absence of the scattering surface S. The scattered fields are defined as the total fields, due to both the external source and presence of the scattering surface, minus the incident fields. Since the scattered fields satisfy Eqs. (2)–(5) with $F_i=0$, they satisfy Curle's equation in the form

$$c_0^2\rho_{sc}(\mathbf{x}) = -\frac{\partial^2}{\partial x_i\partial x_j}\int_D (\sigma_{sc})_{ij}(\mathbf{y})\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}d^3\mathbf{y} - i\omega\rho_0$$
$$\times\int_S (v_{sc})_i(\mathbf{y})\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}dA_i(\mathbf{y}) + \frac{\partial}{\partial x_i}\int_S \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}$$
$$\times[(\sigma_{sc})_{ij}(\mathbf{y}) - \delta_{ij}p_{sc}(\mathbf{y})]dA_j(\mathbf{y}). \quad (14)$$

Since Eqs. (13) and (14) are derived directly from the fluid Eqs. (2)–(5), any solution of a scattering problem based on Eqs. (2)–(5) and Fig. 1 must necessarily satisfy Eqs. (13) and (14) as well. In the following sections we discuss scattering by a rigid sphere in an inviscid fluid, scattering by a rigid sphere in a viscous fluid and scattering by an elastic sphere in a viscous fluid.

III. INVISCID SCATTERING BY A RIGID SPHERE

Suppose that the surface S in Fig. 1 is a stationary rigid sphere of radius R at the origin of a coordinate system $\mathbf{x}=(x,y,z)$, that the fluid in the domain D is inviscid, and that the external source F_i produces a field that can be approximated, in the vicinity of the sphere, by the plane wave

$$\rho_{inc}(\mathbf{x}) = -\frac{1}{c_0^2}\int_D \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}\frac{\partial F_i(\mathbf{y})}{\partial y_i}d^3\mathbf{y} = \alpha\rho_0 e^{ik_0 z}, \quad \alpha \ll 1. \quad (15)$$

Then the scattered field ρ_{sc} satisfies the Helmholz equation with wave number k_0 and zero source, has only outgoing waves at infinity and ensures that the total field, $\rho=\rho_{inc}+\rho_{sc}$, has zero normal derivative at $r=\sqrt{x^2+y^2+z^2}=R$. Since $p=\rho c_0^2$ and $\nabla p = i\omega\rho_0\mathbf{v}$, the last condition ensures no normal flow at the sphere.

If $k_0R \ll 1$, the sphere is acoustically small and its local field is essentially static. Thus, on writing

$$\frac{\rho_{\text{inc}}}{\alpha\rho_0} \sim 1 + ik_0z - \frac{1}{2}k_0^2z^2$$

$$= 1 + ik_0r\cos\theta - \frac{k_0^2r^2}{6}[1 + 2P_2(\cos\theta)],$$

in terms of spherical polar coordinates, it follows from the zero flux condition that the scattered field is approximated locally by

$$\frac{\rho_{\text{sc}}}{\alpha\rho_0} \sim \frac{ik_0R^3}{2r^2}\cos\theta - \frac{k_0^2R^3}{3r},$$

in which the last term is a monopole. The associated outward flux $\frac{4}{3}\pi k_0^2R^3$ of $\nabla(\rho_{\text{sc}}/\alpha\rho_0)$ yields, in this inviscid case, an outward volume flux $\frac{4}{3}\pi\alpha_0(-i\omega)R^3$. The corresponding term in the two-dimensional (2-D) case has the factor $\ln(k_0R)$ in the amplitude, as shown, for example, by Davis and Leppington.[12]

For arbitrary k_0R, the Rayleigh equation (Ref. 13, Eq. 12.4.7)

$$e^{ik_0r\cos\theta} = \sum_{n=0}^{\infty}(2n+1)i^nj_n(k_0r)P_n(\cos\theta), \tag{16}$$

in terms of spherical Bessel functions and Legendre polynomials, yields

$$\frac{\rho_{\text{sc}}}{\alpha\rho_0} = -\sum_{n=0}^{\infty}(2n+1)i^n\frac{j_n'(k_0R)}{h_n^{(1)\prime}(k_0R)}h_n^{(1)}(k_0r)P_n(\cos\theta). \tag{17}$$

Here the first term is a monopole with outward flux

$$4\pi\frac{j_0'(k_0R)}{h_0^{(1)\prime}(k_0R)}\lim_{r\to 0} rh_0^{(1)}(k_0r) = \frac{4\pi j_0'(k_0R)}{ik_0h_0^{(1)\prime}(k_0R)},$$

since $h_0^{(1)}(x) = e^{ix}/ix$. Agreement with the small sphere estimate is confirmed by noting that this flux $\sim\frac{4}{3}\pi k_0^2R^3 + O(k_0^4R^5)$.

In the absence of viscous stresses, $\mathbf{v}$ is a gradient field and Curle's Eq. (14) for the scattered fields reduces to

$$\rho_{\text{sc}}(\mathbf{x}) = -\int_S \frac{\partial\rho_{\text{sc}}}{\partial y_j}(\mathbf{y})\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}dA_j(\mathbf{y})$$

$$-\frac{\partial}{\partial x_i}\int_S \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}\rho_{\text{sc}}(\mathbf{y})dA_i(\mathbf{y}). \tag{18}$$

For the plane wave $\rho_{\text{inc}} = \alpha\rho_0e^{ik_0z}$ incident on a rigid sphere of radius R, the scattered density field is given by Eq. (17). On introducing spherical polar coordinates $\mathbf{x} = (r_x, \theta_x, \phi_x)$, $\mathbf{y} = (r_y, \theta_y, \phi_y)$, as in Ref. 1, the kernel function has the expansion[14]

$$\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|} = ik_0\sum_{n=0}^{\infty}(2n+1)j_n(k_0r_<)h_n^{(1)}(k_0r_>)P_n(\cos\gamma), \tag{19}$$

where $r_<, r_> = \min, \max(r_x, r_y)$ and the addition theorem (Ref. 13, Eq. 12.198) states that

$$P_n(\cos\gamma) = \sum_{m=0}^{n}\epsilon_m\frac{(n-m)!}{(n+m)!}P_n^m(\cos\theta_x)$$

$$\times P_n^m(\cos\theta_y)\cos m(\phi_x - \phi_y), \tag{20}$$

in which $\cos\gamma = \cos\theta_x\cos\theta_y + \sin\theta_x\sin\theta_y\cos m(\phi_x - \phi_y)$ and ϵ_m is Neumann's symbol. But, since ρ_{sc} and its normal derivative at $r=R$ are axisymmetric, only the $m=0$ term in Eq. (20) contributes to Eq. (18), i.e., the m summation effectively reduces to $P_n(\cos\theta_x)P_n(\cos\theta_y)$. Similarly the integral operator simplifies as

$$\frac{1}{4\pi}\int_S dA_i(\mathbf{y}) = \frac{R^2}{2}n_i\int_0^{\pi}\sin\theta_y d\theta_y.$$

Consequently, the substitution of Eqs. (17) and (19) into the right-hand side of Eq. (18) yields

$$\alpha\rho_0\sum_{n=0}^{\infty}(2n+1)i^{n+1}(k_0R)^2\left[j_n(k_0R) - \frac{j_n'(k_0R)}{h_n^{(1)\prime}(k_0R)}h_n^{(1)}(k_0R)\right]$$

$$\times j_n'(k_0R)h_n^{(1)}(k_0r_x)P_n(\cos\theta_x)$$

$$= -\alpha\rho_0\sum_{n=0}^{\infty}(2n+1)i^n\frac{j_n'(k_0R)}{h_n^{(1)\prime}(k_0R)}h_n^{(1)}(k_0r_x)P_n(\cos\theta_x)$$

$$= \rho_{\text{sc}},$$

after using the Wronskian $j_n(Z)h_n^{(1)}(Z) - j_n'(Z)h_n^{(1)}(Z) = i/Z^2$ and comparing the result with Eq. (17). Thus Eq. (18) is satisfied by this flow.

Since the total fields are the sum of the incident and the scattered, Curle's Eq. (13) for the total fields reduces, after using Eq. (15), to

$$0 = -i\omega\rho_0\int_S (v_{\text{inc}})_i(\mathbf{y})\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}dA_i(\mathbf{y})$$

$$-\frac{\partial}{\partial x_i}\int_S \frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}p_{\text{inc}}(\mathbf{y})dA_i(\mathbf{y}),$$

or, using Eqs. (3) and (4),

$$0 = \int_S\left[-\frac{\partial}{\partial y_i}\rho_{\text{inc}}(\mathbf{y})\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|} + \rho_{\text{inc}}(\mathbf{y})\frac{\partial}{\partial y_i}\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}\right]dA_i(\mathbf{y}). \tag{21}$$

Since ρ_{inc} and $e^{ik_0|\mathbf{x}-\mathbf{y}|}/|\mathbf{x}-\mathbf{y}|$ can be continued within the sphere where they satisfy the same source-free Helmholtz equation, an application of Green's theorem to these functions within the sphere shows that Eq. (21) is satisfied; this can also be verified by direct calculation using the expansions (16) and (19). Thus Curle's Eq. (13) for the total fields, equivalent to Eq. (37) of Ref. 1, is also valid, and its correct evaluation includes the monopole term, contrary to the claim made in Ref. 1 in the small-sphere approximation. The monopole is made necessary by the nonzero mean velocity at $r=R$ in the scattered field. This inviscid feature is unfortunately misleading because it deflects attention from the emphasis on radiated fields generated by viscous stresses.

In Ref. 4, Zinoviev and Bies consider the case in which the incident field is given by the converging spherical wave

$$\rho_{\text{inc}}(\mathbf{x}) = \frac{A}{c_0^2}\frac{e^{-ik_0 r}}{r}, \tag{22}$$

where A is a constant and r is the distance from the center of the sphere. The scattered field is then easily obtained as the diverging spherical wave

$$\rho_{\text{sc}}(\mathbf{x}) = \frac{B}{c_0^2}\frac{e^{ik_0 r}}{r}, \tag{23}$$

where

$$B = Ae^{-2ik_0 R}\frac{ik_0 R + 1}{ik_0 R - 1}$$

and R is the radius of the sphere.

The scattered field ρ_{sc} in Eq. (23) satisfies Curle's Eq. (18), as can be shown easily by using Eqs. (19) and (20). However, the incident field ρ_{inc} in Eq. (22) does not satisfy Eq. (21), since ρ_{inc} does not satisfy a source-free Helmholtz equation inside the sphere S. According to Eq. (9), the source $\nabla\cdot\mathbf{F}$ for the field ρ_{inc} is

$$\nabla\cdot\mathbf{F} = c_0^2(\nabla^2 + k_0^2)\left(\frac{A}{c_0^2}\frac{e^{-ik_0 r}}{r}\right) = -4\pi A\,\delta(\mathbf{r}),$$

which implies a point singularity (in this case, a sink for the converging waves) in ρ_{inc} at the center of the sphere S. This example of spherical wave scattering does not, therefore, belong to the class of problems depicted in Fig. 1 and considered here, in which the source $\nabla\cdot\mathbf{F}$ for the field ρ_{inc} is nonzero only in a region outside the surface S. Since Eq. (21) is not satisfied, Curle's Eq. (13) for the total fields is likewise not satisfied.

To illustrate the application of Curle's equation in the presence of viscous stresses, we sought an example that requires computation of at least one nontrivial viscous stress integral in Eq. (13). Our simplest has an inviscid plane wave incident on a small sphere outside a spherical body of viscous fluid. The dominant scattered field is generated by the viscous sphere and can be constructed by the method used in the next section. However, it is expressed in spherical polar coordinates which causes major difficulties in evaluating the Cartesian components of the stress integral over D in Eq. (13). Though the second of the $\mathbf{x}$ derivatives can be arranged as a divergence and converted to spherical polar form, the total algebra is essentially unmanageable, which invites the following observation. By separating the derivatives from the stress components, Curle's equation provides structure and order of magnitude estimates at the expense of direct evaluations. In the next section, we illustrate this observation by calculating both the scatterred field and the force exerted by a plane wave on a rigid sphere in a viscous medium. For an acoustically small scatterer, our dipole term in the scattered pressure is compared to the force-dependent dipole prediction obtained from Curle's equation when viscous terms are neglected.

IV. VISCOUS SCATTERING BY A RIGID SPHERE

In the presence of viscous stresses, $\mathbf{v}$ has gradient and solenoidal field components, associated, respectively, with the dilatation $\nabla\cdot\mathbf{v}$ and the vorticity $\nabla\times\mathbf{v}$. The governing equations are given by Eqs. (6)–(8), with $\mathbf{F}=0$ away from the source region.

Suppose that the incident field is the damped acoustic plane wave $\rho_{\text{inc}} = \alpha\rho_0 e^{ik_a z}$, which is a gradient field. Then

$$\begin{aligned}\mathbf{v}_{\text{inc}} &= \frac{\alpha\omega}{ik_a^2}\nabla\,(e^{ik_a z}) \\ &= \frac{\alpha\omega}{ik_a^2}\nabla\left[\sum_{n=0}^{\infty}(2n+1)i^n j_n(k_a r)P_n(\cos\theta)\right],\end{aligned} \tag{24}$$

according to Eq. (16). The no slip conditions on the sphere therefore require

$$(\mathbf{v}_{\text{sc}}\cdot\mathbf{e}_r)_{r=R} = \frac{\alpha\omega}{k_a}\sum_{n=0}^{\infty}(2n+1)i^{n+1}j_n'(k_a R)P_n(\cos\theta), \tag{25}$$

$$\begin{aligned}(\mathbf{v}_{\text{sc}}\cdot\mathbf{e}_\theta)_{r=R} = &-\frac{\alpha\omega}{k_a^2 R}\sum_{n=1}^{\infty}(2n+1)i^{n+1}j_n(k_a R) \\ &\times P_n'(\cos\theta)\sin\theta,\end{aligned} \tag{26}$$

where $\mathbf{e}_r, \mathbf{e}_\theta$ denote radially and tangentially directed unit vectors. By writing suitable solutions of Eqs. (6) and (7) for the dilatation $\nabla\cdot\mathbf{v}_{\text{sc}}$ and the vorticity $\nabla\times\mathbf{v}_{\text{sc}}$, respectively, the associated gradient and solenoidal field components of the scattered velocity can be readily identified. Thus, in

$$\begin{aligned}\nabla\cdot\mathbf{v}_{\text{sc}} = &-\alpha\omega i\frac{j_0'(k_a R)}{h_0^{(1)\prime}(k_a R)}h_0^{(1)}(k_a r) - \alpha\omega\sum_{n=1}^{\infty}(2n+1)i^{n+1} \\ &\times\left[j_n'(k_a R) - \frac{iC_n}{k_a R h_n^{(1)\prime}(k_a R)}\right]\frac{h_n^{(1)}(k_a r)}{h_n^{(1)\prime}(k_a R)}P_n(\cos\theta),\end{aligned} \tag{27}$$

the differential equation for $h_n^{(1)}$ (Ref. 13, Eq. 11.139) allows us to write

$$\begin{aligned}&-h_n^{(1)}(k_a r)P_n(\cos\theta) \\ &\quad= \left[h_n^{(1)\prime\prime}(k_a r) + \frac{2}{k_a r}h_n^{(1)\prime}(k_a r) - \frac{n(n+1)}{(k_a r)^2}h_n^{(1)}(k_a r)\right] \\ &\qquad\times P_n(\cos\theta) = \frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2\frac{1}{k_a}h_n^{(1)\prime}(k_a r)P_n(\cos\theta)\right] \\ &\qquad+ \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\left[\sin\theta\frac{1}{k_a^2 r}h_n^{(1)}(k_a r)P_n'(\cos\theta)\sin\theta\right].\end{aligned}$$

Similarly, the decomposition

$$
\begin{aligned}
&- i\omega\nu h_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)P_n'(\cos\theta)\sin\theta \\
&\quad = \frac{1}{r}\frac{\partial}{\partial r}\left[r\frac{1}{r}\frac{\partial}{\partial r}\left\{rh_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)P_n'(\cos\theta)\sin\theta\right\}\right] \\
&\qquad + \frac{1}{r}\frac{\partial}{\partial\theta}\left[\frac{n(n+1)}{r}h_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)P_n(\cos\theta)\right]
\end{aligned}
$$

can be used in

$$
\nabla\times\mathbf{v}_{\mathrm{sc}} = -\frac{\alpha\omega^2}{k_a^4R^2\nu}\sum_{n=1}^{\infty}\frac{(2n+1)i^{n+1}}{h_n^{(1)\prime}(k_aR)}C_n\frac{h_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)}{h_n^{(1)}\left(R\sqrt{\frac{i\omega}{\nu}}\right)} \times P_n'(\cos\theta)\sin\theta\mathbf{e}_\phi. \tag{28}
$$

Hence

$$
\begin{aligned}
\mathbf{v}_{\mathrm{sc}} = {}& \frac{\alpha\omega i}{k_a}\frac{j_0'(k_aR)}{h_0^{(1)\prime}(k_aR)}h_0^{(1)\prime}(k_ar)\mathbf{e}_r + \frac{\alpha\omega}{k_a}\sum_{n=1}^{\infty}\frac{(2n+1)i^{n+1}}{h_n^{(1)\prime}(k_aR)}\left\{j_n'(k_aR)h_n^{(1)\prime}(k_ar) + \frac{in(n+1)C_n}{k_a^3R^2}\left[\frac{h_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)}{rh_n^{(1)}\left(R\sqrt{\frac{i\omega}{\nu}}\right)}\right.\right. \\
&\left.\left. - \frac{h_n^{(1)\prime}(k_ar)}{Rh_n^{(1)\prime}(k_aR)}\right]\right\}P_n(\cos\theta)\mathbf{e}_r - \frac{\alpha\omega}{k_a^2r}\sum_{n=1}^{\infty}\frac{(2n+1)i^{n+1}}{h_0^{(1)\prime}(k_aR)}\left\{j_n'(k_aR)h_n^{(1)}(k_ar) + \frac{iC_n}{k_a^2R^2}\left[\frac{\frac{d}{dr}\left[rh_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)\right]}{h_n^{(1)}\left(R\sqrt{\frac{i\omega}{\nu}}\right)}\right.\right. \\
&\left.\left. - \frac{n(n+1)h_n^{(1)}(k_ar)}{k_aRh_n^{(1)\prime}(k_aR)}\right]\right\}P_n'(\cos\theta)\sin\theta\mathbf{e}_\theta,
\end{aligned} \tag{29}
$$

which has the prescribed normal velocity (25). Condition (26) is satisfied by setting

$$
C_n = \left[\frac{\frac{d}{dR}\left[Rh_n^{(1)}\left(R\sqrt{\frac{i\omega}{\nu}}\right)\right]}{h_n^{(1)}\left(R\sqrt{\frac{i\omega}{\nu}}\right)} - \frac{n(n+1)h_n^{(1)}(k_aR)}{k_aRh_n^{(1)\prime}(k_aR)}\right]^{-1} \quad (n\geqslant 1). \tag{30}
$$

Force exerted by the fluid on the sphere

Because the sphere is stationary, the normal and tangential components of the fluid stress at $r=R$ are multiples of the dilatation and vorticity, respectively. On substitution of Eq. (27), these are

$$
-\frac{\alpha\rho_0\omega^2}{k_a^4R^2}\left\{\sum_{n=0}^{\infty}\frac{(2n+1)i^{n+1}}{h_n^{(1)\prime}(k_aR)}P_n(\cos\theta)\mathbf{e}_r + \sum_{n=1}^{\infty}\frac{(2n+1)i^{n+1}C_n}{h_n^{(1)\prime}(k_aR)} \times\left[\frac{n(n+1)h_n^{(1)}(k_aR)}{k_aRh_n^{(1)\prime}(k_aR)}P_n(\cos\theta)\mathbf{e}_r + P_n'(\cos\theta)\sin\theta\mathbf{e}_\theta\right]\right\}.
$$

Evidently, only the $n=1$ terms contribute to the force $-\mathbf{F}e^{-i\omega t}$ on the sphere and it follows that

$$
\mathbf{F} = -\frac{4\pi\alpha\rho_0\omega^2}{k_a^4h_1^{(1)\prime}(k_aR)}\left\{1 + 2C_1\left[\frac{h_1^{(1)}(k_aR)}{k_aRh_1^{(1)\prime}(k_aR)} - 1\right]\right\}\mathbf{e}_z. \tag{31}
$$

Consider the compact body limit, $|k_aR|\ll 1$, in which the approximations

$$
j_n(x) \sim \frac{\Gamma(3/2)}{\Gamma(n+3/2)}\left(\frac{x}{2}\right)^n,
$$

$$
h_n^{(1)}(x) \sim \frac{\Gamma(n+1/2)2^n}{i\Gamma(1/2)x^{n+1}} \quad (|x|\ll 1, n\geqslant 0),
$$

(Ref. 13, Eqs. 11.144, 11.152) yield

$$
\mathbf{F} \sim 2\pi i\alpha\rho_0\frac{\omega^2R^3}{k_a}(1-3C_1)\mathbf{e}_z,
$$

$$
\begin{aligned}
p_s &= \frac{c_0^2\rho_0}{i\omega}\nabla\cdot\mathbf{v}_{\mathrm{sc}} \\
&\sim -\alpha c_0^2\rho_0(k_aR)^3\left[\frac{i}{3}h_0^{(1)}(k_ar) + \frac{1}{2}(1-3C_1)h_1^{(1)}(k_ar)\cos\theta\right].
\end{aligned}
$$

Thus the scattered pressure field is dominated by the monopole and dipole contributions. In the far field, the latter is evidently expressible as

$$p_d \simeq \frac{ze^{i\omega t}}{4\pi c_0 r^2}\left(\frac{k_a}{k_0}\right)^3\left[\frac{\partial}{\partial t}(F_z e^{-i\omega t})\right]_{t-k_a r/\omega}, \tag{32}$$

since $h_1^{(1)}(k_a r) \sim -e^{ik_a r}/k_a r$ at large $k_a r$. By arguing that the viscous terms can be neglected, Howe [Ref. 11, Eq. (2.4.2)] deduced from Curle's equation the above estimate, but with k_a set equal to k_0, of the dipole pressure p_d in the far field in terms of the force exerted by an acoustically compact body in the fluid. The inviscid limit of Eq. (31) yields

$$p_d \simeq \frac{\alpha\rho_0\omega^2 i}{k_0^3 h_1^{(1)\prime}(k_0 R)}\frac{e^{ik_0 r}}{r}\cos\theta,$$

which follows from Eq. (17) when the compact body and far field approximations are made. This expression lacks the viscous decay factor in Eq. (32), showing that the "long range" effect of the small viscosity is significant.

V. NULL-FIELD METHOD

The null-field method[15,16] enables the boundary data to be completed by imposing a zero disturbance within the scattering or radiating rigid body. For example, the application of the condition

$$\int_S \frac{\partial\rho_{sc}}{\partial x_j}(\mathbf{y})\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}dA_j(\mathbf{y}) = \int_S \rho_{sc}(\mathbf{y})\frac{\partial}{\partial y_i}\frac{e^{ik_0|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}-\mathbf{y}|}dA_i(\mathbf{y}) \quad (\mathbf{x}\text{ within } S) \tag{33}$$

to the inviscid scattering considered here yields the values of $(\rho_{sc})_{r=R}$, given by Eq. (17), because $r_>=R$, $r_<=r_x$ in Eq. (19). The total surface pressure is then

$$(p)_{r=R} = c_0^2(\rho_{inc}+\rho_{sc})_{r=R} = \frac{\alpha\rho_0 c_0^2}{(k_0R)^2}\sum_{n=0}^{\infty}\frac{(2n+1)i^{n+1}}{h_n^{(1)\prime}(k_0R)}P_n(\cos\theta). \tag{34}$$

Thus the null field is achieved by application of a suitable surface pressure whose associated net force is the inviscid limit of Eq. (31).

The method is unavailable for the vector potential associated with vorticity. In the viscous case, the sucessful strategy is to retain the inviscid solution with k_0 replaced by k_a, introduce vorticity modes with unknown coefficients, in each mode determine the necessary additional gradient field by applying the null-field method to either one of the no-slip conditions, and finish by using the other no-slip condition to evaluate the unknown coefficients.

VI. RIGID SPHERE/NULL FIELD AS LIMIT OF ELASTIC SPHERE/INTERIOR FIELD

We now replace the rigid sphere by an elastic sphere with density ρ_s and Lamé constants λ, μ. The displacement $\mathbf{u}$ within the sphere satisfies Navier's equation [Ref. 17, Eq. (1.38)]

$$\frac{1}{K^2}\boldsymbol{\nabla}^2\mathbf{u} + \left(\frac{1}{k^2}-\frac{1}{K^2}\right)\boldsymbol{\nabla}(\boldsymbol{\nabla}\cdot\mathbf{u}) = -\mathbf{u}, \tag{35}$$

where

$$k^2 = \frac{\rho_s\omega^2}{\lambda+2\mu}, \quad K^2 = \frac{\rho_s\omega^2}{\mu}. \tag{36}$$

From Eq. (35), it readily follows that

$$(\boldsymbol{\nabla}^2+k^2)\boldsymbol{\nabla}\cdot\mathbf{u} = 0, \quad (\boldsymbol{\nabla}^2+K^2)(\boldsymbol{\nabla}\times\mathbf{u}) = \mathbf{0}, \tag{37}$$

and hence, by comparison with Eqs. (27)–(29),

$$\boldsymbol{\nabla}\cdot\mathbf{u} = \alpha\sum_{n=0}^{\infty} c_n i^n j_n(kr)P_n(\cos\theta), \tag{38}$$

$$\boldsymbol{\nabla}\times\mathbf{u} = \alpha\sum_{n=1}^{\infty} d_n i^n j_n(Kr)P_n'(\cos\theta)\sin\theta\,\mathbf{e}_\phi, \tag{39}$$

$$\mathbf{u} = \frac{\alpha}{k}\sum_{n=0}^{\infty} c_n i^n\left\{-j_n'(kr)P_n(\cos\theta)\mathbf{e}_r + \frac{j_n(kr)}{kr}P_n'(\cos\theta)\sin\theta\,\mathbf{e}_\theta\right\} + \frac{\alpha}{K^2 r}\sum_{n=1}^{\infty} d_n i^n \times\left\{n(n+1)j_n(Kr)P_n(\cos\theta)\mathbf{e}_r - \frac{d}{dr}[rj_n(Kr)]P_n'(\cos\theta)\sin\theta\,\mathbf{e}_\theta\right\}. \tag{40}$$

The interior velocity field $\mathbf{v}^{INT}=-i\omega\mathbf{u}$ must match, at $r=R$, the additional exterior velocity $\mathbf{v}^{EXT}$, due to the elasticity of the sphere. The total velocity in $r>R$ is the sum of Eqs. (24) and (29) and

$$\mathbf{v}^{EXT} = \frac{\alpha\omega}{k_a}\sum_{n=0}^{\infty} a_n i^{n+1}\left\{h_n^{(1)\prime}(k_a r)P_n(\cos\theta)\mathbf{e}_r - \frac{h_n^{(1)}(k_a r)}{k_a r}P_n'(\cos\theta)\sin\theta\,\mathbf{e}_\theta\right\} + \frac{\alpha\omega}{k_a^2 r}\sum_{n=1}^{\infty} b_n i^{n+1} \times\left\{-n(n+1)h_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)P_n(\cos\theta)\mathbf{e}_r + \frac{d}{dr}\left[rh_n^{(1)}\left(r\sqrt{\frac{i\omega}{\nu}}\right)\right]P_n'(\cos\theta)\sin\theta\,\mathbf{e}_\theta\right\}. \tag{41}$$

The continuity of velocity components yields a_0 as a multiple of c_0 and, for each $n\geqslant 1$, a_n, b_n as linear combinations of c_n, d_n. Additional inhomogeneous equations are obtained by requiring the normal and tangential stresses to be continuous. However, the details are superfluous to our current purpose, for which it is sufficient to express the surface displacement vector in terms of the surface stresses and show that the former tends uniformly to zero as the elastic constants, i.e., the elastic wave speeds, tend to infinity.

For conciseness, we write

$$(\mathbf{u})_{r=R} = \alpha\Bigg\{ u_{r0}\mathbf{e}_r + \sum_{n=1}^{\infty} i^n[u_{rn}P_n(\cos\theta)\mathbf{e}_r + u_{\theta n}P_n'(\cos\theta)\sin\theta\mathbf{e}_\theta]\Bigg\}, \tag{42}$$

where, from Eq. (40)

$$u_{r0} = -c_0\frac{j_0'(kR)}{k},$$

$$u_{rn} = -c_n\frac{j_n'(kR)}{k} + d_n\frac{n(n+1)j_n(KR)}{K^2R} \quad (n \geqslant 1),$$

$$u_{\theta n} = c_n\frac{j_n(kR)}{k^2R} - d_n\frac{1}{K^2R}\frac{d}{dR}[Rj_n(KR)] \quad (n \geqslant 1).$$

The recurrence relations

$$j_{n-1}(x) - \frac{n+1}{x}j_n(x) = j_n'(x) = \frac{n}{x}j_n(x) - j_{n+1}(x),$$

(Ref. 13, Eqs. 11.163/164) show that

$$u_{r0} = \frac{c_0}{k}j_1(kR),$$

$$(2n+1)U_n = u_{rn} + nu_{\theta n} = \frac{c_n}{k}j_{n+1}(kR) + \frac{d_n}{K}nj_{n+1}(KR) \quad (n \geqslant 1),$$

$$(2n+1)V_n = u_{rn} - (n+1)u_{\theta n} = -\frac{c_n}{k}j_{n-1}(kR) + \frac{d_n}{K}(n+1)j_{n-1}(KR) \quad (n \geqslant 1), \tag{43}$$

which also defines $\{U_n, V_n; n \geqslant 1\}$ for subsequent convenience. We exclude from Eq. (42) a vector parallel to $\mathbf{e}_z = \cos\theta\mathbf{e}_r - \sin\theta\mathbf{e}_\theta$ as its inclusion would effectively rescale the incident field. Thus we impose

$$u_{r1} = u_{\theta 1}, \tag{44}$$

whose importance becomes apparent below.

The stress components within and on the elastic sphere are given [Ref. 17, Eq. (1.24)] by

$$\sigma_{rr}^{\mathrm{INT}} = 2\mu\frac{\partial u_r}{\partial r} + \lambda\,\boldsymbol{\nabla}\cdot\mathbf{u} = (\lambda+2\mu)\,\boldsymbol{\nabla}\cdot\mathbf{u} - \frac{2\mu}{r}\left[2u_r + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(u_\theta\sin\theta)\right]$$

$$(\sigma_{rr}^{\mathrm{INT}})_{r=R} = \frac{\alpha\rho_s\omega^2}{k^2}\sum_{n=0}^{\infty} c_n i^n j_n(kR)P_n(\cos\theta) - \frac{2\alpha\rho_s\omega^2}{K^2R}\times\Bigg\{2u_{r0} + \sum_{n=1}^{\infty} i^n[2u_{rn} + n(n+1)u_{\theta n}]P_n(\cos\theta)\Bigg\}, \tag{45}$$

$$\sigma_{r\theta}^{\mathrm{INT}} = \mu\left[\left(\frac{\partial}{\partial r} - \frac{1}{r}\right)u_\theta + \frac{\partial u_r}{\partial\theta}\right] = \mu\left[(\boldsymbol{\nabla}\times\mathbf{u})\cdot\mathbf{e}_\phi - \frac{2}{r}\left(u_\theta - \frac{\partial u_r}{\partial\theta}\right)\right]$$

$$(\sigma_{r\theta}^{\mathrm{INT}})_{r=R} = \frac{\alpha\rho_s\omega^2}{K^2}\sum_{n=1}^{\infty} d_n i^n j_n(KR)P_n'(\cos\theta)\sin\theta - \frac{2\alpha\rho_s\omega^2}{K^2R}\times\sum_{n=1}^{\infty} i^n[u_{rn} + u_{\theta n}]P_n'(\cos\theta)\sin\theta, \tag{46}$$

after substitution of Eqs. (36), (38), (39), and (42). The components of the stress exerted by the fluid on the sphere have the similar form

$$(\sigma_{rr}^{\mathrm{EXT}})_{r=R} = \alpha\rho_s\omega^2R^2\left[T_{r0} + \sum_{n=1}^{\infty} i^nT_{rn}P_n(\cos\theta)\right],$$

$$(\sigma_{r\theta}^{\mathrm{EXT}})_{r=R} = \alpha\rho_s\omega^2R^2\sum_{n=1}^{\infty} i^nT_{\theta n}P_n'(\cos\theta)\sin\theta,$$

and so stress continuity at $r=R$ is achieved, according to Eqs. (45) and (46), by setting

$$T_{r0} = c_0\frac{j_0(kR)}{k^2R^2} - \frac{4}{K^2R^3}u_{r0},$$

$$T_{rn} = c_n\frac{j_n(kR)}{k^2R^2} - \frac{2}{K^2R^3}[(n+1)(n+2)U_n - n(n-1)V_n] \quad (n \geqslant 1),$$

$$T_{\theta n} = d_n\frac{j_n(KR)}{K^2R^2} - \frac{2}{K^2R^3}[(n+2)U_n + (n-1)V_n] \quad (n \geqslant 1), \tag{47}$$

after the introduction of $\{U_n, V_n; n \geqslant 1\}$, defined by Eq. (43). We note that V_1 plays no role in the stress balance.

Equations (43) show that the coefficients in Eqs. (47) are given by

$$c_0/k = \frac{u_{r0}}{j_1(kR)},$$

$$\begin{bmatrix} c_n/k \\ d_n/K \end{bmatrix} = \frac{2n+1}{\Delta_n}\begin{bmatrix} (n+1)j_{n-1}(KR) & -nj_{n+1}(KR) \\ j_{n-1}(kR) & j_{n+1}(kR) \end{bmatrix} \times \begin{bmatrix} U_n \\ V_n \end{bmatrix} \quad (n \geqslant 1), \tag{48}$$

where $\Delta_n = (n+1)j_{n-1}(KR)j_{n+1}(kR) + nj_{n+1}(KR)j_{n-1}(kR)$. To examine the limit $KR, kR \to 0$, it is sufficient to deduce from the series expansion for $j_n(x)$ (Ref. 13, Eq. 11.144) that

$$\frac{xj_n(x)}{j_{n+1}(x)} \sim 2n+3 \quad (0 < x \ll 1, n \geqslant 0), \tag{49}$$

whose substitution in Eqs. (48) shows that the required terms in Eqs. (47) are estimated by

$$c_0 \frac{j_0(kR)}{k^2R^2} \sim \frac{3}{k^2R^3} u_{r0},$$

$$\begin{bmatrix} c_n \dfrac{j_n(kR)}{(kR)^2} \\ d_n \dfrac{j_n(KR)}{(KR)^2} \end{bmatrix} \sim \frac{1}{R[n+1+n(K/k)^2]} \begin{bmatrix} \dfrac{n+1}{(kR)^2} & -n\left(\dfrac{K}{k}\right)^2 \\ \dfrac{1}{(kR)^2} & 1 \end{bmatrix} \times \begin{bmatrix} (2n+1)(2n+3)U_n \\ V_n \end{bmatrix} \quad (n \geqslant 1), \quad (50)$$

in which $K^2 > k^2$, according to Eq. (36). Substitution of Eqs. (50) into Eqs. (47) shows that $u_{r0} \to 0$, $U_n \to 0 (n \geqslant 1)$ and $V_n \to 0 (n > 1)$ as $KR, kR \to 0$. Together with Eq. (44) and the definitions in Eq. (43), it follows that $u_{rn} \to 0 (n \geqslant 0)$ and $u_{\theta n} \to 0 (n \geqslant 1)$, which demonstrates, according to Eq. (42), that the surface, and hence interior, displacement vector tends uniformly to zero, yielding a null field, as the elastic wave speeds tend to infinity.

VII. SUMMARY

We present a derivation of Curle's equation for linear time-harmonic acoustics, and give an explicit verification of Curle's equation for two examples of acoustic scattering by a rigid sphere of arbitrary size in an inviscid fluid. We construct the field scattered by a rigid sphere of arbitrary size in a viscous fluid, and show that a previous result for a small sphere, derived from the inviscid form of Curle's equation, is indeed an inviscid estimate. The null-field approach to the sphere scattering problem is discussed, and its extension to the vorticity modes associated with viscosity is indicated. Finally, we construct a solution for the scattering by an elastic sphere in a viscous fluid, and show that the rigid sphere/null-field solution is recovered in the limit of infinite longitudinal and shear wave speeds in the elastic solid.

[1] A. Zinoviev and D. A. Bies, "On acoustic radiation by a rigid object in a fluid flow," J. Sound Vib. **269**, 535–548 (2004).
[2] N. Curle, "The influence of solid boundaries upon aerodynamic sound," Proc. R. Soc. London, Ser. A **231**, 505–514 (1955).
[3] J. E. Ffowcs Williams and D. L. Hawkins, "Sound generation by turbulence and surfaces in arbitrary motion," Philos. Trans. R. Soc. London, Ser. A **264**, 321–342 (1969).
[4] A. Zinoviev and D. A. Bies, "Authors' reply," J. Sound Vib. **281**, 1224–1237 (2005).
[5] A. M. J. Davis and R. J. Nagem, "Acoustic diffraction by a half-plane in a viscous fluid medium," J. Acoust. Soc. Am. **112**, 1288–1296 (2002).
[6] A. M. J. Davis and R. J. Nagem, "Influence of viscosity on the diffraction of sound by a circular aperture in a plane screen," J. Acoust. Soc. Am. **113**, 3080–3090 (2003).
[7] A. M. J. Davis and R. J. Nagem, "Effect of viscosity on acoustic diffraction by a circular disk," J. Acoust. Soc. Am. **115**, 2738–2748 (2004).
[8] M. J. Lighthill, "On sound generated aerodynamically," Proc. R. Soc. London, Ser. A **211**, 564–586 (1952).
[9] F. Farassat, "Comments on the paper by Zinoviev and Bies 'On acoustic radiation by a rigid object in a fluid flow'," J. Sound Vib. **281**, 1217–1223 (2005).
[10] F. Farassat, "Introduction to generalised functions with applications in aerodynamics and aeroacoustics," NASA Technical Paper 3428, 1994 (available at http://techreports.larc.nasa.gov/ltrs/PDF/tp3428.pdf).
[11] M. S. Howe, *Theory of Vortex Sound* (Cambridge University, Cambridge, England, 2003).
[12] A. M. J. Davis and F. G. Leppington, "Scattering of electromagnetic surface waves by circular or elliptic cylinders," Proc. R. Soc. London, Ser. A **353**, 55–75 (1977).
[13] G. Arfken, *Mathematical Methods for Physicists*, 3rd ed. (Academic Press, New York, 1985).
[14] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics II* (McGraw-Hill, New York, 1953).
[15] P. A. Martin, "On the null-field equations for the exterior problems of acoustics," Q. J. Mech. Appl. Math. **33**, 385–396 (1980).
[16] P. A. Martin, "On the null-field equations for water-wave radiation problems," J. Fluid Mech. **113**, 315–332 (1981).
[17] J. A. Hudson, *The Excitation and Propagation of Elastic Waves*, Cambridge University, Cambridge, England, 1980.

Acoustical determination of the parameters governing viscous dissipation in porous media

Raymond Panneton[a)]
GAUS, Department of Mechanical Engineering, Université de Sherbrooke (Qc), J1K 2R1, Canada

Xavier Olny
ENTPE—DGCB URA CNRS 1652-rue Maurice Audin 69518 Vaulx en Velin—France

(Received 9 August 2005; revised 15 December 2005; accepted 8 January 2006)

Analytical solutions are derived to extract from dynamic density the macroscopic parameters governing viscous dissipation of sound waves in open-cell porous media. While dynamic density is obtained from acoustical techniques, the analytical solutions are derived from the model describing this dynamic density. Here, semiphenomenological models by Johnson *et al.* and by Wilson are investigated. Assuming dynamic density, open porosity, and static airflow resistivity known, analytical solutions derived from the Johnson *et al.* model yield geometrical tortuosity and viscous characteristic dimension. For the Wilson model, only dynamic density needs to be known. In this case, analytical solutions yield—for the first time—Wilson's density parameter and vorticity-mode relaxation time. To alleviate constraints on the Johnson *et al.* model, an extrapolation approach is proposed to avoid prior knowledge of static resistivity. This approach may also be used to determine this latter parameter. The characterization methods are tested on three materials covering a wide range of static airflow resistivities (2300–150 100 Ns/m^4), frame rigidities (soft and rigid), and pore geometries (cells and fibers). It is shown that the analytical solutions can be used to assess the validity of the descriptive models for a given material. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2169923]

PACS number(s): 43.20.Gp, 43.20.Jr, 43.20.Ye [LLT] Pages: 2027–2040

I. INTRODUCTION

In this paper, attention is focused on the determination of the macroscopic parameters governing viscous dissipation in air-saturated rigid-frame open-cell porous media. In classical acoustic modeling of rigid-frame porous media,[1] viscous dissipation is taken into account through a dynamic density—also known as complex density or effective density. To date, the three-parameter semiphenomenological model worked out by Johnson *et al.*[2] is commonly used. The three parameters are the static airflow resistivity, geometrical tortuosity, and viscous characteristic dimension. To improve the prediction of the real part of the low-frequency limit, Pride *et al.*[3] introduced an additional parameter. At the same time, Wilson[4] proposed a two-parameter relaxation model for the dynamic density. The two parameters are a density parameter and a vorticity-mode relaxation time. In his paper, Wilson showed the usefulness and applicability of his model, and compared it to the Johnson *et al.* model, notably.

To characterize the parameters governing viscous dissipation in porous media, many authors have worked out methods. They may be divided into nonacoustical and acoustical methods. The main nonacoustical methods are used to characterize static airflow resistivity[5,6] and geometrical tortuosity.[7,8] On the other hand, the main recent acoustical methods can be divided themselves into low-frequency[9,10] and high-frequency methods.[11–13] While low-frequency methods are usually based on the use of an impedance tube, high-frequency methods use ultrasound techniques. Except for low-frequency acoustical methods, existing characterization methods require special apparatuses not usually available in typical acoustic labs. Moreover, the aforementioned methods do not deal with the characterization of the parameters used in the Wilson model. In his paper, Wilson proposed an acoustical method, where the vorticity-mode relaxation time is adjusted in such a way that the relaxation model matches the mid-frequency behavior. Unfortunately, the method was not experimentally tested, and no characterization method was proposed for the density parameter; however, he suggested relations to link his two parameters to the geometrical tortuosity and viscous characteristic dimension.

Our aim in the present work is to develop acoustical methods to determine the classical parameters governing viscous dissipation in porous media and, for the first time, the two Wilson's parameters. These methods are based on previous works by the authors.[14,15] They do not use sophisticated apparatuses and are not inverse methods.[9] They only require acoustical measurements of dynamic density—using an impedance tube technique, and analytical solutions derived from the descriptive models to extract their parameters. In that sense, they are rather indirect characterization methods than inverse ones. Also, unlike ultrasound techniques based on the exact high-frequency limit of the Johnson *et al.* model, the developed methods are based on larger wavelength behavior, where the homogenization assumptions[1] behind the models are verified. This is not necessarily guaranteed when using ultrasound techniques.[16]

[a)]Corresponding author.
Electronic mail: raymond.panneton@usherbrooke.ca

The present paper is organized as follows. In the next section, the descriptive models by Johnson *et al.* and Wilson are recalled to pave to way to the following developments. The third section presents the development of the analytical method, where analytical solutions are derived from the two descriptive models. In the fourth section we present the extrapolation method that applies to the Johnson *et al.* model. In the fifth section we estimate the error of the analytical method in the characterization of the Johnson *et al.* parameters for two virtual materials of theoretically known properties. In the sixth section we present the application of the analytical and extrapolation methods to three porous materials covering a wide range of static airflow resistivities, frame rigidities, and pore geometries. Finally, concluding remarks and prospects are given in the last section.

II. THEORY

Let us consider an isotropic open-cell porous medium. If the vibration of its solid frame is negligible (i.e., motionless), only sound waves propagate in the medium. In this case, the porous medium is seen as an equivalent fluid characterized by an equivalent dynamic density $\tilde{\rho}_{\mathrm{EQ}}$ and a dynamic bulk modulus $\tilde{K}_{\mathrm{EQ}}$—the overtilde indicates that the associated variable is frequency dependent and complex valued. These two dynamic properties are also used in the Biot theory[17,18] to account for the viscoinertial and elastic interactions between air and frame in a poroelastic aggregate.

While propagating through the motionless porous network, the sound waves are attenuated due to viscous and thermal dissipation mechanisms. The dynamic density accounts for the viscous losses, and the dynamic bulk compression modulus for the thermal losses. In the past, many models relating these dynamic properties to the geometrical parameters of the porous network have been worked out. In this work, on the parameters governing the viscous losses, two models are selected to describe the dynamic density: (i) the semiphenomenological model by Johnson *et al.* and (ii) the semiphenomenological model by Wilson. These two models and their parameters are recalled later.

A. Johnson *et al.* model

The semiphenomenological model of Johnson *et al.*[2] is commonly used to describe the complex density of the fluid in a porous medium. It has been found more accurate than previous models for arbitrary pore shapes.[1] In this model, using the $\exp(j\omega t)$ time convention, the dynamic density of the fluid phase may be written as

$$\tilde{\rho} = \rho_0 \alpha_\infty \left(1 - j\frac{1}{\varpi} G_J(M\varpi)\right) \tag{1}$$

with

$$G_J = \sqrt{1 + j\frac{M}{2}\varpi}. \tag{2}$$

Here, M is the dimensionless pore shape factor and ϖ the reduced frequency ω/ω_ν, where ω and ω_ν are the angular frequency and viscous characteristic frequency. M and ω_ν are related to the properties of the saturating fluid and to the geometrical parameters of the porous network by

$$M = \frac{8\alpha_\infty \eta}{\phi \sigma \Lambda^2} \tag{3}$$

and

$$\omega_\nu = \frac{\sigma\phi}{\alpha_\infty \rho_0}. \tag{4}$$

In these equations, ρ_0 and η are the density and dynamic viscosity of the saturating fluid, and ϕ, σ, α_∞, and Λ are the open porosity, static flow resistivity, geometrical tortuosity, and viscous characteristic dimension of the porous medium, respectively. Following Johnson *et al.*, these four latter properties macroscopically describe, in terms of the viscous losses, the geometry of the open porous network.

For comparisons to the Wilson model, the following vorticity-mode relaxation time is also introduced [given by Eq. (A6) in Ref. 4],

$$\tau_J = \frac{M}{2\omega_\nu}. \tag{5}$$

The subscript J is used to indicate that this relaxation time is based on the Johnson *et al.* model.

The connection between the dynamic density of the fluid phase given in (1) and the dynamic density of the equivalent fluid medium (bulk property) is

$$\tilde{\rho}_{\mathrm{EQ}} = \frac{\tilde{\rho}}{\phi}. \tag{6}$$

In the context of this work, the use of the bulk property is preferable since it is the property obtained from acoustical measurements (see Sec. VI). This dynamic density being complex valued, it can be decomposed in real and imaginary parts. After some mathematics, one obtains

$$\tilde{\rho}_{\mathrm{EQ}} = \left(\frac{\rho_0 \alpha_\infty}{\phi} + \frac{\sigma}{\omega} G_I\right) - j\frac{\sigma}{\omega} G_R, \tag{7}$$

with

$$G_R = \frac{1}{\sqrt{2}} \sqrt{1 + \sqrt{1 + \left(\frac{M\varpi}{2}\right)^2}} \quad \text{and} \quad G_I = \frac{M\varpi}{4G_R}. \tag{8}$$

The positive roots are used in (8). This explicit writing of the Johnson *et al.* dynamic density easily shows its high- and low-frequency limits: (i) imaginary low-frequency limit: $-\sigma/\omega$, and (ii) real high-frequency limit: $\rho_0\alpha_\infty/\phi$. These limits are theoretically exact. Only the low-frequency real limit is not exact.

Consequently, the Johnson *et al.* model was worked out to fit the exact high- and low- (imaginary) frequency limits of the dynamic density, with the introduction of a characteristic dimension and function G_J to fit its mid-frequency behavior.

B. Wilson model

The model by Wilson[4] uses only two parameters to define the dynamic density. Unlike other models for arbitrary

pore shapes,[2,19,20] it explicitly formulates the viscous diffusion in the pores as a relaxation process. Using the $\exp(j\omega t)$ time convention, the dynamic density of the equivalent fluid in the Wilson model takes the form

$$\tilde{\rho}_{\mathrm{EQ}}^{W} = \rho_\infty \frac{\sqrt{1 + j\omega\tau_\nu}}{\sqrt{1 + j\omega\tau_\nu} - 1}, \tag{9}$$

where the two Wilson parameters are ρ_∞ the high-frequency limit of $\tilde{\rho}_{\mathrm{EQ}}^{W}$, and τ_ν the vorticity-mode relaxation time. If one constrains the Wilson model to yield the exact low- (imaginary) and high-frequency limits, the following relations are deduced: $\tau_\nu = 2/\omega_\nu = 2\rho_0\alpha_\infty/\phi\sigma$ and $\rho_\infty = \rho_0\alpha_\infty/\phi$. Consequently, there remain no more free parameters to adjust to the mid-frequency range. However, contrary to Johnson *et al.*, the Wilson model is not built to fit the exact high- and low-frequency behaviors of the dynamic density. Instead, it is built to match the mid-frequency range using two degrees of freedom.

In his work, Wilson proposes adjusting τ_ν to match the measured acoustical relaxation characteristic of the medium (see Sec. II of Ref. 4) around the viscous characteristic frequency. If no acoustical measurements are available, using our notations, Wilson proposes to use $\tau_\nu = 4/M\omega_\nu = \rho_0\Lambda^2/2\eta$ to reproduce the medium frequency range. For ρ_∞, he proposes to use $\rho_\infty = \rho_0\alpha_\infty/\phi$—no specific experimental evaluation is proposed. In this work, an experimental evaluation of the Wilson viscous parameters will be investigated.

III. ANALYTICAL METHOD

A. Johnson *et al.* parameters

Assuming dynamic density, static flow resistivity, and open porosity known from prior measurements, only two unknown parameters, α_∞ and Λ, subsist. If the measured equivalent dynamic density $\tilde{\rho}_{\mathrm{EQ}}$ is expressed in terms of its real X and imaginary Y parts, separating analytically (6) into its real and imaginary parts yields a system of two equations:

$$\begin{aligned} &\mathrm{Real} \rightarrow X - \frac{\rho_0\alpha_\infty}{\phi} - \frac{\sigma}{\omega}G_I = 0, \\ &\mathrm{Imag} \rightarrow Y + \frac{\sigma}{\omega}G_R = 0. \end{aligned} \tag{10}$$

Solving this system for α_∞ and Λ, the following admissible analytical solutions are found:

$$\alpha_\infty = \frac{\phi}{\rho_0}\left[X - \sqrt{Y^2 - \left(\frac{\sigma}{\omega}\right)^2}\right] \tag{11}$$

and

$$\Lambda = \alpha_\infty \sqrt{\frac{2\rho_0\eta}{\omega\phi Y(\rho_0\alpha_\infty - \phi X)}}. \tag{12}$$

Consequently, knowing the static flow resistivity, porosity, and dynamic density, these analytical solutions can be used to determine the geometrical tortuosity and viscous characteristic dimension of the Johnson *et al.* model. Here, α_∞ is first determined with (11). Then, it is used in (12) to determine Λ.

According to (11) and (12), the parameters seem to be frequency dependent. However, this is not true if the material follows the behavior given by the model. It is important to underline that the determined α_∞ and Λ are not, strictly speaking, the intrinsic geometrical parameters of the material, except when frequency tends toward infinity, where the Johnson *et al.* model is exact. However, if (11) and (12) yield constant values for α_∞ and Λ in the function of the frequency, then the determined α_∞ and Λ may be seen as intrinsic to the material. Consequently, the constancy of the determined parameters in a given frequency range assesses the validity of the Johnson *et al.* model in this range.

B. Wilson parameters

Once again, assuming the equivalent dynamic density is obtained experimentally, the analytical separation of the real and imaginary parts in (9) yields a system of two equations:

$$\begin{aligned} &\mathrm{Real} \rightarrow -2X\rho_\infty + \rho_\infty^2 - 2\omega\tau_\nu YX + 2\omega\tau_\nu Y\rho_\infty = 0, \\ &\mathrm{Imag} \rightarrow -2Y\rho_\infty + \omega\tau_\nu((X - \rho_\infty)^2 - Y^2) = 0. \end{aligned} \tag{13}$$

Solving this system, three admissible analytical solutions for ρ_∞ and one for τ_ν are found:

$$\rho_\infty = \begin{cases} \dfrac{1}{3}\left[A + 4X + \left(\dfrac{X^2 - 9Y^2}{A}\right)\right], \\ -\dfrac{1}{6}\left\{A - 8X + \left(\dfrac{X^2 - 9Y^2}{A}\right) - j\sqrt{3}\left[A - \left(\dfrac{X^2 - 9Y^2}{A}\right)\right]\right\}, \\ -\dfrac{1}{6}\left\{A - 8X + \left(\dfrac{X^2 - 9Y^2}{A}\right) + j\sqrt{3}\left[A - \left(\dfrac{X^2 - 9Y^2}{A}\right)\right]\right\}, \end{cases} \tag{14}$$

and

$$\tau_\nu = \frac{2Y\rho_\infty}{\omega}\frac{1}{(X - \rho_\infty)^2 - Y^2}, \tag{15}$$

with

$$A = \sqrt[3]{X^3 - 27XY^2 + 3\sqrt{-3X^4Y^2 + 54X^2Y^4 + 81Y^6}}. \tag{16}$$

Once the proper solution for ρ_∞ is found using (14), it is substituted into (15) to determine τ_ν. Since A may be complex valued, the main difficulty lies in the selection of the proper solution for ρ_∞. A valid solution for ρ_∞ has to be real and yields a positive real value for τ_ν. As will be shown in Sec. VI, any of the three solutions may apparently be the valid one.

Contrary to the Johnson *et al.* model, this time the static flow resistivity and porosity are not needed for the determination to proceed—only the knowledge of the equivalent dynamic density is required. Finally, since the main advantage of the Wilson model is to match the medium frequency range, the determination of the parameters should be performed over this range.

IV. EXTRAPOLATION METHOD

In Sec. III A, analytical solutions were developed to extract α_∞ and Λ from the prior knowledge of the static flow resistivity, porosity, and equivalent dynamic density. In the following, a characterization method is proposed to determine α_∞, σ, and Λ from the prior knowledge of the porosity and equivalent dynamic density only.

A. Determination of the geometrical tortuosity

Let the following variable be defined as

$$\Sigma = \frac{\phi}{\rho_0}[\mathrm{Re}(\tilde{\rho}_{\mathrm{EQ}}) + \mathrm{Im}(\tilde{\rho}_{\mathrm{EQ}})]. \tag{17}$$

Now, by the use of (7) and (8), this variable rewrites

$$\Sigma = \alpha_\infty + (G_I - G_R)\frac{\sigma\phi}{\omega\rho_0}. \tag{18}$$

At sufficiently high frequencies, where $\varpi \gg 1$, G_R and G_I tend to the same asymptotic behavior, that is

$$G_I, G_R \simeq \sqrt{\frac{M}{2}\varpi}. \tag{19}$$

This shows that Σ tends to the geometrical tortuosity as ϖ increases. Contrary to ultrasound techniques, based on the exact high-frequency limit, the current method is based on a mid-frequency behavior, where the homogenization assumptions[1] on which the porous equations are based still hold. This is not necessarily guaranteed when using ultrasound techniques.[16]

In the case of low static flow resistivity materials—or, more generally, for low viscous characteristic frequencies, Σ measured in the mid-frequency range (typically around 6000 Hz) is a good estimate of the geometrical tortuosity. However, for higher static airflow resistivity materials, or to increase the accuracy and robustness of the method, an extrapolation of Σ to infinite frequency is preferred. This is explained in the following.

From (18), one can demonstrate that Σ tends to the geometrical tortuosity following a frequency dependence of the form $\omega^{-3/2}$. In fact, for an increasing ϖ greater than unity, (18) approaches

$$\Sigma \doteq \alpha_\infty\left(1 - \sqrt{\frac{1}{M}}\varpi^{-3/2}\right), \tag{20}$$

which is a linear equation of the form $y=mx+b$, with $x=(\omega_\nu\varpi)^{-3/2}=\omega^{-3/2}$, $m=-\alpha_\infty\sqrt{\omega_\nu^3/M}$, and $b=\alpha_\infty$. Consequently, plotting Σ in function of $\omega^{-3/2}$ gives a straight line of slope m and ordinate b. The extrapolation of the line to infinite frequency ($\omega^{-3/2}=0$) yields the geometrical tortuosity ($\alpha_\infty=b$) without prior knowledge of the other material properties, except the porosity.

For the extrapolation method to be valid, (20) requires $\varpi > 1$, or at least in the vicinity of 1. This is usually verified if the following condition is met:

$$\omega > \sqrt[3]{m^2}, \tag{21}$$

where slope m is obtained from a linear regression on the measured variable Σ—as explained in Sec. VI.

B. Determination of the static flow resistivity

Once α_∞ is found from extrapolation, one can obtain from (11) the static flow resistivity,

$$\sigma = \omega\sqrt{Y^2 - \left(X - \frac{\rho_0\alpha_\infty}{\phi}\right)^2}. \tag{22}$$

In the case of high static flow resistivity media, the determination of the tortuosity from extrapolation may be difficult since the high-frequency behavior ($\varpi \gg 1$) may be far from the measured frequency range for $\tilde{\rho}_{\mathrm{EQ}}$ (or X and Y). In this case, the viscous characteristic frequency is usually high, and a good estimate of the resistivity may be obtained at low frequencies from the dynamic resistivity, defined as

$$\tilde{\sigma} = j\omega\tilde{\rho}_{\mathrm{EQ}}. \tag{23}$$

The real part of the low-frequency limit of the dynamic resistivity is the static resistivity.

C. Determination of the viscous characteristic dimension

Once σ is found, Λ is inferred from (20), and the linear regression on the measured variable Σ (see Sec. VI A for an application). Substituting (3) and (4) into (20), the following is obtained:

$$\Lambda = -mb\sqrt{\frac{8\eta\rho_0^3}{\sigma^4\phi^4}}, \tag{24}$$

where m and b are the slope and ordinate ($\doteq\alpha_\infty$) of the linear regression.

V. ESTIMATION OF THE ERROR USING JOHNSON ET AL MODEL

Estimating the error of the proposed characterization method is tricky since several assumptions are used, and the precision highly relies on the ability to measure the equivalent dynamic density. However, one important point can be tackled. It deals with the error made on the real intrinsic parameters of the material using the analytical solutions (11), (12), and (22) derived from the Johnson *et al.* model.

In this section, two virtual materials are studied. The first one is a layer of oblique parallel circular cylindrical pores of radius R. The second one is a layer of oblique parallel slits of width $2a$. The angle between the pore axis and the normal to the layer is denoted θ. For such simple pore geometries, exact solutions for the dynamic density are known.[1] They are recalled in Appendix A. Using these exact solutions, measurements can be simulated and the viscous parameters of the Johnson *et al.* model can be determined using the developed analytical solutions (11), (12), and (22).

Let us now consider the relative error made on the determination of a parameter by

$$\delta_u = \frac{\hat{u} - u}{u}, \tag{25}$$

with u the theoretical value and $\hat{u}$ the value determined using (11), (12), and (22). For the studied virtual materials, the theoretical viscous characteristic dimensions are given by $\Lambda_c = R$ for the circular pores and $\Lambda_s = 2a$ for the slits. The theoretical tortuosity is given by (A5) for both materials, and the theoretical resistivity by (A6) and (A11). Figure 1 shows the relative errors of the analytical determinations for both materials in function of the reduced frequency.

In Fig. 1(a), one can note that for both materials the error on the tortuosity passes through a maximum (2.5% in absolute value), and then decreases to zero as the frequency increases. This is compatible with the fact the high-frequency limit of the Johnson *et al.* dynamic density is exact. As for the maximum error at mid-frequency, it is linked to the maximum difference between the approximate dissipation function G_J given in (2) and the exact dissipation functions G_c and G_s given in (A2) and (A8). At the lower limit of the spectrum, the error approaches zero for the circular pore and to approximately 1.5% for the slit. These low errors seem to be in contradiction with the fact the real part of the Johnson *et al.* dynamic density is not exact at the low-frequency limit. For these virtual materials, the real part predicted by (1) is 6.25% and 2.78% lower than the exact low-frequency limit for the circular pore and the slit, respectively. This is much more than the relative error observed in Fig. 1(a). This is explained by the fact that the analytical determination using (11) does not only depend on the real part X, but also depends on the imaginary part Y, which is exact at the low-frequency limit.

In Fig. 1(b), a similar analysis applies to the viscous characteristic dimension. However, the error at low frequencies is larger than the one observed for the tortuosity. In the worst case, the maximum error (in absolute value) is less than 20%, and occurs at medium frequencies. This is quite acceptable since this parameter is difficult to obtain with accuracy using other existing methods.

In Fig. 1(c), one can note that the relative error on the determination of the resistivity using (22) is very small (less than 1%) at low frequencies ($\varpi < 1$), and then rapidly increases with the frequency.

To conclude, the previous error analysis showed that extraction using analytical solutions should proceed preferably at high reduced frequencies ($\varpi > 1$) for α_∞ and Λ, and at low reduced frequencies ($\varpi < 1$) for σ. This would minimize the relative errors with respect to the real parameters of the material to characterize. If this cannot be respected, systematic errors on the determination of the parameters may be introduced. Also, for the determination of Λ, the frequency range in the vicinity of $\varpi \approx 1$ should be avoided.

VI. RESULTS

The proposed characterization methods are now applied to three porous materials: (i) low-resistivity polyurethane foam, (ii) medium-resistivity metal foam, and (iii) high-resistivity rock wool. The polyurethane foam has a relatively

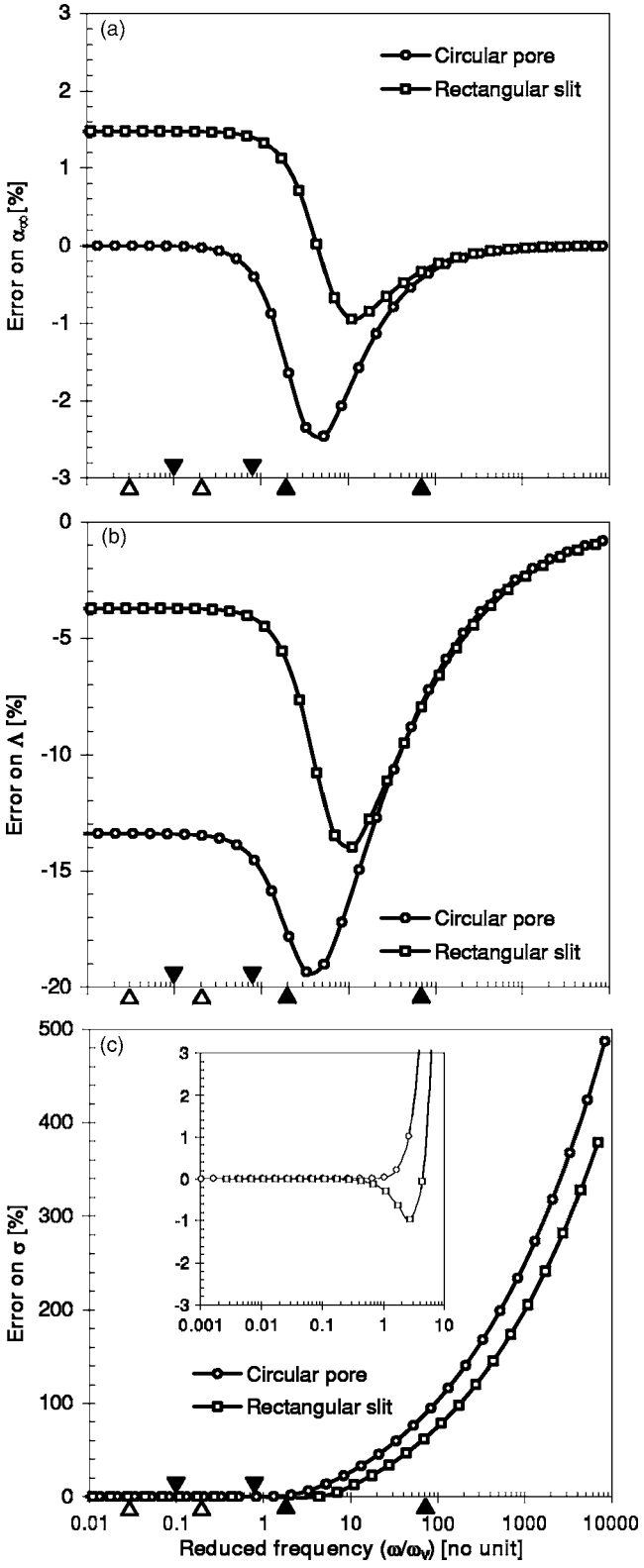


FIG. 1. Errors on the Johnson *et al.* parameters determined from the analytical solutions. The errors are relative to the theoretical parameters of two virtual porous materials having (◦) cylindrical pores of circular cross-section and (□) rectangular slits. (a) Error on the geometrical tortuosity. (b) Error on the viscous characteristic dimension. (c) Error on the static resistivity. Symbols ▲, ▼, and △ indicate the limits of the frequency range used during extraction of the parameters in Sec. VI for the low resistivity, medium resistivity, and high resistivity materials, respectively. The inlaid graph in (c) is a zoom in the vicinity of $\omega/\omega_\nu = 1$.

TABLE I. Physical properties of the tested porous specimens. Open porosity and static airflow resistivity are measured using direct techniques (Refs. 5, 21, and 22). Static pressure and relative humidity prevailing during impedance tube measurements are used in the evaluation of the properties of the saturating moist air (Refs. 26 and 27).

	Symbol	Polyurethane foam (low resistivity)	Metal foam (medium resistivity)	Rock wool (high resistivity)	Units
Static airflow resistivity @ 100 sccm	σ	2 300	50 010	150 100	N s/m^4
Open porosity	ϕ	0.960	0.890	0.935	
Density of saturating air	ρ_0	1.163	1.149	1.187	kg/m^3
Dynamic viscosity of saturating air	η	1.84×10^{-5}	1.83×10^{-5}	1.83×10^{-5}	N s/m^2

soft frame. The metal foam has a rigid nickel-based frame. The rock wool has a random fibrous structure. Preliminary tests have been performed to measure the open porosity and static airflow resistivity of the three materials using published direct techniques.[5,21,22] The results of these tests are given in Table I.

Also, based on the impedance tube technique by Utsono *et al.*,[23] the characteristic impedance $\tilde{Z}_c$ and propagation constant $\tilde{\gamma}$ of each material were measured, and the equivalent dynamic density deduced from $\tilde{\rho}_{\mathrm{EQ}}=-j\tilde{Z}_c\tilde{\gamma}/\omega$. The measurements used either a small (29 mm), or medium (44 mm), diameter tube to obtain nice results in the frequency range of interest (500–4000 Hz). Since the proposed characterization methods assume a rigid-frame behavior (i.e., motionless), the method proposed by Iwase *et al.*[24] to suppress or minimize the resonant vibrations of the frame is used. Briefly, this method consists in sticking nails (or needles) normally into the porous specimen. Due to the friction force between the nails and the skeleton, the resonant vibrations are minimized. Few nails are usually needed since only the first resonant mode is to be suppressed. Higher resonant modes are usually not excited since they occur for frequencies higher than the fluid-solid decoupling frequency.[25] Sticking thin nails into the material modifies only slightly its properties. For example, 20 nails of 0.58 mm diameter accounts for only 1% of the total surface of a 29 mm diameter specimen. Moreover, in this latter example, tests showed no modification in the static resistivity. Finally, in addition to the previous precautions to eliminate frame vibrations, care should also be taken to prevent edge-constrained and air-leakage effects during impedance tube measurements.[28,29]

The real X and imaginary Y parts of the equivalent dynamic densities obtained following the aforementioned method and precautions are shown in Fig. 2. For the polyurethane foam and rock wool, 10 and 20 needles of 0.58 mm in diameter were used to minimize the resonant vibrations, respectively. The metal foam being rigid, no nails were used. As expected, the measured dynamic densities are typical of rigid-frame porous materials. However, the polyurethane foam still shows some resonant vibrations at low frequencies. To see their influence on the proposed determination method, no additional effort was made to suppress them.

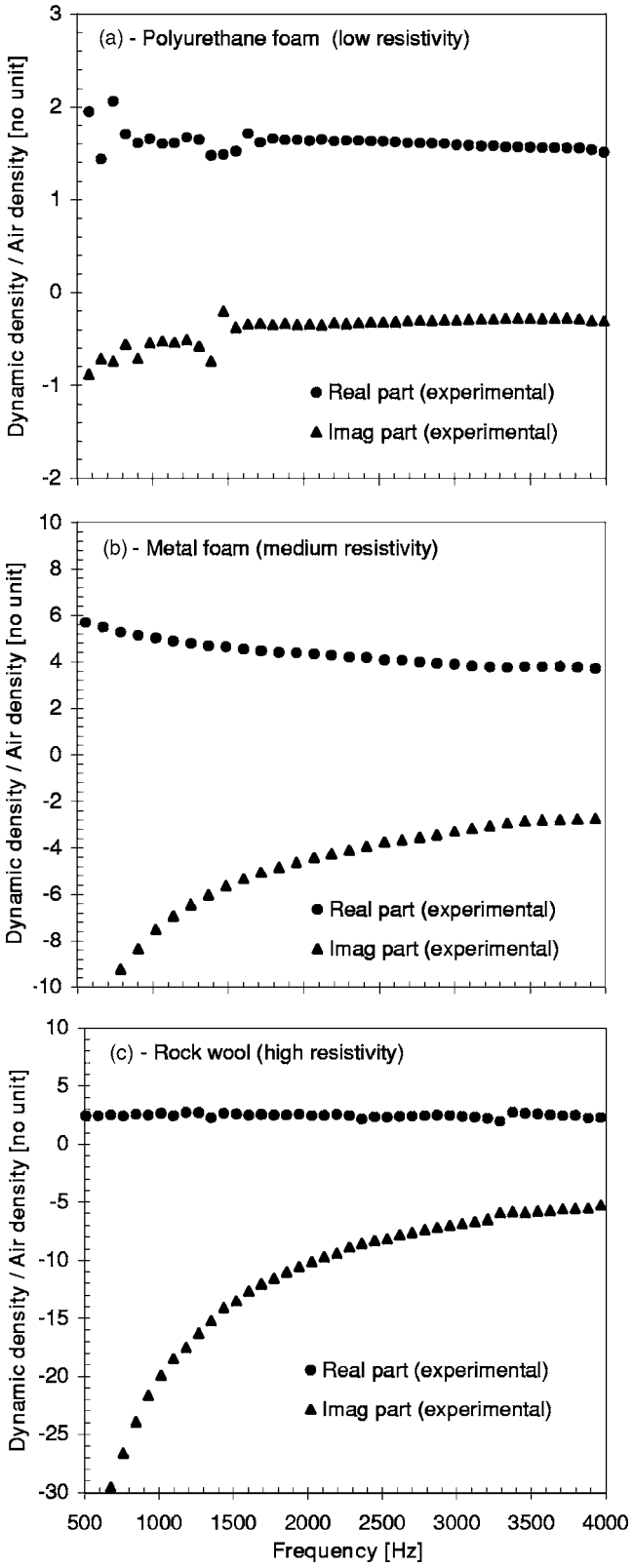


FIG. 2. Measured equivalent dynamic density normalized by the air density (ρ_0).

A. Extrapolation method—Johnson *et al.* model

1. Geometrical tortuosity

Following the extrapolation method explained in Sec. IV, the measured variable Σ is first plotted as a function of $\omega^{-3/2}$; see Fig. 3. Here, Σ is obtained by substituting the real X and imaginary Y parts of the measured equivalent dynamic density into (17), i.e., $\Sigma=\phi(X+Y)/\rho_0$, where ϕ and ρ_0 are given in Table I. Next, a frequency range where the evolution is linear is selected (black dots in Fig. 3). For the three materials, the range 1500–3000 Hz is selected. The nonlinear behavior in the range 3000–4000 Hz is attributed to the lack of accuracy in the measurements of the dynamic density. In the selected frequency range, a linear regression is applied on the measured Σ. For each material, the equation of the resulting linear regression ($\Sigma=m\omega^{-3/2}+b$) is given in the figure. Validity of the extrapolation method is first verified using condition (21):

Low resistivity material:

$\omega > 1614$ rad/s (257 Hz) → verified;

Medium resistivity material:

$\omega > 15849$ rad/s (2522 Hz) → verified;

High resistivity material:

$\omega > 43267$ rad/s (6886 Hz) → not verified.

The condition is verified for the two lower resistive materials. Consequently, coefficient b of the linear regression should be a good estimate of their geometrical tortuosity: $\alpha_\infty=1.29$ for the low resistivity material, and $\alpha_\infty=1.27$ for the high resistivity material. For the high resistivity material, condition (21) is not verified and the extrapolation method cannot be used—the data are obviously in the low-frequency behavior ($\varpi<1$) of the material. In this case, increasing the frequency range, for example, up to 6000 Hz, may be sufficient to use the method. Here no attempt is done to extend the frequency range. Instead, the tortuosity for this latter material will be estimated indirectly in the next section.

The previously determined values for α_∞ are copied into Table II. Substituting these values together with the properties given in Table I into (4), the viscous characteristic frequencies for the three materials are estimated to, respectively, 234, 4854, and 18 660 Hz. Consequently, the studied frequency range 500–4000 Hz corresponds to the following reduced frequency ranges: [2.1–17.1] for the low resistivity material, [0.1–0.8] for the medium resistivity material, and [0.03–0.21] for the high resistivity material. These limits are shown in Fig. 1. The following analyses will refer to these limits.

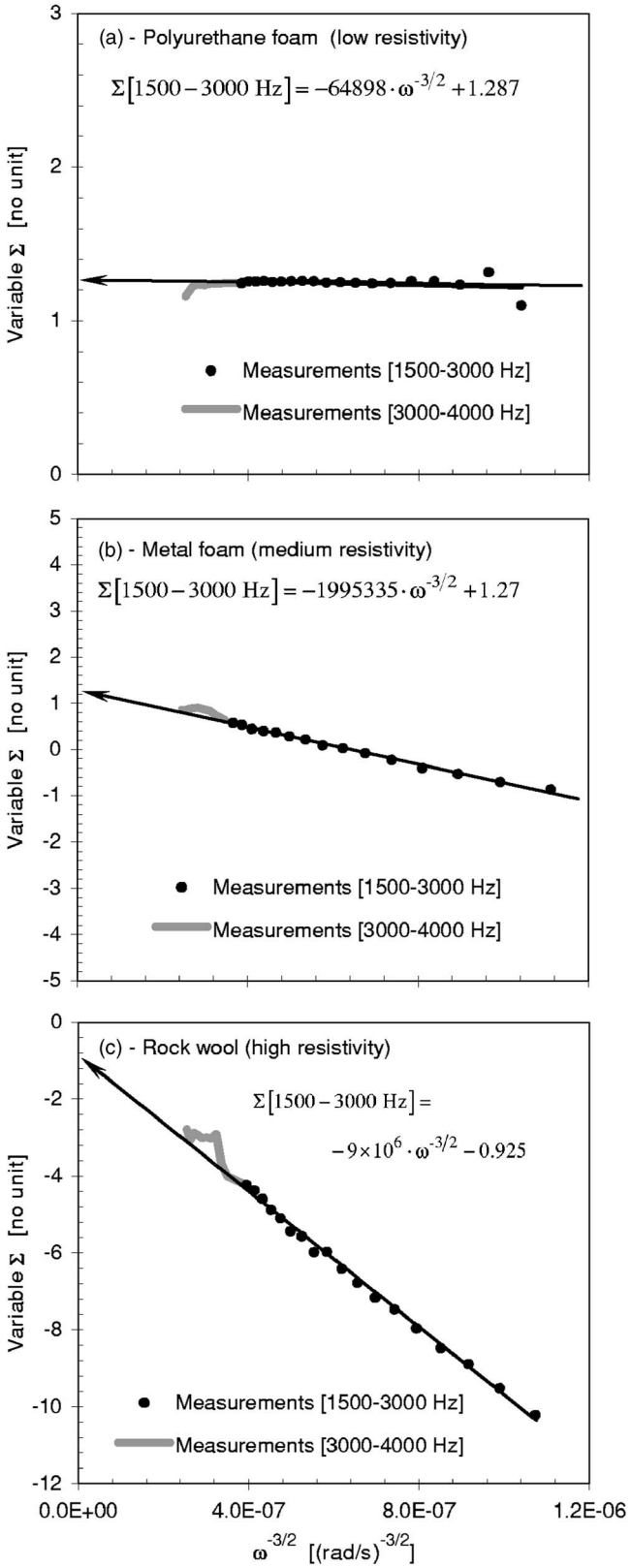


FIG. 3. Linear regression (→) applied on variable Σ as a function of $\omega^{-3/2}$. The linear regression is performed in the range 1500–3000 Hz.

2. Static airflow resistivity

σ is determined using the analytical solution (22) or the low-frequency limit of $\mathrm{Re}(\tilde{\sigma})$ given in (23). For the analytical solution, ϕ given in Table I and the previously found α_∞ are used. Figure 4 presents the results of the analytical solution (▲) and the real part of the dynamic resistivity (•) as a function of the frequency. The statistics (mean±standard deviation) in the range 1500–3000 Hz are given in Table II. Note that a unit value for α_∞ is used in (22) in the case of the high resistivity material—this value was adjusted so that the

TABLE II. Johnson *et al.* viscous parameters found using the extrapolation method. The statistics (mean ± standard deviation) are performed on the frequency range 1500–3000 Hz.

	Symbol	Polyurethane foam (low resistivity)	Metal foam (medium resistivity)	Rock wool (high resistivity)	Units
Geometrical tortuosity	α_∞	1.29	1.27	1.00	
Static airflow resistivity	σ	2 609±387	49 579±788	150 919+904	N s/m^4
Viscous characteristic dimension	Λ	202.8±6	20.0±0.3	7.1±0.1	μm

analytical solution fits the low-frequency limit of the real part of the dynamic resistivity. It is a typical value for fibrous material.

As observed in Fig. 4, the analytical solution is relatively constant at low frequencies and gives a good estimate of the directly measured σ(---). This is particularly true for the two higher resistivity materials, since their extraction range (△, ▼), shown in Fig. 1(c), is in the low-frequency viscous regime. For these two materials, the relative errors on the average static resistivities given in Table II are less than 1%. This is consistent with Fig. 1(c).

On the other hand, the analytical solution for the low resistivity material starts diverging rapidly from 2500 Hz. This is also consistent with its extraction range (▲) shown in Fig. 1(c). In this case, the relative error on the average resistivity is 13%. To reduce the error, the range should be limited to lower frequencies (e.g., 1500–2500 Hz).

For the high resistivity material, ω_ν being much larger than 4000 Hz, good estimates of σ are obtained in the whole studied frequency range by the two methods. Consequently, for highly resistive materials for which α_∞, cannot be obtained by extrapolation, σ can be first predicted using (23) and α_∞ adjusted so that the analytical solution (22) yields the same value.

3. Viscous characteristic dimension

So far, the extrapolation method yielded estimates for σ and α_∞. Using the found α_∞, the σ spectrum (▲) in Fig. 4, and slope m of the linear regression in Fig. 3, analytical solution (24) is used to obtain Λ. The statistics for the three materials are given in micrometer in Table II. Note that the value found for the high resistivity material should be underestimated since the amplitude of the slope in Fig. 3(c) is not good, and should be greater. This will be verified later on.

B. Analytical method—Johnson *et al.* model

1. Geometrical tortuosity

Using analytical solution (11) with the measured real X and imaginary Y parts of the equivalent dynamic density given in Fig. 2, and the properties given in Table I, α_∞ is determined. Its determination for the range 500–4000 Hz is shown in Fig. 5. For the two lower resistivity materials, one can note that the determined tortuosity is relatively constant with frequency. Statistics performed in the range 1500–3000 Hz are given in Table III for these two materials. The average geometrical tortuosities are α_∞=1.28 for the low

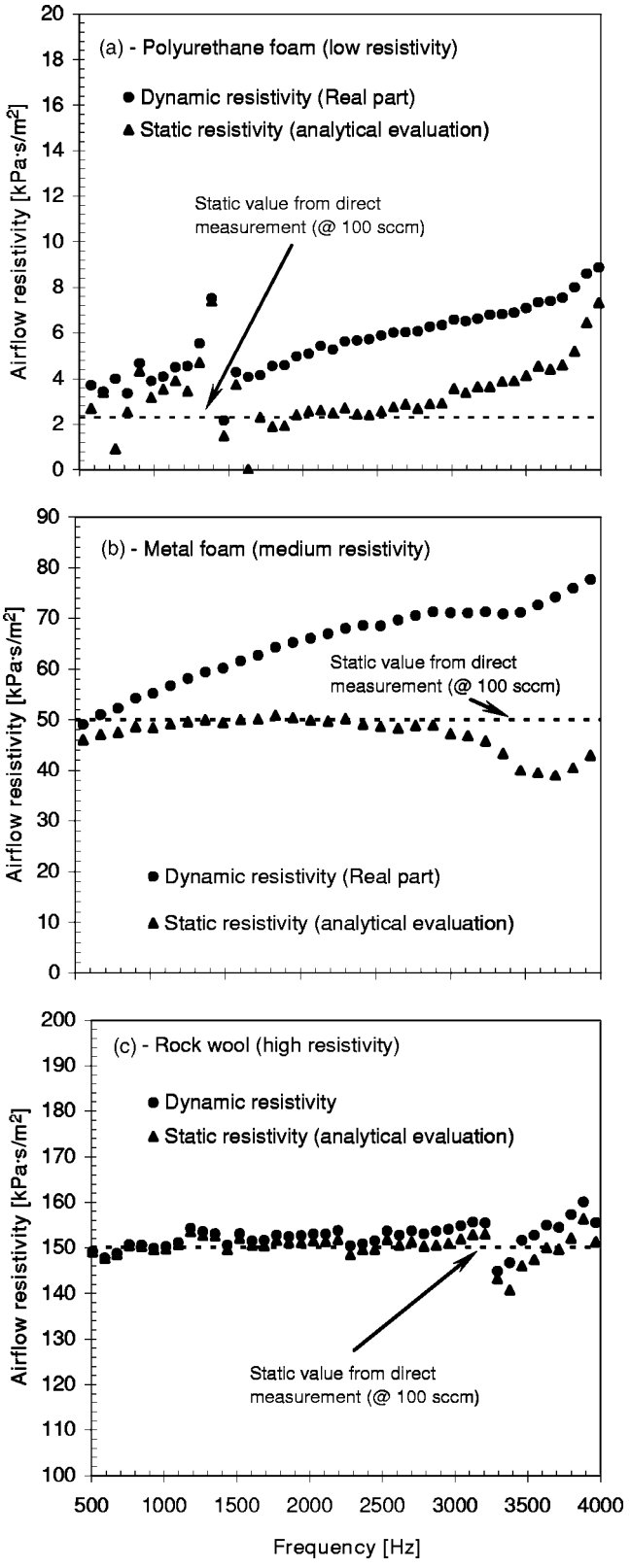


FIG. 4. Airflow resistivity as a function of the frequency. The real part of the dynamic airflow resistivity (•). Analytical solutions for the static airflow resistivity (▲).

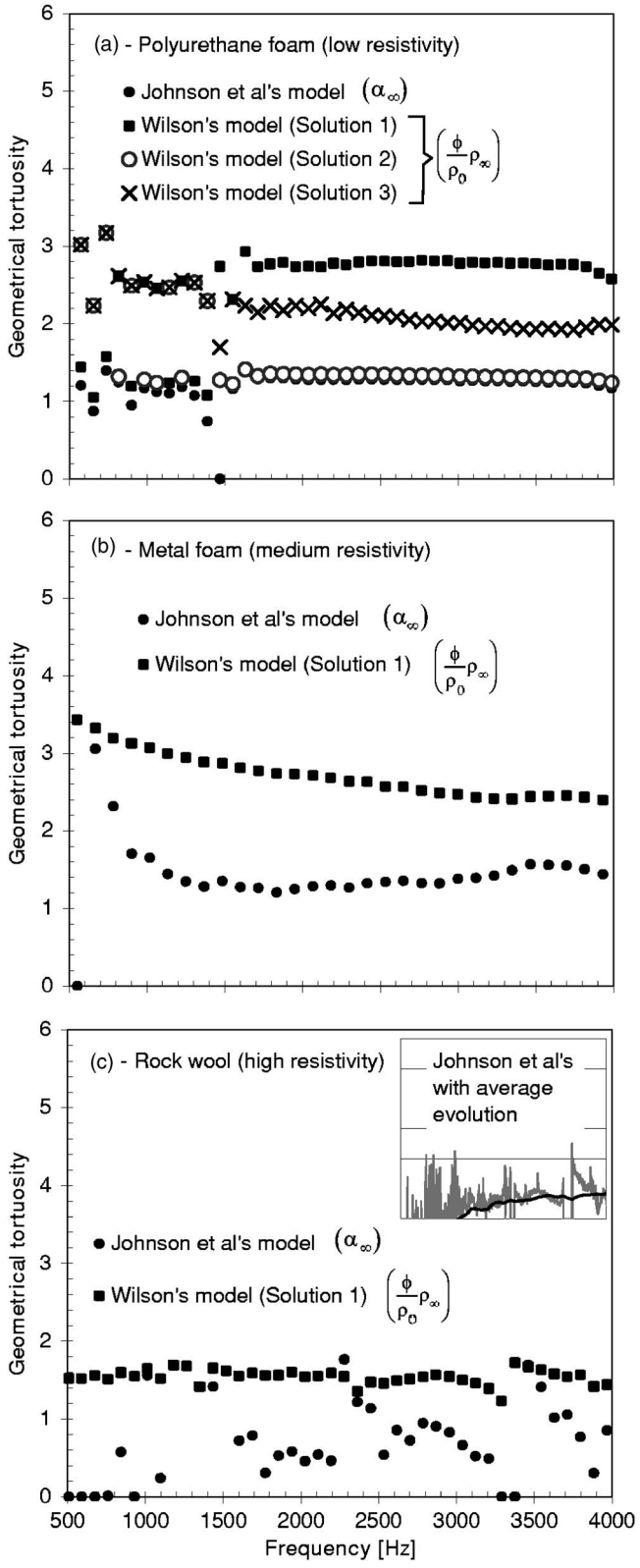


FIG. 5. Geometrical tortuosity as a function of the frequency. Analytical solutions for the Johnson *et al.* model (•) and Wilson model (■, ∘, ×). *x* axis for the inset in (c) ranges from 500 to 4000 Hz, and *y* axis from 0 to 6. In the inset, all the frequency points are plotted (gray solid line), and the average evolution (dark solid line) uses 255 frequency points per average.

resistivity material, and $\alpha_\infty = 1.29$ for the medium resistivity material. They are statistically identical to the values found with the extrapolation method—see Table II.

TABLE III. Johnson *et al.* viscous parameters found using the analytical method. Statistics (mean ± standard deviation) are performed on the frequency range 1500–3000 Hz.

	Symbol	Polyurethane foam (low resistivity)	Metal foam (medium resistivity)	Rock wool (high resistivity)	Units
Geometrical tortuosity	α_∞	1.28±0.01	1.29±0.04	1.00	
Viscous characteristic dimension	Λ	201.7±14.4	20.4±1.1	14.1±0.5	μm

For the high resistivity material, the results are noisy up to 4000 Hz and makes the estimation of α_∞ difficult. Nevertheless, if one looks at the average evolution (255 frequency points per average)—see the inset in Fig. 5(c), the evolution seems to converge slowly to a unit value. This value is similar to the one deduced with the extrapolation method. Consequently, its tortuosity is arbitrarily fixed to unity in Table III. Once again, extending the range up to 6000 Hz would probably help.

At low frequencies ($\varpi < 1$), inaccuracy in the measurements and errors due to the model (see Sec. V) may explain the divergence of the results and noisy data. Also, noisy data may be attributed to the evaluation of the root term in (11) when $Y \approx \sigma/\omega$—at the machine precision. More accuracy in the measurement of $\tilde{\rho}_{EQ}$ may alleviate the problem.

In the previous results, the constancy of α_∞ according to the frequency seems to indicate that the Johnson *et al.* model is appropriate, and correctly accounts for the viscoinertial behavior of the tested materials. However, as noted in Sec. V and Fig. 1, a systematic error (less than 1.5% in absolute value) could be induced for the two higher resistivity materials, since their extraction range (△, ▼) is below their characteristic viscous frequency ($\varpi < 1$). On the other hand, the low resistivity material is tested in its mid-frequency regime, where the error on the method is larger. However, since the statistics given in Table III is performed in the range 1500–3000 Hz (i.e., reduced frequency range 6.4–12.81), one can expect from Fig. 1(b) a relative error between −2.3% and −0.5%.

2. Viscous characteristic dimension

Using the properties given in Table I and the measured $\tilde{\rho}_{EQ}$ in Fig. 2, analytical solution (12) yields estimates for Λ. They are presented in Fig. 6. These estimates either use the geometrical tortuosity spectrum of Fig. 5, or the average value of Table III. In both cases, constancy in Λ is observed in the range 1500–3000 Hz. However, using the average value for α_∞, the constancy is stressed. This is particularly true for the high-resistivity material. For the latter, the noisy results stems from the noisy data in the tortuosity spectrum in Fig. 5(c).

For the two lower resistive materials, statistics on Λ in the range 1500–3000 Hz are given in Table III. The average values are very close to those obtained with the extrapolation method. For the high resistivity material, the average value

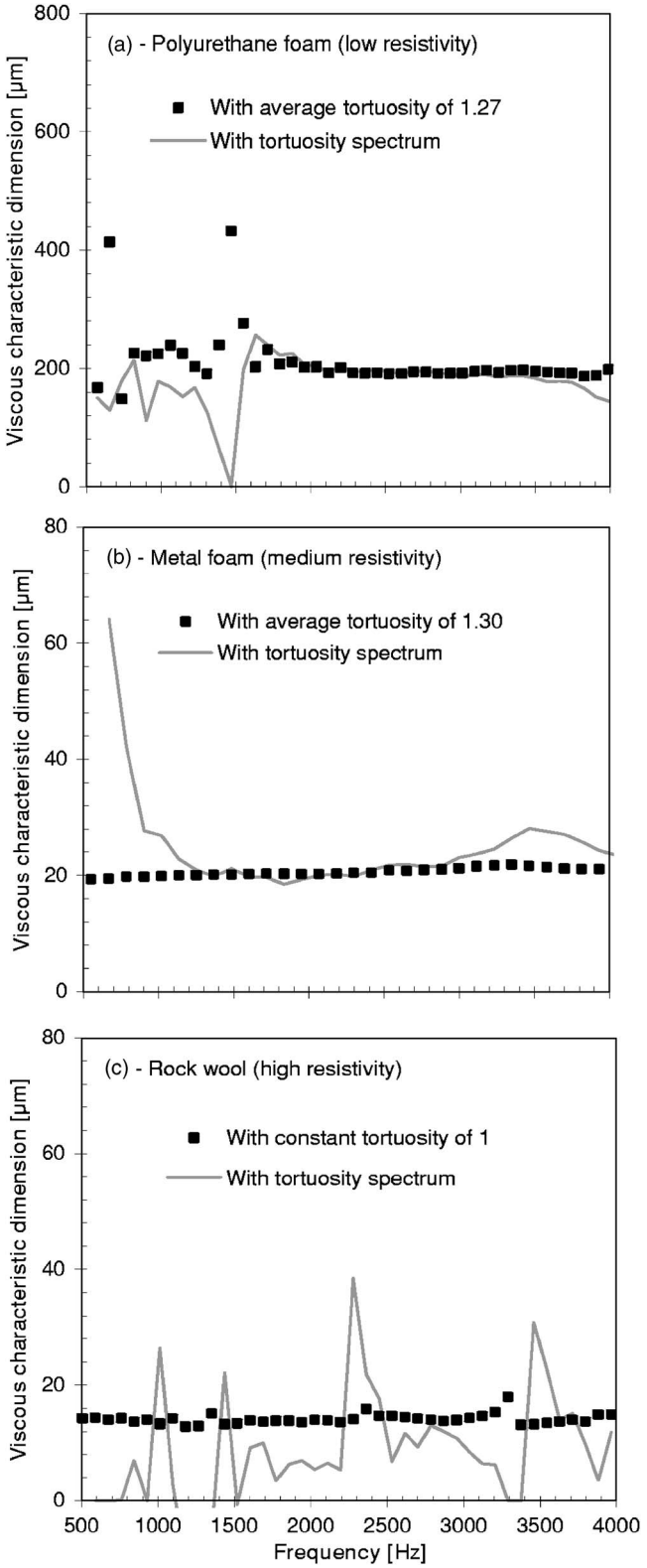


FIG. 6. Viscous characteristic dimension as a function of the frequency. Analytical solutions for the Johnson *et al.* model using the constant-average tortuosity values of Table III (•), and the tortuosity spectra of Fig. 5 (solid line).

for Λ is twice the value found with the extrapolation method. This is in agreement with the fact the slope m in Fig. 3(c) is underestimated—see the discussion in Sec. VI A 3.

As mentioned previously, the low resistivity material is tested in its mid-frequency regime, where the error on the method is larger. However, since the statistics given in Table III are performed in the frequency range 1500–3000 Hz (i.e., reduced frequency range 6.4–12.81), one can expect from Fig. 1(b) a relative error between −17% and −10%. Similarly, for the two other materials, the characterization operates in their low-frequency regime. In this case, Fig. 1(b) predicts a systematic relative error between −13.5% to 4%.

C. Analytical method—Wilson model

1. Density parameter

Substituting the real X and imaginary Y parts of the measured equivalent dynamic density into analytical solution (14), estimates are obtained for ρ_∞. To facilitate the comparison with the Johnson *et al.* geometrical tortuosity, the following relation is used: $\rho_\infty = \rho_0 \alpha_\infty / \phi$. Using this relation, one finds corresponding α_∞ values for given ρ_∞ values. These corresponding values are added to the geometrical tortuosity spectra in Fig. 5.

As noted in Fig. 5(a), analytical solution (14) yields three possible solutions for the Wilson density parameter. The second solution yields a spectrum similar to the one obtained for the Johnson *et al.* tortuosity. Since the material has a low viscous characteristic frequency, the existence of three possible solutions may be explained by the fact the imaginary part Y of $\tilde{\rho}_{\mathrm{EQ}}$ is small enough in the studied frequency range so that variable A in (16) is always real and positive. This is not true for the two other materials for which the viscous characteristic frequencies are greater than the studied frequency range.

Similar to the Johnson *et al.* model, results found for ρ_∞ are relatively constant with the frequency. Except for the low resistivity material, no simple correspondence between ρ_∞ and α_∞ is found. Consequently, it may be inappropriate to use relation $\rho_\infty = \rho_0 \alpha_\infty / \phi$ with the Wilson model, as proposed by the author. A characterization, as the one proposed here, is a secure method to infer ρ_∞ in the Wilson model.

The statistics on ρ_∞ in the range 1500–3000 Hz are given in Table IV. Note that for the low resistivity material, only the second solution is given. In the next section, it will be shown that only the second solution is admissible.

2. Vorticity-mode relaxation time

By the use of an analytical solution (15), the vorticity-mode relaxation time τ_ν is found. For the low resistivity material, the first and third solutions previously obtained for ρ_∞ yield negative values for τ_ν. Consequently, they cannot be admissible solutions; only the second solution yields real positive values for τ_ν. The results obtained by (15) using the average ρ_∞ in Table IV are shown in Fig. 7. Also, the Johnson *et al.* vorticity-mode relaxation time τ_{J}, given by (5), is plotted in Fig. 7.

The statistics over the range 1500–3000 Hz are given in Table IV. One can note that τ_ν and τ_{J} are of the same order of magnitude, and very close for the high resistivity material. Moreover, τ_ν seems relatively constant with frequency—especially for higher resistivity values. However, as for ρ_∞, it may be inappropriate to use relation $\tau_\nu = \rho_0 \Lambda^2 / 2\eta$ or τ_ν

TABLE IV. Wilson viscous parameters found using the analytical method. Statistics (mean ± standard deviation) are performed on the frequency range 1500–3000 Hz.

		Symbol	Polyurethane foam (low resistivity)	Metal foam (medium resistivity)	Rock wool (high resistivity)	Units
Vorticity-mode relaxation time	Density parameter	ρ_∞	1.62±0.04	3.41±0.15	1.94±0.08	kg/m^3
	Wilson	τ_ν	1.24±0.17	0.11±0.01	0.03±0.00	ms
	Johnson *et al.*	τ_J	0.69±0.09	0.17±0.01	0.03±0.01	ms
	Wilson (approx.)	$\rho_0\Lambda^2/2\eta$	1.28	0.013	0.006	ms
	Wilson (approx.)	$2\rho_0\alpha_\infty/\phi\sigma$	1.356	0.066	0.017	ms

$=2\rho_0\alpha_\infty/\phi\sigma$ in the Wilson model. In fact, Table IV gives the values $\rho_0\Lambda^2/2\eta$ and $2\rho_0\alpha_\infty/\phi\sigma$ for the three materials using the properties given in Table I and Table III. For higher resistivity materials, one can note that these values may be one order of magnitude lower than the values found using (15)—here we assume the values in Table III are good. Consequently, it seems more appropriate to use the following relation $\tau_\nu \approx \tau_J$, or to use the proposed analytical equation (15) as a secure method to infer the value of τ_ν in the Wilson model.

D. Comparison to measurements

Now, let us compare to measurements the equivalent dynamic density predicted by Johnson *et al.* [Eq. (6)] and by Wilson [Eq. (9)]—see Fig. 8. The average parameters in Table III and Table IV and the properties in Table I are used for the predictions. One can note that the predicted results for both models compare well with measurements.

In Fig. 8(a), for the low resistivity material, both models yield similar results. This may be explained by the fact that the relation $\rho_\infty=\rho_0\alpha_\infty/\phi$ is verified in this case—see Fig. 5(a). This is logical since the studied frequency range is in the mass-controlled region ($\varpi>1$), where the effect of tortuosity dominates over the vorticity-relaxation process. Also, one can note that the measured resonant behavior at low frequencies is not predicted due to the rigid-frame assumption of the predictive models. In this case, the Biot model[1,17,18] should be used to predict the resonant behavior.

In Fig. 8(b), for the medium-resistivity material, a better agreement is obtained with the Johnson *et al.* model. As expected, the Johnson *et al.* model underestimates the real part at low frequencies. The divergence between the two models may be attributed to the fact neither $\rho_\infty=\rho_0\alpha_\infty/\phi$ nor $\tau_\nu=\tau_J$ are verified since the studied frequency range is close to the medium frequency behavior of the material ($\varpi\approx1$).

In Fig. 8(c), for the high-resistivity material, the two models yield identical results and compare very well with measurements. In this case, where the low-frequency viscosity-controlled regime dominates, the Wilson's relaxation time is identical to the Johnson *et al.* relaxation time ($\tau_\nu=\tau_J$)—see Table IV. Also, contrary to the medium resistivity material, no underestimation of the real part is observed for the Johnson *et al.* model. This may be explained by the fact the Johnson *et al.* model is more accurate for this fiber-like material of unit tortuosity.

VII. CONCLUSION

In this work, acoustical methods were derived from the Johnson *et al.* and Wilson dynamic density models to characterize the parameters governing viscous dissipation in open-cell porous media. The methods were tested on three materials covering a wide range of static flow resistivities (2300–150 100 Ns/m^4), frame rigidities (soft and rigid), and pore geometries (cells and fibers). To operate, the methods require an acoustical measurement of the dynamic density of the tested material.

From the Johnson *et al.* model, two methods were derived: (i) an analytical method and (ii) an extrapolation method. To use the analytical method, the static airflow resistivity and open porosity of the material must be known. To use the extrapolation method, only the open porosity needs to be known. Both methods allow characterization of the geometrical tortuosity and viscous characteristic dimension. In addition, the extrapolation method can be used to characterize the static airflow resistivity of the material. Tests on the three porous materials have shown that both methods yield statistically the same results. Moreover, the static flow resistivities found from the extrapolation method were very close to the directly measured ones.

To estimate the precision of the analytical method using the Johnson *et al.* model, the method was applied on two virtual materials of theoretically known properties (dynamic density, geometrical tortuosity, viscous characteristic dimension, and static airflow resistivity). From this study, it was found that the method should yield small relative errors on the tortuosity (maximum error less than 2% in absolute value). Similarly, small relative errors (less than 1% in absolute value) are expected for the static airflow resistivity if the extraction proceeds at low frequencies. For the viscous characteristic dimension, if the method is applied in the low-frequency viscosity-controlled regime, a systematic error is introduced. From the tested virtual materials, this systematic error is expected to be no more than 20% (in absolute value)—which is quite acceptable since this parameter is difficult to obtain with accuracy using other existing methods. However, if the method is applied in the higher-frequency mass-controlled regime, the error tends rapidly to zero. In addition to the assessment of the precision of the method, Fig. 1 obtained from this error analysis can be used to select the frequency range in which the extraction should proceed to minimize the error of the method.

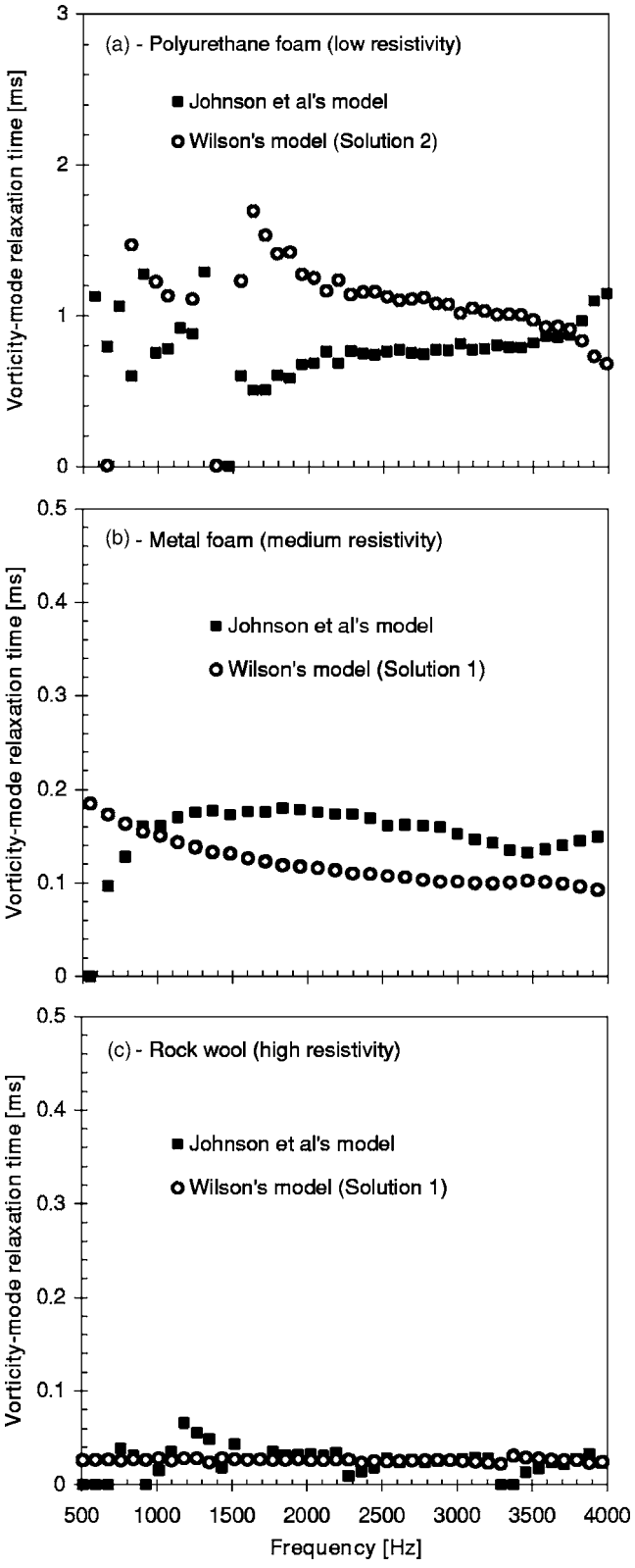


FIG. 7. Vorticity-mode relaxation time in function of the frequency. Analytical solutions for the Wilson model (◦), and the Johnson *et al.* model (•).

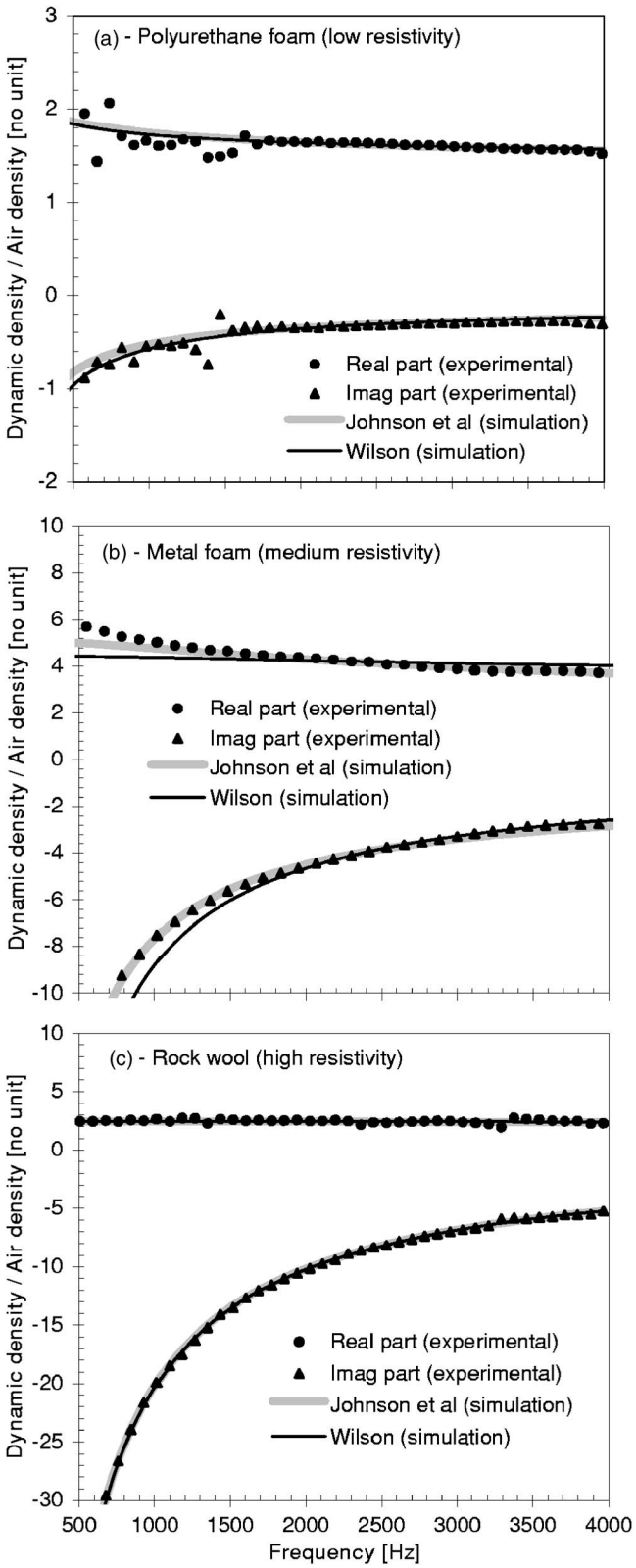


FIG. 8. Equivalent dynamic density in function of the frequency. A comparison between experimentation (•, ▲), the Johnson *et al.* model (—), and the Wilson model (—).

For the Wilson model, the derived analytical method was found to be a secure characterization method to infer representative values for ρ_∞ and τ_ν in the Wilson dynamic density. Also, it was shown that relations $\rho_\infty=\rho_0\alpha_\infty/\phi$ and, more particularly, $\tau_\nu=\rho_0\Lambda^2/2\eta$ or $\tau_\nu=2\rho_0\alpha_\infty/\phi\sigma$, as proposed by Wilson, may be misleading. For the second relation, it was found to be more appropriate to use $\tau_\nu\approx\tau_J$. One advantage of the Wilson model over the Johnson *et al.* model

is that the characterization of the two Wilson's parameters does not require prior knowledge of any other physical properties.

Moreover, it was found that both models, once their parameters are carefully characterized, predict similar dynamic densities and compare well with measurements. Also, for both models, relative constancy in the characterized parameters in the function of the frequency was noted. This constancy can be used to assess the validity of (1) the descriptive models in a given frequency range; and (2) the characterized parameters found from the proposed characterization methods. Also, it supports the relaxation process as suggested by Wilson and the Johnson *et al.* dissipation function G_J.

Once again, these good correlations with measurements, together with the noted constancy in the found parameters with the frequency, reinforce the fact that the proposed characterization methods offer an elegant alternative to existing characterization methods. Moreover, since the methods only rely on equations and a widespread apparatus (impedance tube), this makes the characterization of porous materials possible for many acoustic laboratories.

To conclude, the authors believe that comparisons with ultrasound techniques are necessary to complete this work—this was not possible here using the available laboratory equipments. Also, similar methods[15] could be developed for characterizing the parameters relative to the thermal losses (i.e. thermal characteristic dimension, static thermal permeability, Wilson entropy-mode relaxation time, Wilson compressibility parameter). To improve the reliability of the proposed methods, work on improving the accuracy of the measurements of the dynamic properties (characteristic impedance, propagation constant, dynamic density) and on the minimization of frame vibrations, is necessary.

ACKNOWLEDGMENTS

We would like to thank Dr. Keith Wilson for his valuable comments; and N.S.E.R.C. Canada, F.Q.N.R.T. Quebec, C.Q.R.D.A. Quebec, and REGAL for their financial support.

APPENDIX A: THEORETICAL MODELS

In this appendix we recall the theoretical expressions of the dynamic density for (i) a porous material with oblique circular cylindrical pores, and (ii) a porous material with oblique rectangular slits.

1. Porous material with oblique circular cylindrical pores

The equivalent theoretical dynamic density of a layer having parallel oblique cylindrical pores of a circular cross section is given by

$$\tilde{\rho}_{\mathrm{EQ}}^{\mathrm{cyl}} = \frac{\rho_0 \alpha_\infty}{\phi}\left(1 - j\frac{1}{\varpi}G_c(\varpi)\right), \tag{A1}$$

with

$$G_c = -\frac{s\sqrt{-j}}{4}\frac{J_1(s\sqrt{-j})}{J_0(s\sqrt{-j})}\Bigg/\left(1 - \frac{2}{s\sqrt{-j}}\frac{J_1(s\sqrt{-j})}{J_0(s\sqrt{-j})}\right) \tag{A2}$$

and

$$s = (8\varpi)^{1/2}. \tag{A3}$$

In (A2), J_0 and J_1 are the Bessel functions of the first kind of order 0 and 1. Considering n pores of radius R per unit area of surface, the open porosity, tortuosity, and static flow resistivity are, respectively,

$$\phi = \frac{n\pi R^2}{\cos\theta}, \tag{A4}$$

$$\alpha_\infty = \frac{1}{\cos^2\theta}, \tag{A5}$$

$$\sigma = \frac{8\eta}{n\pi R^4 \cos\theta}, \tag{A6}$$

where θ is the angle between the axis of the oblique pores and the normal to the layer.

2. Porous material with oblique rectangular slits

The equivalent theoretical dynamic density of a layer having parallel oblique rectangular slits is given by

$$\tilde{\rho}_{\mathrm{EQ}}^{\mathrm{slit}} = \frac{\rho_0 \alpha_\infty}{\phi}\left(1 - j\frac{1}{\varpi}G_s(\varpi)\right), \tag{A7}$$

with ϖ given by

$$G_s = \frac{s'\sqrt{j}}{3}\tanh(s'\sqrt{j})\Bigg/\left(1 - \frac{\tanh(s'\sqrt{j})}{s'\sqrt{j}}\right) \tag{A8}$$

and

$$s' = (3\varpi)^{1/2}. \tag{A9}$$

Considering n slits of width $2a$ per unit length, the open porosity and static flow resistivity are, respectively,

$$\phi = \frac{2na}{\cos\theta}, \tag{A10}$$

$$\sigma = \frac{3}{2}\frac{\eta}{na^3\cos\theta}. \tag{A11}$$

The tortuosity is given by (A5).

[1] J.-F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Elsevier Applied Science, New York, 1993).
[2] 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).
[3] S. R. Pride, F. D. Morgan, and A. F. Gangi, "Drag forces of porous-medium acoustics," Phys. Rev. B **47**, 4964–4975 (1993).
[4] D. K. Wilson, "Relaxation-matched modeling of propagation through porous media, including fractal pore structure," J. Acoust. Soc. Am. **94**, 1136–1145 (1993).
[5] M. R. Stinson, and G. A. Daigle, "Electronic system for the measurement of flow resistance," J. Acoust. Soc. Am. **83**, 2422–2428 (1988).
[6] Standard test method for airflow resistance of acoustical materials, ASTM C 522-80, 1980.

[7]R. J. S. Brown, "Connection between formation factor for electrical resistivity and fluid-solid coupling factor in Biot equations for acoustic waves in fluid-filled media," Geophysics **45**, 1269–1275 (1980).
[8]D. L. Johnson, T. J. Plona, C. Scala, F. Pasierb, and H. Kojima, "Tortuosity and acoustic slow waves," Phys. Rev. Lett. **49**, 1840–1844 (1982).
[9]Y. Atalla and R. Panneton, "Inverse acoustical characterization of open-cell porous media using impedance tube measurements," accepted for publication in Can. Acoust. **33**(1), 11–24 (2005).
[10]N. Sellen, M.-A. Galland, and O. Hilbrunner, "Identification of the characteristic parameters of porous media using active control," *8th AIAA/CEAS Aeroacoustics Conference*, AIAA Paper 2002-2504, 2002.
[11]J. F. Allard, B. Castagnede, and M. Henry, "Evaluation of tortuosity in acoustic porous materials saturated by air," Rev. Sci. Instrum. **65**, 754–755 (1994).
[12]P. Leclaire, L. Kelders, W. Lauriks, M. Melon, N. Brown, and B. Castagnède, "Determination of the viscous and thermal characteristic lengths of plastic foams by ultrasonic measurements in helium and air," J. Appl. Phys. **80**, 2009–2012 (1996).
[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]R. Panneton, X. Olny, and J. Tran Van, "Analytical solutions for characterizing tortuosity and characteristic lengths of porous media using acoustical measurements: Extrapolation model," ASA 142nd Meeting, 2001.
[15]X. Olny, J. Tran Van, and R. Panneton, "Analytical solution for characterizing tortuosity and characteristic lengths: Indirect model," ASA 142nd Meeting, 2001.
[16]P. Leclaire, L. Kelders, W. Lauriks, J. F. Allard, and C. Glorieux, "Ultrasonic wave propagation in reticulated foams saturated by different gases: High frequency limit of the classical models," Appl. Phys. Lett. **69**, 2641–2643 (1996).
[17]M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[18]M. A. Biot, "The theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher-frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[19]K. Attenborough, "On the acoustic slow wave in air-filled granular media," J. Acoust. Soc. Am. **81**, 93–102 (1982).
[20]J. F. Allard, C. Depollier, and A. Aknine, "Acoustical properties of partially reticulated foams with high and medium flow resistance," J. Acoust. Soc. Am. **79**, 1734–1740 (1986).
[21]R. Panneton and E. Gros, "A missing mass method to measure the open porosity of porous solids," Acta Acust. (Beijing) **91**, 342–348 (2005).
[22]Y. Champoux, M. R. Stinson, and G. A. Daigle, "Air-based system for the measurement of the porosity," J. Acoust. Soc. Am. **89**, 910–916 (1990).
[23]H. Utsuno, T. Tanaka, and T. Fujikawa, "Transfer function method for measuring characteristic impedance and propagation constant of porous materials," J. Acoust. Soc. Am. **86**, 637–643 (1989).
[24]T. Iwase, Y. Izumi, and R. Kawabata, "A new measuring method for sound propagation constant by using sound tube without any air spaces back of a test material," paper read at *Internoise 98*, Christchurch, New Zealand, 1998.
[25]O. C. Zwikker and C. W. Kosten, *Sound-Absorbing Materials* (Elsevier, Amsterdam, 1949).
[26]P. Giacomo, "Equation for the determination of the density of moist air (1981)," Metrologia **18**, 33–40 (1982).
[27]R. S. Davis, "Equation for the Determination of the Density of Moist Air (1981/91)," Metrologia **29**, 67–70 (1992).
[28]D. Pilon, R. Panneton, and F. Sgard, "Behavioral criterion quantifying the edge-constrained effects on foams in the standing wave tube," J. Acoust. Soc. Am. **114**, 1980–1987 (2003).
[29]D. Pilon, R. Panneton, and F. Sgard, "Frame acoustical excitability: a decoupling criterion for poroelastic materials," *Proceedings of Euronoise 2003*, Paper ID: 9.

Scattering of Rayleigh-Lamb waves by a surface breaking crack in an elastic plate

Miguel A. Flores-López[a)]
Colegio de Ciencia y Tecnología, Universidad Autónoma de la Ciudad de México, México D.F., México

R. Douglas Gregory
Department of Mathematics, The University of Manchester, M13 9PL, United Kingdom

(Received 31 October 2005; revised 2 February 2006; accepted 6 February 2006)

This work studies the problem of scattering of Rayleigh-Lamb waves by a thin surface-breaking crack. The problem is solved by the projection method, which exploits some orthogonality relations deduced from the reciprocal theorem. Numerical results obtained by the projection method are compared with the results obtained using a finite element method reported by other authors for the case of a flat crack, finding good agreement. From the numerical results it is observed that, as the wave number of the incident wave, S_0 for instance, is smaller, the reflected and transmitted energy corresponding to the converted mode, A_0 in this case, are almost equal to each other. An explanation for this result is given, based on the asymptotic nature of the modes S_0 and A_0 for small wave numbers. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2180209]

PACS number(s): 43.20.Mv, 43.20.Gp [YHB] Pages: 2041–2049

I. INTRODUCTION

The scattering problem of Rayleigh-Lamb waves is important in the field of nondestructive evaluation, where elastic waves are used to evaluate material properties as well as to locate and measure defects in critical structures such as tubing in plants, pipelines in chemical processing facilities, and so on. Guided-waves are being used to find tiny defects over large distances, in structures with insulations and coatings. Chimenti[1] made a review of such waves employed to evaluate composite materials and to detect defects in, and the integrity of bonded joints. In his work there is an interesting account of the application of nondestructive evaluation in the aviation industry.

Among the work done for solving some important problems involving Rayleigh-Lamb waves in the last years, it is worth mentioning the following: the Wiener-Hopf technique was used by Rokhlin[2] who studied the case of a horizontal thin crack on the midsurface of the plate. A combination of normal mode expansions and finite element methods was employed by Al-Nassar *et al.*,[3] who studied the scattering of Rayleigh-Lamb waves by a normal strip welded joint in an isotropic plate. Alleyne and Cawley[4] studied numerically, using finite element techniques, the interaction of Rayleigh-Lamb modes with notches in elastic and isotropic plates. They also undertook some experimental measurements which confirmed their numerical results. With the use of a Fourier transform technique they quantified the wave interaction with the defects as function of parameters such as frequency-thickness product, the type of incoming mode and the geometry of the notch. Using modal decomposition and collocation (imposing conditions along the crack), Castaings *et al.*[5] studied the interaction of Rayleigh-Lamb waves with breaking cracks and with internal cracks symmetric with respect to the midplane of the plate. A hybrid method, called a mode-exciting method, which solves the problem in a finite region containing the crack by a numerical method (finite element method or boundary element method), combined with mode expansion in the complementary infinite plate was presented by Gunawan and Hirose.[6]

Other approaches to the scattering of Rayleigh-Lamb waves, based on boundary elements, have been presented by Cho *et al.*[7] and by Cho and Rose,[8] both for a finite surface breaking crack. Lowe and Diligent[9] used the finite element method and some experimental work, to study in particular the reflection of the fundamental symmetric Rayleigh-Lamb mode from a rectangular notch in an elastic and isotropic plate. Problems involving composite materials have been addressed by Datta *et al.*[10] and by Karunasena *et al.*[11]

This work studies the scattering problem of a Rayleigh-Lamb wave by a thin surface-breaking crack in an isotropic elastic plate, using the projection method. The crack is assumed to be orthogonal to the surface of the plate. The method of projection was introduced by Gregory and Gladwell,[12] for computing stress and strain fields in a cantilever beam. Later it was also applied to problems in wave dynamics by Gregory and Gladwell,[13] to determine reflected Rayleigh-Lamb waves at the edge of a semi-infinite elastic plate. An improved version of this method was presented by Gregory *et al.*,[14] for computing effective shear and moment resultants of a cantilever strip plate.

II. RAYLEIGH-LAMB WAVES IN A PLATE WITH A CRACK

This section introduces the problem of a Rayleigh-Lamb wave traveling from $-\infty$ to the right in a plate with a crack orthogonal to its surface, as shown in Fig. 1. For definiteness it is assumed that the incident wave corresponds to the fundamental (zeroth) symmetric mode, S_0.

[a)]Electronic mail: maflores@math.cinvestav.mx

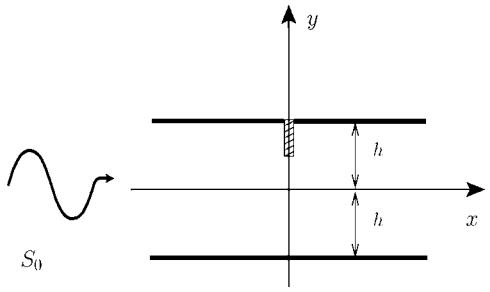


FIG. 1. The geometry for the plate and the crack.

It is supposed that the crack lies on the y axis with its tip at the point $(0,a)$. When the incident wave interacts with the crack it produces a scattered wave. The displacement of the scattered wave is denoted by $\mathbf{u}$ and by $\boldsymbol{\tau}$ the stress. The plate is considered to be elastic, isotropic and the linear theory equations may be used for the field variables.

A. Rayleigh-Lamb modes

Rayleigh and Lamb,[15] both studied the case when waves of plane strain (no z dependence) propagate in the plate subject to free surface conditions,

$$\tau_{yy} = \tau_{xy} = \tau_{zy} = 0, \quad y = \pm h. \tag{1}$$

The displacement vector, $(u,v)^T$, can be split into symmetric and antisymmetric components with respect to the midsurface $y=0$.

In the case of the symmetric modes (longitudinal modes) the frequency relation is

$$\frac{\tanh \delta}{\tanh \gamma} = \frac{4\alpha^2\gamma\delta}{(2\alpha^2 - K^2)^2}, \tag{2}$$

and for the antisymmetric modes (flexural modes)

$$\frac{\tanh \delta}{\tanh \gamma} = \frac{(2\alpha^2 - K^2)^2}{4\alpha^2\gamma\delta}, \tag{3}$$

where $\gamma = \sqrt{\alpha^2 - k^2}$, and $\delta = \sqrt{\alpha^2 - K^2}$, with $k = \omega/c_L$ and $K = \omega/c_T$. Here c_L is the velocity of the compressional wave and c_T of the transversal in the elastic plate.

For symmetric modes the displacements are given by

$$u(x,y,t) = U_n^S(y)\exp[i(\alpha_n x - \omega t)],$$

$$v(x,y,t) = V_n^S(y)\exp[i(\alpha_n x - \omega t)],$$

with

$$U_n^S(y) = i\alpha_n \cosh(\gamma_n y) + \delta_n B_n \cosh(\delta_n y), \tag{4}$$

$$V_n^S(y) = \gamma_n \sinh(\gamma_n y) - i\alpha_n B_n \sinh(\delta_n y), \tag{5}$$

where

$$B_n = -i(2\alpha_n^2 - K^2)\cosh \gamma_n / 2\alpha_n\delta_n \cosh \delta_n,$$

$\gamma_n = \gamma(\alpha_n)$, and $\delta_n = \delta(\alpha_n)$. The stresses are similarly

$$\tau_{xx} = \mu R_n^S(y)\exp[i(\alpha_n x - \omega t)],$$

and

$$\tau_{xy} = \mu S_n^S(y)\exp[i(\alpha_n x - \omega t)],$$

where μ is the shear modulus of the plate and

$$R_n^S(y) = -(2\gamma_n^2 + K^2)\cosh(\gamma_n y) + 2i\alpha_n\delta_n B_n \cosh(\delta_n y), \tag{6}$$

and

$$S_n^S(y) = 2i\alpha_n\gamma_n \sinh(\gamma_n y) + (2\alpha_n^2 - K^2)B_n \sinh(\delta_n y). \tag{7}$$

The corresponding field variables for antisymmetric modes $U_n^A(y)$, $V_n^A(y)$, R_n^A, and S_n^A are obtained from the above by replacing cosh by sinh and sinh by cosh everywhere.

The following notation is used to describe the displacement and stress vectors of the nth symmetric and antisymmetric modes, with the term $e^{-i\omega t}$ factored out. By

$$\mathbf{U}_n^P \exp(i\alpha_n x)$$

and

$$\mathbf{S}_n^P \exp(i\alpha_n x),$$

will be denoted the displacement and the stress vectors corresponding to the nth mode, where if $P=S$ it will denote the symmetric mode and when $P=A$ it will be the antisymmetric mode.

For the case of the mode P_n, where P can be S or A according to the parity of the mode, the mean energy flux across a surface orthogonal to the plate is given by

$$\langle \mathcal{P}_n^P \rangle = \frac{\omega\mu}{2}\int_{-h}^{h} \mathrm{Im}\{\mathbf{S}_n^P(y)\cdot\mathbf{U}_{-n}^P(y)\}dy, \tag{8}$$

where the subscript $-n$ denotes the corresponding mode traveling to the left of the x axis.

It will be convenient to introduce the expressions

$$J_n = 2\int_{-h}^{h} \mathrm{Im}\{\mathbf{S}_n^S(y)\cdot\mathbf{U}_{-n}^S(y)\}dy \tag{9}$$

and

$$K_n = 2\int_{-h}^{h} \mathrm{Im}\{\mathbf{S}_n^A(y)\cdot\mathbf{U}_{-n}^A(y)\}dy. \tag{10}$$

An application of the reciprocal theorem (Gregory and Gladwell[13]) produces

$$\int_{-h}^{h} \{\mathbf{S}_m^{P_1}(y)\cdot\mathbf{U}_n^{P_2} - \mathbf{S}_n^{P_2}(y)\cdot\mathbf{U}_m^{P_1}\}dy = \begin{cases} 0 & \text{if } n \neq -m, \text{ or } P_1 \neq P_2, \\ 2J_m & \text{if } n = -m,\ P_1 = P_2 = S, \\ 2K_m & \text{if } n = -m,\ P_1 = P_2 = A. \end{cases}$$

The last equation represents a bi-orthogonality relation among the Rayleigh-Lamb modes.

B. Basic equations

Let u and v be the x and y components of the displacement and τ_{xx}, τ_{yy}, and τ_{xy} the relevant stress components. Assuming $\mathbf{u}$ to be time-harmonic, the equation governing the displacement becomes

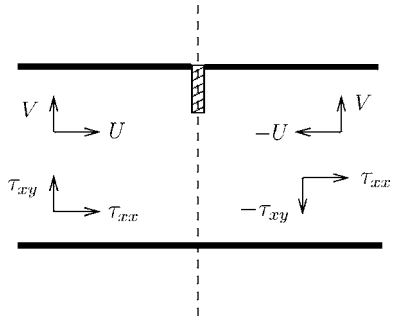


FIG. 2. The even part of displacements and the corresponding stresses.

$$k^2\nabla^2\mathbf{u} + (K^2 - k^2)\,\nabla\,\nabla\cdot\mathbf{u} + k^2K^2\mathbf{u} = 0, \tag{11}$$

where the wave numbers k and K are defined by $k=\omega/c_L$ and $K=\omega/c_T$, with ω the angular frequency.

Then the boundary value problem for $\mathbf{u}$ can be defined by Eq. (11), together with the free boundary conditions

$$\tau_{yy} = \tau_{xy} = \tau_{zy} = 0, \quad y = \pm h, \tag{12}$$

and the crack conditions

$$\tau_{xx} = \tau_{xy} = 0, \quad \text{for } x = 0,\ a \leqslant y \leqslant h. \tag{13}$$

In order to apply the projection method on a semi-infinite domain, it is convenient to convert this problem into a boundary value problem in the domain $x \geqslant 0, -h \leqslant y \leqslant h$. This is done by considering the following: The incident wave, $\mathbf{u}^{\text{inc}}$, has components

$$\mathbf{u}^{\text{inc}} = \begin{pmatrix} U_0^S(y) \\ V_0^S(y) \end{pmatrix} \exp[i(\alpha_0 x - \omega t)],$$

where U_0^S and V_0^S are the expressions (4) and (5), evaluated for the first symmetric wave number α_0. It is noticed that $\mathbf{u}^{\text{inc}}$ can be split as

$$\mathbf{u}^{\text{inc}} = \mathbf{u}_E^{\text{inc}} + \mathbf{u}_O^{\text{inc}},$$

where

$$\mathbf{u}_E^{\text{inc}} = \frac{1}{2}\left[\begin{pmatrix} U_0^S(y) \\ V_0^S(y) \end{pmatrix} \exp(i\alpha_0 x) + \begin{pmatrix} -U_0^S(y) \\ V_0^S(y) \end{pmatrix} \exp(-i\alpha_0 x) \right]$$

and

$$\mathbf{u}_O^{\text{inc}} = \frac{1}{2}\left[\begin{pmatrix} U_0^S(y) \\ V_0^S(y) \end{pmatrix} \exp(i\alpha_0 x) - \begin{pmatrix} -U_0^S(y) \\ V_0^S(y) \end{pmatrix} \exp(-i\alpha_0 x) \right],$$

and the subscripts E and O stand for even and odd. Therefore, $\mathbf{u}_E^{\text{inc}}$ can be seen as the sum of a Rayleigh-Lamb wave traveling to the right plus a Rayleigh-Lamb wave propagating to the left (respectively the first and second terms). In a similar way, $\mathbf{u}_O^{\text{inc}}$ is written as a Rayleigh-Lamb wave traveling to the right minus the same Rayleigh-Lamb mode traveling to the left.

Figure 2 illustrates the displacement components of $\mathbf{u}_E^{\text{inc}}$ and the corresponding stress components. It is noticed that the even part of the incoming wave has no x component at $x=0$. Similarly, $\tau_{xy}^E=0$ is zero at $x=0$.

Then if $\mathbf{u}^E$ and $\boldsymbol{\tau}^E$ are the displacements and stresses which solve Eqs. (11)–(13) for an even incident wave $\mathbf{u}_E^{\text{inc}}$, then the additional conditions that $\mathbf{u}^E$ and $\boldsymbol{\tau}^E$ must satisfy, by symmetry, are

$$u^E = \tau_{xy}^E = 0 \quad \text{on } -h \leqslant y < a,\ x = 0. \tag{14}$$

Analogously, if $\mathbf{u}^O$ and $\boldsymbol{\tau}^O$ are the displacements and stresses of the solution to the problem when the incident wave is $\mathbf{u}_O^{\text{inc}}$, then $\mathbf{u}^O$ and $\boldsymbol{\tau}^O$ must satisfy

$$v^O = \tau_{xx}^O = 0 \quad \text{on } -h \leqslant y < a,\ x = 0. \tag{15}$$

So the problem for the incident wave $\mathbf{u}^{\text{inc}}$ can be divided in two subproblems, the first problem given by Eqs. (11)–(14) and the second problem given by Eqs. (11)–(13) and (15). Each of these subproblems is now a reflection problem on a semi-infinite plate.

Summarizing, it will be referred to $\mathbf{u}^E$ as the solution of the even subproblem

$$\begin{aligned} &k^2\nabla^2\mathbf{u} + (K^2 - k^2)\,\nabla\,\nabla\cdot\mathbf{u} + k^2K^2\mathbf{u} = 0, \\ &\tau_{yy} = \tau_{xy} = 0, \quad \text{for } x \geqslant 0,\ y = \pm h, \\ &\tau_{xx} = \tau_{xy} = 0, \quad \text{for } x = 0,\ a \leqslant y \leqslant h, \\ &u = \tau_{xy} = 0, \quad \text{for } x = 0,\ -h \leqslant y < a, \end{aligned} \tag{16}$$

and $\mathbf{u}^O$ will be associated with the solution of the odd subproblem

$$\begin{aligned} &k^2\nabla^2\mathbf{u} + (K^2 - k^2)\,\nabla\,\nabla\cdot\mathbf{u} + k^2K^2\mathbf{u} = 0, \\ &\tau_{yy} = \tau_{xy} = 0, \quad \text{for } x \geqslant 0,\ y = \pm h, \\ &\tau_{xx} = \tau_{xy} = 0, \quad \text{for } x = 0,\ a \leqslant y \leqslant h, \\ &v = \tau_{xx} = 0, \quad \text{for } x = 0,\ -h \leqslant y < a. \end{aligned} \tag{17}$$

The following section will show the procedure for solving the even subproblem (16).

III. SOLUTION BY THE PROJECTION METHOD

In this section the projection method is applied to solve the even subproblem. To achieve that, some orthogonality relations are going to be used. Those relations may be obtained after an application of the elastic reciprocal theorem (Sokolnikoff, Ref. 16). The form of the singularities may be incorporated directly into the solution method, by using weighted inner products, in such a way that the singular factors are scaled out whenever possible. The convergence of the method is assessed by computing the energy conservation identity. The results so obtained show that good accuracy can be achieved, thereby offering an inexpensive computational procedure.

A. The projection method

The notation will be simplified in the following manner:

$$\mathbf{u} = \begin{pmatrix} u(x,y) \\ v(x,y) \end{pmatrix}$$

will denote $\mathbf{u}^E$ and

$$\boldsymbol{\tau} = \begin{pmatrix} \tau_{xx}(x,y) \\ \tau_{xy}(x,y) \end{pmatrix}$$

will denote the corresponding stress $\boldsymbol{\tau}^E$.

Following Gregory and Gladwell,[13] an application of the elastic reciprocal theorem on the rectangle $[0,x]\times[-h,h]$, $x>0$, yields that

$$L_n \equiv e^{i\alpha_n x}\int_{-h}^{h}[\mu^{-1}\boldsymbol{\tau}(x,y)\cdot\mathbf{U}_n^P(y) - \mathbf{u}\cdot\mathbf{S}_n^P(y)]dy$$

is independent of x for both $P=S$ and $P=A$, with μ the shear modulus. Making $x\to 0+$ and using the boundary conditions on $\boldsymbol{\tau}$,

$$L_n \to \int_{-h}^{h}\left[\mu^{-1}\begin{pmatrix}\tau_{xx}(0,y)\\ 0\end{pmatrix}\cdot\mathbf{U}_n^P(y) - \begin{pmatrix}u(0,y)\\ v(0,y)\end{pmatrix}\cdot\mathbf{S}_n^P(y)\right]dy.$$

For $x>0$ it is assumed that $\mathbf{u}$ and $\boldsymbol{\tau}$ have a representation given by

$$u(x,y) = -U_0^S e^{-i\alpha_0 x} + \sum_{j=0}^{\infty} a_j U_j^S(x,y)e^{i\alpha_j x} + \sum_{j=0}^{\infty} b_j U_j^A(x,y)e^{i\beta_j x}, \tag{18}$$

$$v(x,y) = V_0^S e^{-i\alpha_0 x} + \sum_{j=0}^{\infty} a_j V_j^S(x,y)e^{i\alpha_j x} + \sum_{j=0}^{\infty} b_j V_j^A(x,y)e^{i\beta_j x}, \tag{19}$$

$$\tau_{xx}(x,y) = R_0^S e^{-i\alpha_0 x} + \sum_{j=0}^{\infty} a_j R_j^S(x,y)e^{i\alpha_j x} + \sum_{j=0}^{\infty} b_j R_j^A(x,y)e^{i\beta_j x}, \tag{20}$$

$$\tau_{xy}(x,y) = -S_0^S e^{-i\alpha_0 x} + \sum_{j=0}^{\infty} a_j S_j^S(x,y)e^{i\alpha_j x} + \sum_{j=0}^{\infty} b_j S_j^A(x,y)e^{i\beta_j x}, \tag{21}$$

where α_j and β_j represent the jth root of the symmetric and antisymmetric modes, respectively. The first and second summations correspond to the reflected symmetric and antisymmetric modes, respectively. This representation is valid if the Rayleigh-Lamb modes are a complete family.

Using the representations given above and the biorthogonality relations for the Rayleigh-Lamb modes, it is obtained, when $x\to\infty$

$$L_n \to \begin{cases} 0 & \text{for } n\geq 0, \quad P=A, \\ 0 & \text{for } n\geq 1, \quad P=S, \\ -J_0 & \text{if } n=0, \text{ and } \quad P=S, \end{cases}$$

where J_0 is given in Eq. (9). This means that

$$\mathbf{F} = \begin{pmatrix} \mu^{-1}\tau_{xx}(0,y) \\ -u(0,y) \\ -v(0,y) \end{pmatrix}$$

is orthogonal to the sets

$$\left\{\begin{pmatrix} U_n^S(y) \\ R_n^S(y) \\ S_n^S(y)\end{pmatrix}^*\right\}_{n\geq 1} \quad \text{and} \quad \left\{\begin{pmatrix} U_n^A(y) \\ R_n^A(y) \\ S_n^A(y)\end{pmatrix}^*\right\}_{n\geq 0}$$

under the inner product defined by

$$\langle\mathbf{P},\mathbf{Q}\rangle = \int_{-h}^{h}\mathbf{P}\cdot\mathbf{Q}^* dy,$$

where $\mathbf{Q}^*$ denotes the complex conjugate of $\mathbf{Q}$. Therefore, the unknown vector function $\mathbf{F}$ must belong to the orthogonal complement of the space

$$\mathcal{R} \equiv \bar{\mathcal{L}}\left\{\begin{pmatrix} U_n^S(y) \\ R_n^S(y) \\ S_n^S(y)\end{pmatrix}^*, \begin{pmatrix} U_n^A(y) \\ R_n^A(y) \\ S_n^A(y)\end{pmatrix}^*\right\}_{n\geq 0}.$$

Assuming completeness of the set

$$\left\{\begin{pmatrix} U_n^S(y) \\ R_n^S(y) \\ S_n^S(y)\end{pmatrix}^*, \begin{pmatrix} U_n^A(y) \\ R_n^A(y) \\ S_n^A(y)\end{pmatrix}^*\right\}_{n\geq 0},$$

can be deduced that $\mathcal{R}^{\perp}$ is one dimensional. Then, the procedure to find $\mathbf{F}$ will be to start with an arbitrary vector $\mathbf{h}\notin\mathcal{R}$ and compute the orthogonal projection of $\mathbf{h}$ on $\mathcal{R}$, $\mathbf{h}_p$. So, $\mathbf{h}-\mathbf{h}_p$ must be in $\mathcal{R}^{\perp}$. Therefore for some scalar γ,

$$\mathbf{h} - \mathbf{h}_p = \gamma\mathbf{F}.$$

The value of γ can be found from the normalization relation

$$L_0 = -J_0.$$

B. Singularities of the solution

The behavior of the fields close to the tip of the crack can be found by considering the static case, as these equations retain the most singular terms. Karp and Karal[17] and Williams[18] have studied this behavior. The equations for elastostatics, in terms of the stress under plane strain are

$$\frac{\partial\tau_{xx}}{\partial x} + \frac{\partial\tau_{xy}}{\partial y} = 0, \tag{22}$$

$$\frac{\partial\tau_{xx}}{\partial x} + \frac{\partial\tau_{xy}}{\partial y} = 0. \tag{23}$$

Introducing an Airy stress potential, the components of stress are computed according to

$$\tau_{xx} = \frac{\partial^2 \Phi}{\partial y^2}, \quad \tau_{xy} = -\frac{\partial^2 \Phi}{\partial x\, \partial y}, \quad \tau_{yy} = \frac{\partial^2 \Phi}{\partial x^2}.$$

To ensure that the strain components derived from this stress field are compatible, the potential Φ must satisfy the biharmonic equation

$$\nabla^2 \nabla^2 \Phi = 0. \tag{24}$$

Following Williams,[18] it will be assumed an Airy stress potential in the form

$$\Phi = r^{\lambda+1} f(\theta),$$

which must satisfy Eq. (24), subject to the boundary conditions for the stress on the crack, and where r and θ represent polar coordinates. It will be convenient to introduce the following notation:

$$\tau_{xx+} \equiv \tau_{xx}(0,y), \quad \tau_{xy+} \equiv \tau_{xy}(0,y), \quad \text{for } y > a,$$

$$\tau_{xx-} \equiv \tau_{xx}(0,y), \quad \tau_{xy-} \equiv \tau_{xy}(0,y), \quad \text{for } y < a,$$

and

$$u_+(y) \equiv u(0,y), \quad v_+(y) \equiv v(0,y), \quad \text{for } y > a,$$

$$u_-(y) \equiv u(0,y), \quad v_-(y) \equiv v(0,y), \quad \text{for } y < a.$$

After solving Eq. (24) with the boundary conditions at the crack, it is found that

$$\tau_{xx-} = Ar^{-1/2} + \cdots \quad \text{for } y \leqslant a,$$

where A is a constant. Because of the continuity of the displacements, the integration constant found for u_+, after integrating the corresponding stresses, must be zero, so

$$u_+ = \frac{4A(1+\nu)(1-\nu)}{E} r^{1/2} + \cdots,$$

where ν is Poisson's ratio and E is Young's modulus. However this argument cannot be used for the displacements v_+ and v_-, then an unknown integration constant, C, must be taken into account so that

$$v_- = C + \frac{2A(1+\nu)(-1+2\nu)}{E} r^{1/2} + \cdots,$$

$$v_+ = C + \frac{4A(1+\nu)(1-\nu)}{E} r + \cdots.$$

In terms of the variables τ_{xx-}, u_+, v_-, and u_+ the orthogonality relations mean that

$$\begin{pmatrix} \mu^{-1}\tau_{xx-} \\ -u_+ \\ -v_- \\ -v_+ \end{pmatrix}$$

is orthogonal to

$$\boldsymbol{G}_j = \begin{pmatrix} U_j^S \\ R_j^S \\ S_j^S \\ S_j^S \end{pmatrix}^* \tag{25}$$

for $j \geqslant 1$ and to

$$\boldsymbol{H}_j = \begin{pmatrix} U_j^A \\ R_j^A \\ S_j^A \\ S_j^A \end{pmatrix}^*, \tag{26}$$

for $j \geqslant 0$ under the usual complex inner product. Define the scaled variables, $\hat{\tau}_{xx-}$ and $\hat{u}_+$ such that

$$\tau_{xx-} = \hat{\tau}_{xx-} r^{-1/2},$$

$$u_+ = \hat{u}_+ r^{1/2}.$$

In terms of the scaled variables the orthogonality relation is written as

$$\mathbf{F} = \begin{pmatrix} \mu^{-1}\hat{\tau}_{xx-} \\ -\hat{u}_+ \\ -\hat{v}_- \\ -\hat{v}_+ \end{pmatrix}, \tag{27}$$

is orthogonal to $\boldsymbol{G}_k$ and to $\boldsymbol{H}_j$ for $k \geqslant 1$ and $j \geqslant 0$ under the weighted inner product

$$\langle \mathbf{P}, \mathbf{Q} \rangle = \int_{-h}^{a} p_1 q_1^* (a-y)^{-1/2} dy + \int_a^h p_2 q_2^* (y-a)^{-1/2} dy + \int_{-h}^{a} p_3 q_3^* dy + \int_a^h p_4 q_4^* dy,$$

where it is assumed that $\mathbf{P} = (p_1(y), p_2(y), p_3(y), p_4(y))$ and $\mathbf{Q} = (q_1(y), q_2(y), q_3(y), q_4(y))$. In terms of these scaled variables the unknown $\mathbf{F}$ can be approximated by $\mathbf{F}_n$ if a finite dimensional space $\mathcal{R}$ is taken. In that case the orthogonality relation can be established as a finite system of equations. If $\mathbf{h}$ is any vector not in $\mathcal{R}$, then

$$\gamma \mathbf{F}_n = \mathbf{h} - \sum_{j=1}^{n} d_j^S \boldsymbol{G}_j - \sum_{j=0}^{n} d_j^A \boldsymbol{H}_j,$$

the coefficients d_j^S and d_j^A are obtained from the linear equations

$$\sum_{j=1}^{n} d_j^S \langle \boldsymbol{G}_j, \boldsymbol{G}_m \rangle + \sum_{j=0}^{n} d_j^A \langle \boldsymbol{H}_j, \boldsymbol{G}_m \rangle = \langle \mathbf{h}, \boldsymbol{G}_m \rangle, \tag{28}$$

for $m = 1, \ldots, n$, and

$$\sum_{j=1}^{n} d_j^S \langle \boldsymbol{G}_j, \boldsymbol{H}_m \rangle + \sum_{j=0}^{n} d_j^A \langle \boldsymbol{H}_j, \boldsymbol{H}_m \rangle = \langle \mathbf{h}, \boldsymbol{H}_m \rangle, \tag{29}$$

for $m = 0, \ldots, n$. The normalization constant, γ, may be found from the relation

$$\gamma\langle \mathbf{F}_n, \boldsymbol{G}_0\rangle = -J_0, \tag{30}$$

where J_0 is given by Eq. (9). Besides Eqs. (28)–(30), additional conditions satisfied by the components of $\mathbf{F}_n$ at the crack tip can be used. From the definition of the variables it is obtained that

$$\hat{\tau}_{xx-}(a-) = A,$$

$$\hat{u}_+(a+) = 4A(1-\nu)(1+\nu)/E, \tag{31}$$

$$v_-(a-) = v_+(a+).$$

The last condition is only the statement of continuity of the vertical displacements at the crack tip. Using the relation $\mu = E/2(1+\nu)$ for the elastic constants of the material, $\hat{\tau}_{xx-}(a-)$ can be related to $\hat{u}_+(a+)$ as

$$\hat{u}_+(a+) = 2(1-\nu)\mu^{-1}\hat{\tau}_{xx-}(a-). \tag{32}$$

More information about the solution can be obtained. For instance, at the upper corner, the point $(0,h)$, $\tau_{xx} \to 0$ and $\tau_{yy} \to 0$, so $\partial v/\partial y \to 0$. At the lower corner, $(0,-h)$, it can be found that v and τ_{xx} are related by

$$2\mu\frac{\partial v}{\partial y} = -\nu\tau_{xx}.$$

Writing these two relations in terms of the scaled variables yields

$$\frac{\partial \hat{v}_+}{\partial y}(h) = 0, \tag{33}$$

and

$$-\frac{2}{\nu}\frac{\partial \hat{v}_-}{\partial y}(-h) = \mu^{-1}(a+h)^{-1/2}\hat{\tau}_{xx-}(-h), \tag{34}$$

respectively. So, four additional equations, (31)–(34), have been produced for which it is needed four more variables to include in the system of equations so as not to be overprescribed. These may be obtained by writing $\mathbf{h}$ in the form

$$\mathbf{h} = \mathbf{h}_0 + c_1\mathbf{h}_1 + c_2\mathbf{h}_2 + c_3\mathbf{h}_3 + c_4\mathbf{h}_4,$$

with the vectors $\mathbf{h}_0, \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3$, and $\mathbf{h}_4$ chosen to be linearly independent. For instance, these vectors have been taken as

$$\mathbf{h}_0 = (1,1,1,2)^T, \quad \mathbf{h}_1 = (0,1,0,2)^T,$$

$$\mathbf{h}_2 = (0,1,1,0)^T, \quad \mathbf{h}_3 = (0,0,y,0)^T,$$

$$\mathbf{h}_4 = (0,0,0,y)^T.$$

The inclusion of the singularities will produce an improvement of the convergence of the approximation based only on the normal modes. Any expansion of the solution to this problem, in terms of normal modes is an expansion for a singular solution in terms of a finite number of continuous functions. So, to mimic such singular behavior it could be needed to use a large number of normal modes.

IV. OBTAINING THE AMPLITUDES

The amplitudes are obtained by taking advantage of the orthogonality of the modes. The amplitudes a_j and b_j in the expansions given by Eqs. (18)–(21) for the even problem, may be obtained by applying reciprocity to the fields $\boldsymbol{\tau}$ and $\mathbf{u}$ and $\mathbf{U}^P_{-n}\exp(-i\alpha_n x)$, and $\mathbf{S}^P_{-n}\exp(-i\alpha_n x)$. After doing this, it is obtained

$$a_m = \frac{\langle \mathbf{F}, \boldsymbol{G}_m\rangle}{J_m},$$

$$b_m = \frac{\langle \mathbf{F}, \boldsymbol{H}_m\rangle}{K_m},$$

where $\mathbf{F}, \boldsymbol{G}_m$, and $\boldsymbol{H}_m$ are given by Eqs. (25)–(27). The quantities J_m and K_m are as defined in Eqs. (9) and (10), respectively.

Once the amplitudes of the even and odd subproblems are evaluated it is possible to compute the amplitudes of the complete problem using the relation

$$\mathbf{u}(x,y) = \begin{cases} \mathbf{u}^E(-x,y) - \mathbf{u}^O(-x,y) & \text{for } x < 0, \\ \mathbf{u}^E(x,y) + \mathbf{u}^O(x,y) & \text{for } x > 0, \end{cases} \tag{35}$$

where $\mathbf{u}^E(x,y)$ are the displacements of the even problem and $\mathbf{u}^O(x,y)$ the displacements of the corresponding odd problem.

V. RECIPROCAL RELATIONS

The reciprocal theorem has been utilized so far to obtain orthogonality relations. However it can provide another interesting relation between the incident and the scattered field for different modes.

Consider a region $\mathcal{B}$ in an elastic plate containing a flaw, which is of arbitrary shape. It is assumed that all the boundaries of the body are stress-free. In addition, it is assumed that the S_m Rayleigh-Lamb mode is incident on the scattering flaw S, from $-\infty$. In the particular case of a scatterer S, symmetric with respect to the y axis the fraction of energy transmitted by the m mode for an incoming k mode is equal to the fraction of energy transmitted in the k mode by an incoming m mode. The relation is also true for any combination of symmetric and antisymmetric modes (Auld, Ref. 19). In a similar manner, the corresponding reciprocal principle is satisfied for the reflected energy and for any combination of symmetric and antisymmetric modes. These identities can be obtained from the application of the reciprocal theorem and they may be useful to check numerical computations.

VI. NUMERICAL RESULTS

In Figs. 3 and 4 are plotted the transmission and reflection coefficients of the fundamental modes, when S_0 or A_0 are the incident fields, respectively, as functions of the ratio of the crack depth with respect to the plate thickness $2h$. The data used to produce such results were the following: $c_L = 5.94$ mm/μs, $c_T = 3.2$ mm/μs, $\rho = 7.8$ g/cm^3, and the frequency was such that $2fh = 1$ MHz mm, where f is the linear frequency and $2h$ is the plate thickness. These parameters give for the Poisson's ratio the value $\nu = 0.286\,542$. They

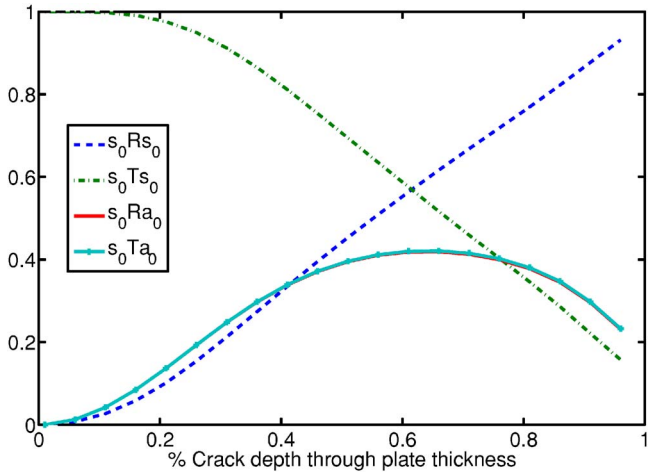


FIG. 3. (Color online) Transmission and reflection coefficients for an S_0 incident mode, obtained by projection. The curves corresponding to ${}_{S_0}\mathrm{R}_{A_0}$ and ${}_{S_0}\mathrm{T}_{A_0}$ are overlapping.

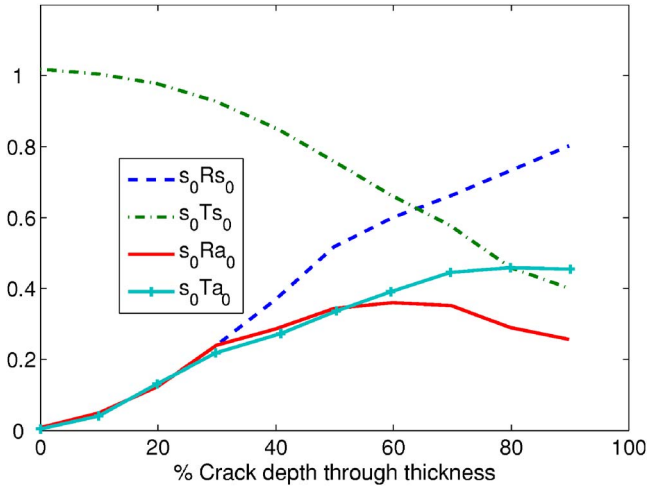


FIG. 5. (Color online) Transmission and reflection coefficients for an S_0 incident mode, obtained by Cho and Rose by boundary element method, for a semielliptical crack.

were chosen in this way in order to compare with the results obtained by Cho and Rose.[8] In their work, Cho and Rose used Boundary Element Methods to solve the scattering problem. Instead of considering flat cracks they modeled the cracks by semiellipses. However, the data produced by them for the flattest semiellipse, Figs. 5 and 6, which corresponds to an aspect ratio of the horizontal semiaxis to the vertical semiaxis equal to 0.2, are qualitatively similar to the results obtained by projection for a flat crack at the same frequency.

In both of these graphs, Figs. 3 and 4, the reciprocal principle discussed in the previous section appears. If ${}_{A_0}\mathrm{T}_{S_0}$ is the transmitted energy corresponding to the mode S_0, when the mode A_0 is incident and ${}_{S_0}\mathrm{T}_{A_0}$ is the transmitted energy of the A_0 mode when the incident mode is S_0, then

$$ {}_{A_0}\mathrm{T}_{S_0} = {}_{S_0}\mathrm{T}_{A_0} $$

similarly, if ${}_{A_0}\mathrm{T}_{S_0}$ and ${}_{S_0}\mathrm{T}_{A_0}$ are the corresponding fractions of reflected energy, then

$$ {}_{A_0}\mathrm{R}_{S_0} = {}_{S_0}\mathrm{R}_{A_0}. $$

Another important aspect to remark on is that Figs. 3 and 4 also show an overlapping between the reflection and transmission coefficients of the A_0 mode when the incident mode is S_0. The same observation applies to the reflection and transmission coefficients of the S_0 mode due to an incoming mode A_0. This could be explained by the behavior of the S_0 and A_0 modes in the low frequency range. For very low frequency, the S_0 mode is fundamentally a compressional mode and the fundamental antisymmetric mode, A_0, is mainly a flexural mode (see Achenbach, Ref. 20); so when the thickness of the plate is very small compared to the wavelength, $2\pi/k$, the boundary conditions for the odd subproblem (17), $\tau_{xy}=0$, $v=0$, are automatically satisfied if the incoming wave is S_0. The remaining boundary condition, $\tau_{xx}=0$, can be satisfied if the reflected mode is the same compressional mode S_0 with negative sign. It means that the odd subproblem (17) will give no amplitudes for the antisymmetric modes. From Eq. (35) for the amplitudes in terms

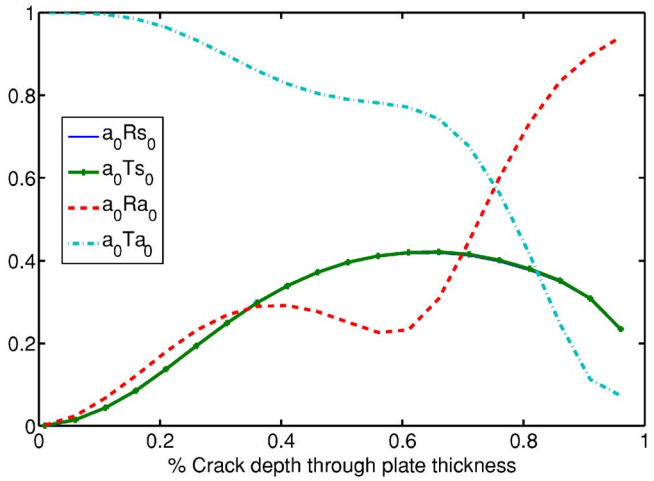


FIG. 4. (Color online) Transmission and reflection coefficients for an A_0 incident mode, obtained by projection. The curves corresponding to ${}_{A_0}\mathrm{R}_{S_0}$ and ${}_{A_0}\mathrm{T}_{S_0}$ are overlapping. Also, the curves corresponding to these variables are the same as the curves corresponding to the values of ${}_{S_0}\mathrm{R}_{A_0}$ and ${}_{S_0}\mathrm{T}_{A_0}$, shown in the previous figure.

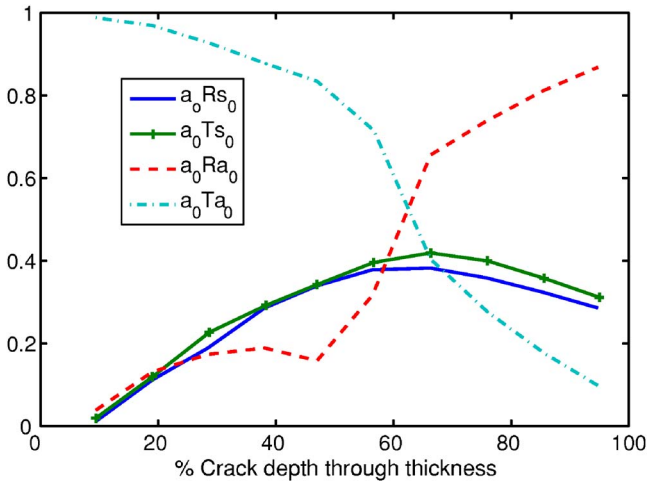


FIG. 6. (Color online) Transmission and reflection coefficients for an A_0 incident mode, obtained by Cho and Rose by the boundary element method, for a semielliptical crack.

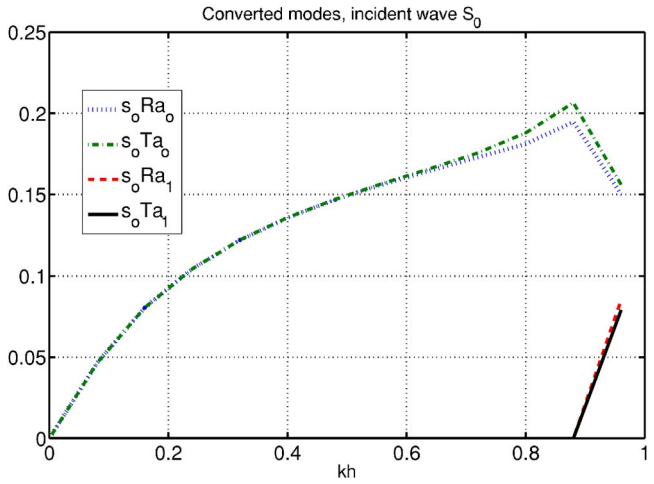


FIG. 7. (Color online) Behavior of the converted energy for an S_0 incident mode, when the dimensionless wave number kh is small. As $kh \to 0$, ${}_{S_0}R_{A_0}$ and ${}_{S_0}T_{A_0}$ are very close to each other.

of the even and odd subproblems, it is obtained that the reflected and transmitted energy corresponding to the antisymmetric mode, A_0, must be the same for the complete problem. A similar argument applies for the case when the A_0 mode is incident. In Fig. 7 it is possible to notice that the reflected and transmitted energy of the converted mode A_0 is the same, when the dimensionless wave number, kh, is small and the incident mode is S_0. The number of terms used for these computations in the eigenfunction expansion was of 20 symmetric modes and 20 antisymmetric modes. The accuracy obtained in energy conservation (total reflected energy plus total transmitted must be equal to the incident energy), was of three decimal places.

Finally, Fig. 8 shows the variation of the reflection coefficient as a function of the crack depth and frequency. These data were obtained by the projection method, using ten symmetric and ten antisymmetric modes in the eigenfunction expansion. There is good agreement with the results published by Lowe and Diligent[9] using a finite element method. The elastic constants are the same as those employed in the previous figures.

The previous graph shows the advantage of including the singular behavior of the solution in the approximation.

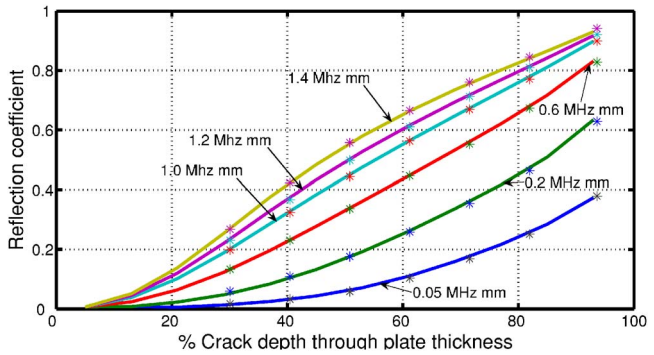


FIG. 8. (Color online) Reflection coefficient for an S_0 incident mode as a function of the crack depth for several excitation frequencies. (–) Results obtained by the projection method; (*) results from Lowe and Diligent (Ref. 9), computed by the finite element method.

The number of required modes in a normal mode expansion may be large in order to get a modest accuracy. Castaings *et al.*,[5] mention that they have to solve linear systems of size at least 40×40, in order to get an error within 5% in the energy balance equation.

VII. CONCLUSION

The method of projection has been used to solve the problem of scattering of Rayleigh-Lamb waves by a surface breaking crack, and different principles have been used to assess the numerical results obtained. The first is the principle of energy conservation. The second is the reciprocal principle, which the results show clearly. From the numerical results it is also observed that as kh, the product of the wave number times half the thickness of the plate, approaches zero, the reflected and transmitted energy of the fundamental converted mode, under a fundamental incident mode, are both equal. This observation was then validated, using the asymptotic behavior of the fundamental modes, for small wave numbers.

Besides the accuracy offered by this semianalytical method, another point to remark on is that its convergence is very fast, as the singular behavior about the crack tip is incorporated directly. Then with a small number of modes in the expansion it is possible to achieve accurate results, with a modest time for computation. These characteristics make the projection method a very suitable tool for solving inverse problems related to the detection and characterization of flaws in materials by ultrasonic techniques.

ACKNOWLEDGMENTS

The first author is very grateful to Professor Federico Sabina and the Departamento of Matemáticas y Mecánica, at the IIMAS-UNAM, for their hospitality during a short stay as a visiting researcher.

[1]D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–283 (1997).
[2]S. Rokhlin, "Diffraction of Lamb waves by a finite crack in an elastic layer," J. Acoust. Soc. Am. **67**, 1157–1165 (1980).
[3]Y. N. Al-Nassar, S. K. Datta, and A. H. Shah, "Scattering of Lamb waves by a normal rectangular strip weldment," Ultrasonics **29**, 125–132 (1991).
[4]D. N. Alleyne and P. Cawley, "The interaction of Lamb waves with defects," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 381–397 (1992).
[5]M. Castaings, E. Le Clezio, and B. Hosten, "Modal decomposition method for modeling the interaction of Lamb waves with cracks," J. Acoust. Soc. Am. **112**, 2567–2582 (2002).
[6]A. Gunawan and S. Hirose, "Mode-exciting method for Lamb wave-scattering analysis," J. Acoust. Soc. Am. **115**, 996-1005 (2004).
[7]Y. Cho, D. D. Hongerholt, and J. L. Rose, "Lamb wave scattering analysis for reflector characterization," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 44–52 (1997).
[8]Y. Cho and J. Rose, "An elastodynamic hybrid boundary element study for elastic guided wave interactions with a surface breaking defect," Int. J. Solids Struct. **37**, 4103–4124 (2000).
[9]M. S. Lowe and O. Diligent, "Low-frequency reflection characteristics of the s_0 Lamb wave from a rectangular notch in a plate," J. Acoust. Soc. Am. **111**, 64–74 (2002).
[10]S. K. Datta, A. H. Shah, R. L. Bratton, and T. Chakraborty, "Wave propagation in laminated composite plates," J. Acoust. Soc. Am. **83**, 2020–2026 (1988).
[11]W. M. Karunasena, K. M. Liew, and S. Kitipornchai, "Hybrid analysis of Lamb wave reflection by a crack at the fixed end of a composite," Com-

put. Methods Appl. Mech. Eng. **125**, 221–233 (1995).
[12]R. D. Gregory and I. Gladwell, "The cantilever beam under tension, bending or flexure at infinity," J. Elast. **12**, 317–343 (1982).
[13]R. D. Gregory and I. Gladwell, "The reflection of a symmetric Rayleigh-Lamb wave at the fixed or free edge of a plate," J. Elast. **13**, 185–206 (1983).
[14]R. D. Gregory, Ch. Gu, and F. Y. M. Wan, "The cantilever strip plate under torsion, bending or flexure at infinity," J. Elast. **43**, 109–136 (1996).
[15]H. Lamb, "On waves in an elastic plate," Proc. R. Soc. London, Ser. A **93**, 114–128 (1917).
[16]I. S. Sokolnikoff, *Mathematical Theory of Elasticity* (McGraw-Hill, New York, 1956).
[17]S. N. Karp and F. C. Karal, "The elastic field behavior in the neighborhood of a crack of arbitrary angle," Commun. Pure Appl. Math. **15**, 413–421 (1962).
[18]M. L. Williams, "Stress singularities resulting from various boundary conditions in angular corners of plates in extension," J. Appl. Mech. **19**, 526–528 (1952).
[19]B. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, 1990), Vol. 2.
[20]J. Achenbach, *Wave Propagation in Elastic Solids* (North-Holland, Amsterdam, 1973).

Mode-matching without root-finding: Application to a dissipative silencer[a)]

Jane B. Lawrie[b)]
School of Information Systems, Computing and Mathematics, Mathematical Sciences, Brunel University, Uxbridge UB8 3PH, United Kingdom

Ray Kirby[c)]
School of Engineering and Design, Mechanical Engineering, Brunel University, Uxbridge UB8 3PH, United Kingdom

(Received 17 June 2005; revised 2 December 2005; accepted 12 January 2006)

This article presents an analytic mode-matching approach suitable for modelling the propagation of sound in a two-dimensional, three-part, ducting system. The approach avoids the need to find roots of the characteristic equation for the middle section of the duct (the component) and is readily applicable to a broad class of problems. It is demonstrated that the system of equations, derived via analytic mode-matching, exhibits certain features which ensure that they can be recast into a form that is independent of the roots of the characteristic equation for the component. The precise details of the component are irrelevant to the procedure; it is required only that there exists an orthogonality relation, or similar, for the eigenmodes corresponding to the propagating wave forms in this region. The method is applied here to a simple problem involving acoustic transmission through a dissipative silencer of the type commonly found in heating ventilation and air-conditioning ducts. With reference to this example, the silencer transmission loss is computed, and the power balance for the silencer is investigated and is shown to be an identity that is necessarily satisfied by the system of equations, regardless of the level of truncation.
 [DOI: 10.1121/1.2172168]

PACS number(s): 43.20.Mv, 43.50.Gf [DKW] Pages: 2050–2061

I. INTRODUCTION

Guided waves can be observed in a wide range of physical situations. Examples include water waves propagating along a channel, seismological waves travelling through a layer of rock, the transmission of electromagnetic waves along an optical fiber and acoustical waves manifesting as noise in heating ventilation and air conditioning (HVAC) ducts. Very often the physical problems of interest can be formulated as boundary value problems that are amenable to analytic solution methods such as the Wiener-Hopf technique or mode-matching. A precursor to these methods, however, is accurate knowledge of sufficient admissible wave numbers. These are usually defined in terms of a characteristic equation, the complexity of which depends on both the equations governing the wave guide media and on the boundary conditions. Only in the simplest of problems can the wave numbers be expressed in closed form and for problems involving elastic or porous media and/or flexible walled ducts numerical "root-finding" is a normal procedure. Root-finding, while tedious, is by no means an impossible task but the difficulties should not be underestimated. Of course, the Wiener-Hopf and mode-matching techniques are not necessarily the only viable analytic solution methods. A wide range of other techniques are available including those based on Green's theorem and a variety of Fourier integral approaches. Some of these approaches successfully avoid the process of root-finding, see, for example, Huang;[1,2] however, such methods are not always easily generalized.

In the following discussion attention is restricted to two-dimensional waveguides. Associated with every mode is an axial and at least one transverse wave number which are related through the governing equation(s). Further, it is assumed that the waveguide boundary conditions contain only even derivatives in the axial direction.[3] Then, broadly speaking, the characteristic equation, which will be meromorphic and expressed in terms of the transverse wave number, can be broken down into four categories. First, there is the simple trigonometric form, such as that for a rigid walled acoustic duct, for which the roots can be written down in closed form. Second, there are those that involve a linear combination of trigonometric terms and possibly algebraic terms but which have only real parameters and involve only even powers of at most one transverse wave number. For such characteristic functions the roots always lie on either the real axis or the imaginary axis. Furthermore, the asymptotic form of the roots is usually known. The characteristic equations for flexible-walled acoustic ducts[3] are typical of this class. The third class of characteristic function is much the same as the second other than it will involve even powers of two related transverse wave numbers. Waveguides comprising slabs of an elastic material typically have this

[a)]Portions of this work were presented in "On analysing a dissipative silencer: a mode-matching approach" Proceedings of IUTAM 2002/04, Liverpool, UK, 2002.
[b)]Author to whom correspondence should be addressed. Electronic mail: jane.lawrie@brunel.ac.uk
[c)]Electronic mail: ray.kirby@brunel.ac.uk

type of characteristic function. In this case the roots may be real, imaginary or complex yet much may be known about their location[4] and asymptotic form. The fourth class of characteristic equation has all the properties of those in class three but, in addition, the physical parameters are complex. This class of characteristic equation arises in the study of dissipative silencers in which the absorbent material is modelled as an equivalent fluid.

The latter class of characteristic equation is known to be problematic when it comes to root-finding. This is because the roots do not necessarily lie on or close to any specified curve in the complex plane which makes it difficult to choose appropriate initial guesses for the root-finding algorithm. Further, the complex parameters of the dispersion relation are usually frequency dependent and thus the location of the roots can vary, often significantly and in arbitrary direction, with small variations in frequency. If one desires to find the roots at one specific frequency then usually, with perseverance, this can be done. Then, if one wishes to plot a given physical property against frequency it is usual to adopt a tracking approach. That is, the roots successfully located for frequency f_0 are used as initial guesses for frequency $f_0+\epsilon$, $\epsilon \ll 1$. Although this approach increases the likelihood of finding all required roots as frequency varies, it can fail even for frequency shifts of $\epsilon < 0.01$ Hz. Not only is it difficult to locate all the roots in a specified region of the complex plane at all desired frequencies, but there are very few reliable techniques by which to determine whether all the roots have been found. The Argument Principle is the only rigorous test, but if one or more roots are found to be missing it is of limited help in locating them. Traditional algorithms used for solving this class of characteristic equation include the Newton-Raphson method and variations such as the secant[5] and Muller's[6] methods. None of these are robust in the sense that modes are easily missed if the initial guesses are not sufficiently close to the actual root. Although the effect of a missing root on the results depends on the location of the root, it is always undesirable and often the cause of significant inaccuracies.[7]

Analytic solution methods that avoid root-finding are highly desirable but few and far between. Huang,[1,2] for example, considers a class of problem whereby a membrane is inserted into a finite section, say $-\ell \leqslant x \leqslant \ell$, of an otherwise infinite rigid duct. He employs a technique whereby the sound field within the duct is represented by an infinite sum of Fourier integrals, each one forced by a velocity distribution $\sin[n\pi(x+\ell)/(2\ell)]$ on the duct surface for $-\ell \leqslant x \leqslant \ell$. The Fourier coefficients of the pressure field are then determined by substituting an evaluated form of the Fourier integrals into the membrane condition. This works well for the class of problem that Huang considers but the method is limited to situations in which the membrane forms part of the duct surface. The method is not applicable to situations whereby there is a structure, such as a finite length membrane, positioned parallel to the duct wall within the fluid region or to situations in which part of the fluid region is layered comprising, for example, alternate layers of fluid and porous material. Further, the extension of Huang's approach to the situation in which the membrane is replaced by an elastic plate is straightforward *only* for the case in which the plate edges are pin-jointed. There is a clear need for analytic techniques that can tackle a wide range of such geometries and do not depend on the roots of the characteristic equation.

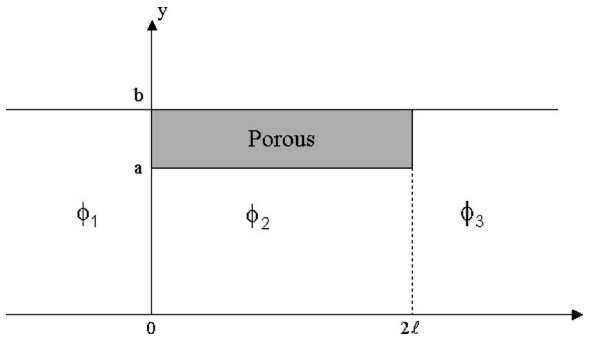


FIG. 1. Silencer geometry.

This article offers an approach by which root-finding can be avoided for a broad class of problems: the propagation of sound in a two-dimensional, three-part ducting system see, for example, Fig. 1. The class of problem is described in Sec. II and a general mode-matching solution is presented. This solution exhibits certain features that ensure it can always be recast, using a contour integral technique, into a form that is independent of the roots of the characteristic equation for the middle region. This approach completely bypasses the need to solve the characteristic equation, and provides robust and accurate expressions by which to compute any physical quantity of interest. In Sec. III attention is focussed on a particular example: the propagation of sound through a simple silencer comprising a finite length of lined duct. The contour integral technique is demonstrated and a detailed analysis of the power balance is presented. Further analysis proves that the power balance is an identity which, regardless of the severity of the truncation, is automatically satisfied by the system of equations relating the reflection and transmission coefficients. The latter result is well established for nondissipative systems[8,9] but, to the authors' knowledge, has not previously been proven for a dissipative system. Numerical results, in terms of transmission loss and power absorbed across each surface of the liner, are presented for two different silencer configurations in Sec. IV. In Sec. V a discussion of the potential extensions and limitations of this "root-free" approach is presented.

II. THE GENERAL PROBLEM

In this section the general problem of determining the sound field in a two-dimensional, three-part ducting system is discussed. It is demonstrated that, for the class of problem to be considered, the system of equations derived via mode-matching have a particular structure that ensures that they can always be recast into a form that is independent of the roots of the characteristic equation. The system comprises inlet and outlet ducts lying in the regions $0 \leqslant \bar{y} \leqslant \bar{b}$, $\bar{x} < 0$ and $0 \leqslant \bar{y} \leqslant \bar{b}$, $\bar{x} > 2\bar{\ell}$ of a Cartesian frame of reference. A "silencer-type" component is sandwiched between them occupying the region $0 \leqslant \bar{y} \leqslant \bar{h}$, $0 \leqslant \bar{x} \leqslant 2\bar{\ell}$ where $\bar{h}$ may be

greater than, less than or equal to $\bar{b}$. The vertical line segments joining the points $y=h$ to $y=b$ at $x=0$ and $x=2\ell$ (when $h\neq b$) are assumed to be rigid. Thus, the duct geometry is symmetrical about the vertical line $x=\ell$. Figure 1 shows a specific example of a typical three-part ducting system in which $h=b$. The analysis that follows is, however, in no way restricted to this case. A compressible fluid of sound speed c_f and density ρ fills the interior of the duct and region exterior to the duct is in vacuo. The incident sound field is assumed to have harmonic time dependence, $e^{+i\omega\bar{t}}$, where ω is related to the frequency by $\omega=2\pi f$, and propagates through the fluid in the positive $\bar{x}$ direction towards $\bar{x}=0$.

It is convenient to nondimensionalize the boundary value problem using typical length and time scales k^{-1} and ω^{-1} where $\omega=c_f k$. Thus, $x=k\bar{x}$, $y=k\bar{y}$, etc., where the "barred" quantities are dimensional. The time-independent fluid velocity potentials for each duct region are $\phi_1(x,y)$, $x<0$; $\phi_2(x,y)$, $0<x<2\ell$ and $\phi_3(x,y)$, $x>2\ell$ (see Fig. 1). For the inlet and outlet ducts ($x<0$ and $x>2\ell$, respectively) the velocity potentials are governed by Helmholtz's equation and must satisfy specified boundary conditions at $y=0$ and $y=b$. Thus,

$$\phi_1 = \sum_{j=0}^{\infty} F_j Z_j(y) e^{-\eta_j x} + \sum_{j=0}^{\infty} A_j Z_j(y) e^{\eta_j x}, \tag{1}$$

$$\phi_3 = \sum_{j=0}^{\infty} D_j Z_j(y) e^{-\eta_j (x-2\ell)}, \tag{2}$$

where A_j, D_j, $j=1,2,3,\ldots$ are the complex amplitudes of the reflected and transmitted modes and η_j is the axial wave number of the duct modes. The incident sound field in (1) accommodates multimodal forcing for which the modal amplitudes F_n will inevitably depend on the characteristics of the noise source (typically a fan). The eigenmodes $Z_j(y)$, $j=0,1,2,\ldots$ depend on the inlet and/or outlet duct walls which, for example, may be rigid, soft or wave bearing. In the latter case the eigenmodes $Z_j(y)$, $j=0,1,2,\ldots$ will satisfy a non-Sturm-Liouville eigensystem and the orthogonality relation (OR) will be of the form (4) together with (5) discussed below. This point is revisited in Sec. V.

It is assumed that the component lying in the region $0\leqslant x\leqslant 2\ell$ is of very much more complicated structure than the inlet and/or outlet ducts, possibly comprising wave-bearing walls and layers of absorbent material. The precise details of the component are irrelevant to the mode-matching procedure, it is required only that there exists an orthogonality relation, or similar, for the eigenmodes corresponding to the propagating wave forms in the component region. Under such circumstances the velocity potential in the component region is given by

$$\phi_2 = \sum_{j=0}^{\infty} (B_j e^{-s_j x} + C_j e^{s_j x}) Y_j(y), \tag{3}$$

where B_j, C_j are the complex amplitudes of the waves in the component region. The quantities s_j and $Y_j(y)$, $j=1,2,3,\ldots$ are the wave number of the jth mode and the corresponding eigenfunction. Although details of the eigensystem are not required for the process described here, it should be noted that the wave numbers, s_j, $j=0,1,2,\ldots$ are defined as the roots of the characteristic equation $K(s)=0$ and that the eigenfunctions must satisfy an OR, or similar, of the form

$$(Y_n, Y_m) = \delta_{mn} E_n, \tag{4}$$

where δ_{mn} is the usual Kronecker delta function. It is worthwhile pointing out that such relationships exist not only for Sturm-Liouville eigensystems but also for systems in which the boundary conditions contain high order, even derivatives with respect to the axial direction of the waveguide, that is, x. [A consequence of the latter point is that $K(-s)=K(s)$.] Such boundary conditions arise if wave-bearing surfaces, such as a membrane or elastic plate, form part of the component. In such cases the OR will comprise the usual integral term together with a linear combination of the eigenfunctions and their derivatives. A typical form for the left-hand side of the OR is thus

$$(Y_n, Y_m) = \int_0^h Y_n(y) Y_m(y) dy + P(\gamma_n^2, \gamma_m^2) Y_n'(y_0) Y_m'(y_0), \tag{5}$$

where $\gamma_n=(-1-s_n^2)^{1/2}$ and $P(u,v)=P(v,u)$ is a polynomial in the variables u and v, the coefficients of which depend on the boundary condition describing the wave-bearing surface which is taken to lie along the line segment $y=y_0$, $0\leqslant x\leqslant 2\ell$. A general class of such expressions was derived by Lawrie and Abrahams[3] and these have subsequently been utilized in a variety of situations.[9–11] Expression (4) is written as an inner product. This formulation is accurate for Sturm-Liouville systems and also those containing membrane boundaries. For systems with higher-order derivatives present in the boundary conditions the left-hand side of (4) will not be an inner product in the usual sense. This, however, is immaterial to the following analysis.

Crucial to the process of expressing the mode-matching equations in root-free form is the knowledge that E_n is related to the characteristic equation through an expression of the form

$$E_n = \left. \frac{Y_n'(y_0)}{2s_n} \frac{d}{ds} K(s) \right|_{s=s_n}, \tag{6}$$

where the prime indicates differentiation with respect to y and y_0 is constant, $0\leqslant y_0\leqslant h$. The physical significance of the line segment $y=y_0$, $0\leqslant x\leqslant 2\ell$ depends on the details of the component but very often represents an interface between fluid and, for example, a porous material or, as indicated above, the line along which a wave-bearing surface lies. The precise form of (6) is subject to minor variations, sometimes $Y_n(y_0)$ appears rather than its derivative, however, E_n is invariably proportional to the derivative of the characteristic equation and inversely proportional to the wave number.

Equations (1)–(3) give the form of the wave field in each region of the duct. The modal coefficients A_j-D_j, $j=0,1,2,\ldots$ are, however, as yet undetermined. Mode matching is usually achieved by applying continuity of pressure and normal velocity at the interfaces between the component and inlet and/or outlet duct. This process is now demon-

strated. For ease of exposition, the inlet and outlet ducts are assumed to be rigid and plane wave forcing is to be assumed (although neither simplification is a requirement of the solution method). Thus, the duct modes are given by $Z_j(y)=\cos(j\pi y/b)$ and $F_j=\delta_{j0}$. It is convenient to make use of the symmetry of the duct geometry and consider separately the symmetric and antisymmetric subproblems. In both cases only the left-hand side of the system need be considered, and the conditions $\phi_{2x}(\ell,y)=0$ or $\phi_2(\ell,y)=0$ are applied along the line of symmetry for the symmetric and antisymmetric cases, respectively. There are two continuity conditions to be applied at $x=0$ and two ORs by which to do this. Which condition is enforced using which OR depends on the geometry. For $h>b$ continuity of normal velocity is enforced using (4) and continuity of pressure is enforced using the OR for the standard duct modes $\{\cos(n\pi y/b)|n=0,1,2,\ldots\}$. If $h<b$ the situation is reversed while for $h=b$ there is a choice. It is henceforth assumed, *without loss of generality*, that $h>b$. Thus, for the symmetric problem, continuity of pressure yields the following expression:

$$A_n^S=\frac{4}{b\epsilon_n}\sum_{j=0}^{\infty}B_j^S\cosh(s_j\ell)R_{nj}-\delta_{n0}. \tag{7}$$

Here A_n^S are the reflection coefficients, B_j^S are the amplitude of the waves inside the component for this subproblem, and

$$R_{jm}=\int_0^b\cos\left(\frac{j\pi y}{b}\right)Y_m(y)dy. \tag{8}$$

On using (5), continuity of normal velocity yields

$$B_m^S=\frac{1}{2s_mE_m\sinh(s_m\ell)}\times\left\{-iR_{0m}+\sum_{j=0}^{\infty}A_j^S\eta_jR_{jm}-2Y'_m(y_0)\sum_{n=0}^{N}a_n^S\gamma_m^{2n}\right\}. \tag{9}$$

It is important to note that for a component with a Sturm-Liouville eigensystem $a_n^S=0$, $n=0,1,\ldots,N$. In contrast, for one involving, for example, an elastic plate, these coefficients are nonzero and must be determined by applying appropriate edge conditions such as zero-displacement and gradient at the plate edges. It is straightforward to eliminate B_m^S between (9) and (7) to obtain

$$A_m^S=\frac{2}{b\epsilon_m}\sum_{n=0}^{\infty}A_n^S\eta_n\Omega_{nm}-\frac{2i}{b\epsilon_m}\Omega_{0m}-\frac{4}{b\epsilon_m}\sum_{n=0}^{N}a_n^S\sum_{j=0}^{\infty}\frac{\coth(s_j\ell)}{s_jE_j}\gamma_j^{2n}Y'_j(y_0)R_{mj}-\delta_{m0}, \tag{10}$$

where

$$\Omega_{nm}=\sum_{j=0}^{\infty}\frac{\coth(s_j\ell)}{s_jE_j}R_{mj}R_{nj}. \tag{11}$$

For the antisymmetric problem a system of equations with identical structure is obtained, the only difference is that Ω_{nm} is replaced by Λ_{nm} which is given by (11) but with $\coth(s_j\ell)$ replaced by $\tanh(s_j\ell)$. This replacement also applies to any sums on the right-hand side of (10). Note that the form of Ω_{nm} may vary slightly. For example, if continuity of normal velocity and pressure are enforced using the opposite choice of OR to that taken here the wave number s_j will occur in the numerator of the summand rather than the denominator and this does occur in the example implemented in the next section. The solution to the full problem is obtained from the symmetric and antisymmetric subproblems simply by noting that $A_n=A_n^S+A_n^A$ and $D_n=A_n^S-A_n^A$. Equation (10) and its antisymmetric counterpart are of a form that is generic to any three-part ducting system of the class considered here. The systems may be solved by truncation and numerical inversion of the matrix. Accurate evaluation of the quantities Ω_{jn} and Λ_{jn} [and any further sums on the right-hand side of (10)] for all required frequencies is crucial to this process but, in their current form, this depends on locating sufficient roots, s_j. Note, however, that these roots occur only within these sums and that the quantity E_m can be eliminated using (6). Thus, Ω_{nm} can be expressed as

$$\Omega_{nm}=2\sum_{j=0}^{\infty}\frac{\coth(s_j\ell)}{Y'_j(y_0)\frac{d}{ds}K(s)\big|_{s=s_j}}R_{mj}R_{nj}. \tag{12}$$

Once expressed in this form it can be seen that all such sums exhibit two important features. First, the summands are even functions of s_j. This is due to the fact that the governing equation (Helmholtz') contains only even powers of both x and y, and all boundary conditions for the component region are either Sturm-Liouville or contain only even higher derivatives in x. Second, each sum has the form of an infinite sum of the residues arising from an *odd* integrand. It is the latter fact that ensures that each sum can be recast in a form that is independent of the roots, s_j, of the characteristic equation for the component region. All that is required is to choose an appropriate contour integral. The process is demonstrated in the next section.

III. A SIMPLE DISSIPATIVE SILENCER

In this section the process of expressing Ω_{jn} and Λ_{jn} in root-free form is demonstrated for a two dimensional, three-part ducting system in which the middle component is a simple dissipative silencer of the type that is commonly found in a typical HVAC ducting system. Thus, the section of duct lying in the region $0<\bar{x}\leq 2\bar{\ell}$ is lined with a porous material which occupies the space $a\leq y\leq b$, $b>a$. The porous media is modelled as an equivalent fluid with complex speed of propagation c_p and complex density $\rho_p=Z_a/c_p$ where Z_a is the impedance of the material. Much experimental work has been done on relating c_p and ρ_p to the bulk acoustic properties of real porous materials and the complex values of these parameters are known in principle.[12]

A. The mode-matching solution

The velocity potential in the silencer region ($0\leq x\leq 2\ell$) must satisfy Helmholtz's equation with unit wave number in the central passage, $0\leq y<a$ and with wave number c_f/c_p in the porous media $a<y\leq b$. At the interface be-

tween the porous media and the fluid ($y=a$) it is assumed that the pressure and normal velocity are continuous while at the rigid duct walls the normal velocity is zero. The travelling wave forms in a duct of this type are easily determined by using separation of variables. A typical mode travelling in the positive x direction has the form $\phi_{2n}(x,y)=Y_n(y)e^{-s_n x}$ where

$$Y_n(y)=\begin{cases} Y_{1n}(y), & 0\leqslant y<a, \\ Y_{2n}(y), & a\leqslant y\leqslant b, \end{cases} \tag{13}$$

and the wave number s_n is defined below as a root of (21). The eigenfunctions $Y_n(y)$, $n=0,1,2,\ldots$ are the solutions to the eigensystem

$$Y_1''-\gamma^2 Y_1=0, \quad 0\leqslant y<a, \tag{14}$$

$$Y_1'(0)=0, \tag{15}$$

$$Y_2''-\lambda^2 Y_2=0, \quad a\leqslant y<b, \tag{16}$$

$$Y_2'(b)=0, \tag{17}$$

$$Y_1(a)=\beta Y_2(a), \tag{18}$$

$$Y_1'(a)=Y_2'(a), \tag{19}$$

where $\beta=\rho_p/\rho$, $\gamma(s)=(-1-s^2)^{1/2}$ and $\lambda(s)=(\Gamma^2-s^2)^{1/2}$ with $\Gamma=ic_f/c_p$, the branches being chosen such that $\gamma(0)=+i$ and $\lambda(0)=+\Gamma$.

It is readily shown that

$$Y_n(y)=\begin{cases} \cosh(\gamma_n y), & 0\leqslant y<a, \\ \dfrac{\gamma_n \sinh(\gamma_n a)}{\lambda_n \sinh(\lambda_n d)}\cosh[\lambda_n(y-b)], & a\leqslant y\leqslant b. \end{cases} \tag{20}$$

Here, $d=b-a$, $\gamma_n=\gamma(s_n)$, $\lambda_n=\lambda(s_n)$, and s_n, $n=0,1,2,\ldots$ are those roots of the dispersion relation $K(s)=0$ with $\Re(s_n)\geqslant 0$, where

$$K(s)=\cosh(\gamma a)+\frac{\beta\gamma\sinh(\gamma a)}{\lambda\sinh[\lambda(b-a)]}\cosh[\lambda(b-a)]. \tag{21}$$

Note that the roots s_n, $n=0,1,2,\ldots$ are numbered by increasing real part. Thus, the mode with wave number s_0 is the least attenuated.

The orthogonality relation[13] for this set of eigenfunctions is

$$\int_0^a Y_{1n}Y_{1m}\,dy+\beta\int_a^b Y_{2n}Y_{2m}\,dy=\delta_{mn}E_n, \tag{22}$$

where δ_{mn} is the usual Kronecker delta and E_n is given by (6) with $y_0=a$.

The velocity potentials ϕ_j, $j=1,2,3$ are given by (1)–(3) with $Z_n(y)=\cos(n\pi y/b)$, $n=0,1,2,\ldots$ and $\eta_j=(j^2\pi^2/b^2-1)^{1/2}$ with $\eta_0=i$, these definitions being appropriate for the choice of time dependence. Numerical results for both plane wave forcing and a multimodal incident field will be presented in the next section. Thus, the following analysis is carried out under the assumption of multimodal forcing, but without stating explicitly the form of the modal amplitudes. The appropriate form for both types of forcing are stated in Sec. IV. The mode-matching procedure described in Sec. II yields the following systems of equations:

$$\chi_n=F_n\eta_n^{1/2}-\frac{2}{b\epsilon_n\eta_n^{1/2}}\sum_{j=0}^{\infty}(\chi_j+F_j\eta_j^{1/2})\frac{\Lambda_{jn}}{\eta_j^{1/2}} \tag{23}$$

and

$$\psi_n=F_n\eta_n^{1/2}-\frac{2}{b\epsilon_n\eta_n^{1/2}}\sum_{j=0}^{\infty}(\psi_j+F_j\eta_j^{1/2})\frac{\Omega_{jn}}{\eta_j^{1/2}}, \tag{24}$$

where $\chi_n=(A_n+D_n)\eta_n^{1/2}$; $\psi_n=(A_n-D_n)\eta_n^{1/2}$ and R_{jm} is given by (8). Note that, the explicit form for R_{jm} for this problem is given in Eq. (33). For this coupled system of equations

$$\Lambda_{jn}=\sum_{m=0}^{\infty}\frac{s_m}{E_m}\tanh(s_m\ell)R_{jm}R_{nm}, \tag{25}$$

and Ω_{jn} is given by (25) with $\tanh(s_m\ell)$ replaced by $\coth(s_m\ell)$. The coefficients B_n and C_n, $n=0,1,2,\ldots$ are expressed in terms of A_n and D_n by

$$B_m+C_m=\frac{1}{E_m}\sum_{j=0}^{\infty}(A_j+F_j)R_{jm} \tag{26}$$

and

$$B_m e^{-2s_m\ell}+C_m e^{2s_m\ell}=\frac{1}{E_m}\sum_{j=0}^{\infty}D_j R_{mj}. \tag{27}$$

B. Recasting the matrix elements

As discussed in Sec. II, expression (25) can be recast in a form whereby the summation takes place over eigenvalues that can be expressed in closed form. The key is expression (6) which relates the quantity E_n to the derivative of the dispersion relation. This enables (25) to be recognized as the sum over a family of poles for a carefully selected integral. The appropriate integral is

$$I_{jm}=\lim_{X\to\infty}\int_{-iX}^{iX}\frac{\ell\Upsilon(s)L_j(s)L_m(s)}{\coth(s\ell)}ds, \tag{28}$$

where $X\gg 1$ is real. The quantities $\Upsilon(s)$ and $L_j(s)$ are defined by

$$\Upsilon(s)=\frac{s^2\gamma\sinh(\gamma a)}{\ell K(s)} \tag{29}$$

and

$$L_j(s)=\frac{P_j(s)}{\gamma^2+(j\pi/b)^2}-\frac{Q_j(s)}{\lambda^2+(j\pi/b)^2}, \tag{30}$$

with $\gamma=(-1-s^2)^{1/2}$ and $\lambda=(\Gamma^2-s^2)^{1/2}$, and

$$P_j(s)=\cos\left(\frac{j\pi a}{b}\right)+\frac{j\pi}{b}\sin\left(\frac{j\pi a}{b}\right)\frac{\cosh(\gamma a)}{\gamma\sinh(\gamma a)}, \tag{31}$$

$$Q_j(s) = (-1)^j \left\{ \cos\left(\frac{j\pi d}{b}\right) + \frac{j\pi}{b} \sin\left(\frac{j\pi d}{b}\right) \frac{\cosh(\lambda d)}{\lambda \sinh(\lambda d)} \right\}. \tag{32}$$

The choices of $P_j(s)$ and $Q_j(s)$ are such that $L_j(s)$ does not have poles when $\gamma^2+(j\pi/b)^2=0$ or $\lambda^2+(j\pi/b)^2=0$ (provided $j>0$). Further, although $Q_j(s)$ seems to be written in a slightly cumbersome form, it is this form that ensures that $Q(\tau_j)=0$ by inspection (i.e., without recourse to any trigonometric relations) where $\tau_j=(\Gamma_2+j^2\pi^2/b^2)^{1/2}$. Note that, the function $L_j(s)$ is related to the quantity R_{jn}, see (8), by

$$\gamma_n \sinh(\gamma_n a) L_j(s_n) = R_{jn}. \tag{33}$$

The path of integration in (28) lies along the imaginary axis and is indented to the left (right) of any poles on the upper (lower) half of the imaginary axis. Therefore, since the integrand is an odd function of s, $I_{jm}=0$ as $|X| \to \infty$. Note also that the integrand is $0(s^{-3})$ as $|s| \to \infty$. Thus, on deforming the path of integration onto a semicircular arc of radius X on the right-hand half-plane, the sum of the residues of all the poles crossed is zero as $X \to \infty$. The integrand has families of poles when (i) $K(s)=0$, i.e., $s=s_n$; (ii) $\cosh(s\ell)=0$, i.e., $s=\sigma_{1n}=(2n+1)i\pi/(2\ell)$; (iii) $\sinh(\lambda d)=0$, i.e., $s=\nu_n=(\Gamma^2+n^2\pi^2/d^2)^{1/2}$; (iv) $\gamma \sinh(\gamma a)=0$, i.e., $s=\theta_n=(n^2\pi^2/a-1)^{1/2}$ where, in each case, $n=0,1,2,\ldots$. Note that, as mentioned above, the forms of $P_j(s)$ and $Q_j(s)$ ensure that the quantity $L_j(s)$ has no poles when $\gamma^2+(j\pi/b)^2=0$ or $\lambda^2+(j\pi/b)^2=0$ provided $j>0$. The cases $j=0$, $m\neq 0$; $m=0$, $j\neq 0$ and $m=j=0$, however, need separate consideration. For these values of m and j the families of poles denoted (ii) and (iii) above vanish leaving, in both cases, only the simple pole corresponding to $n=0$.

The first family of poles yields Λ_{jm} and, on evaluating all the other pole contributions, it is found that

$$\begin{aligned} \frac{1}{2}\Lambda_{jm} = &- \sum_{n=0}^{\infty} \Upsilon(\sigma_{1n}) L_j(\sigma_{1n}) L_m(\sigma_{1n}) \\ &+ \frac{(-1)^{j+m}}{\beta d} \sum_{n=0}^{\infty} \frac{\nu_n V_{jn}(d) V_{mn}(d)}{\epsilon_n \coth(\nu_n \ell)} \\ &+ \sum_{n=0}^{\infty} \frac{\theta_n V_{jn}(a) V_{mn}(a)}{a\epsilon_n \coth(\theta_n \ell)}, \end{aligned} \tag{34}$$

where

$$V_{mn}(x) = \begin{cases} \dfrac{\frac{m\pi}{b}\sin(m\pi x/b)}{[(m\pi/b)^2 - (n\pi/x)^2]}, & m \neq n, \quad mx \neq nb, \\ x, & m = n = 0, \\ \dfrac{x}{2}(-1)^{m+n}, & mx = nb, \quad m \neq 0 \end{cases} \tag{35}$$

with x real. Note first that the second sum on the right-hand side of (34) is of identical structure to the third sum but with θ_n replaced by ν_n and a by $d=b-a$. Had $Q_j(s)$ not been defined as in (32) the trigonometric terms in the numerator of the second summand would have contained a rather than d producing an incorrect value when $md=nb$ and $m=j=0$. Note second that, for the cases $j=0$, $m\neq 0$; $m=0$, $j\neq 0$ and $m=j=0$ the second and third sums of (34) both reduce to one term. This is consistent with the vanishing of the families of poles denoted by (ii) and (iii) for these values of m and j as discussed above.

A similar expression for Ω_{jm} can be obtained using the integral (28) but with the term $\coth(s\ell)$ in the denominator of the integrand replaced with $\tanh(s\ell)$. The expression for Ω_{jm} is identical in structure to (34) but every occurrence of coth is replaced by tanh and every occurrence of σ_{1n} is replaced by $\sigma_{2n}=n\pi i/\ell$, $n=0,1,2,\ldots$. Having recast the matrix elements into a form that is independent of the roots of (21), Eqs. (23) and (24) may be truncated and solved numerically to determine the amplitudes A_n and D_n.

C. The power balance

In the preceding section it was shown that it is possible to fully determine the complex amplitudes A_n and D_n without solving the complicated dispersion relation $K(s)=0$ where $K(s)$ is given by (21). Clearly the incident, reflected and transmitted powers are independent of the roots of the dispersion relation and can thus easily be determined. The expressions for these quantities are well known, that for the incident power is

$$P_{\text{Inc}} = \frac{1}{2}\Re\left\{ \sum_{m=0}^{\infty} |F_m|^2 i \eta_m^* \epsilon_m \right\}, \tag{36}$$

where * indicates the complex conjugate. The expressions for the reflected and transmitted powers, P_{ref} and P_{Trans}, are identical but with F_m, $m=0,1,2,\ldots$ replaced by A_m and D_m, respectively. Note that these expressions are given in the form of power flux per unit height. Further, F_m, $m=0,1,2,\ldots$ are defined such that the incident power as given by (36) is unity.

It is a trivial matter to deduce that the power absorbed by the layer of lining is $P_{\text{Inc}}-P_{\text{Ref}}-P_{\text{Trans}}$. Such an expression, however, is inadequate in two respects. First, it does not give any indication of the power flux across each of the three surfaces of the absorbent lining. Second, it does not complete the power balance and thus, the opportunity to implement an algorithmic check on the algebra is missed. In fact, it is straightforward to calculate the power flux across the vertical surfaces at $a\leqslant y\leqslant b$, $x=0$ and $a\leqslant y\leqslant b$, $x=2\ell$ using the nondimensional flux integral

$$P = \frac{1}{b}\Re\left\{ -i\int_a^b \phi_j \left(\frac{\partial \phi_j}{\partial x}\right)^* dy \right\}, \tag{37}$$

where $j=1$ for the surface at $x=0$ and $j=3$ for the surface at $x=2\ell$. It is found that the power absorbed across $x=0$ is

$$P_{x=0} = \Re\left\{ \frac{i}{b} \sum_{n=0}^{\infty} \sum_{j=0}^{\infty} (F_n + A_n)(F_j - A_j)^* \eta_j^* I_{jn} \right\}, \tag{38}$$

while that absorbed across the surface at $x=2\ell$ is

$$P_{x=2\ell} = \Re\left\{ -\frac{i}{b}\sum_{n=0}^{\infty}\sum_{j=0}^{\infty} D_n D_j^* \eta_j^* I_{jn} \right\}, \tag{39}$$

where

$$I_{jn} = \int_a^b \cos\left(\frac{j\pi y}{b}\right)\cos\left(\frac{n\pi y}{b}\right) dy \tag{40}$$

which is easily evaluated and independent of the roots $s_n, n=0,1,2,....$ The expressions for $P_{x=0}$ and $P_{x=2\ell}$ given above were calculated using ϕ_1 and ϕ_3, respectively, and are the simplest formulas for the power flux across the vertical surfaces. It would, however, have been equally appropriate to use ϕ_2 [in which case (37) must be multiplied by $\beta=\rho_p/\rho$ to account for the difference in density of the absorbent lining and the acoustic medium]. This yields equivalent expressions in terms of the amplitudes B_n and C_n. It is found that

$$P_{x=0} = \Re\left\{ i\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \frac{\Delta_{mn} G_{nm}^*(0)}{\lambda_m^2 - (\lambda_n^*)^2} \right\}, \tag{41}$$

where

$$\Delta_{mn} = \frac{\beta}{\beta^*}\gamma_m \sinh(\gamma_m a)\cosh(\gamma_n^* a) - \gamma_n^* \sinh(\gamma_n^* a)\cosh(\gamma_m a) \tag{42}$$

and

$$G_{mn}(x) = s_m (C_m e^{2s_m x} - B_m e^{-2s_m x})(B_n^* e^{-2s_n^* x} + C_n^* e^{2s_n^* x}). \tag{43}$$

The power flux across the surface at $x=2\ell$ is given by (41) with $G_{mn}^*(0)$ replaced by $-G_{mn}^*(\ell)$. These expressions depend on the wave numbers s_n and the amplitudes B_n and C_n, however, they can be recast in a form that is independent of these quantities. The new forms depend only on A_n, D_n, F_n and the independent eigenvalues ν_n, σ_{1n}, and σ_{2n} defined in Sec. III B. Details of the rearrangement are given in the appendix where it is shown that

$$P_{x=0} = \Re\left\{ \frac{i}{b}\sum_{j=0}^{\infty}\sum_{n=0}^{\infty} [\tilde{A}_j^* \tilde{A}_n S1_{jn} - \tilde{A}_j^* D_n S2_{jn}] \right\} \tag{44}$$

and

$$P_{x=2\ell} = \Re\left\{ \frac{i}{b}\sum_{j=0}^{\infty}\sum_{n=0}^{\infty} [D_j^* D_n S1_{jn} - \tilde{A}_n D_j^* S2_{jn}] \right\}, \tag{45}$$

where $\tilde{A}_j = (A_j + F_j)$ and

$$S1_{jn} = \sum_{m=0}^{\infty}\left(-\frac{\Upsilon(\sigma_m) L_n(\sigma_m) Q_j(\sigma_m)}{\lambda^2(\sigma_m) + (j\pi/b)^2} - \frac{2(-1)^{n+j}}{\beta d}\frac{\nu_m V_{nm}(d) V_{jm}(d)}{\epsilon_m \tanh(2\nu_m \ell)} \right) \tag{46}$$

and

$$S2_{jn} = \sum_{m=0}^{\infty}\left(\frac{\Upsilon(\sigma_m) L_n(\sigma_m) Q_j(\sigma_m)}{(-1)^{m+1}[\lambda^2(\sigma_m) + (j\pi/b)^2]} - \frac{2(-1)^{n+j}}{\beta d}\frac{\nu_m V_{nm}(d) V_{jm}(d)}{\epsilon_m \sinh(2\nu_m \ell)} \right). \tag{47}$$

Note that, $\sigma_m = im\pi/(2\ell)$ which comprises both σ_{1m} and σ_{2m} as defined in Sec. III B.

The power flux across the horizontal surface of the lining can be expressed only in terms of ϕ_2. The appropriate, non-dimensional, flux integral can be evaluated to obtain

$$P_{y=a} = \Re\left\{ \frac{i}{b}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\cosh(\gamma_m a)\gamma_n^* \sinh(\gamma_n^* a)}{s_m^2 - (s_n^*)^2} \times [G_{nm}^*(\ell) + G_{mn}(0) - G_{mn}(\ell) - G_{nm}^*(0)] \right\}. \tag{48}$$

This expression can also be re-expressed in a form that is independent of B_m, C_m, and s_m. The new form depends on A_n, D_n, F_n and the independent eigenvalues θ_n, σ_{1n}, and σ_{2n} defined in Sec. III B. Again, details of the rearrangement, which involves some complicated manipulations, are presented in the appendix. The appropriate form for the power flux across the horizontal surface of the lining is stated here as

$$P_{y=a} = \Re\left\{ \frac{i}{b}\left(\sum_{n=0}^{\infty}\sum_{j=0}^{\infty} [\tilde{A}_n D_j^* + D_n \tilde{A}_j^*] S3_{jn} - [D_n D_j^* + \tilde{A}_n \tilde{A}_j^*] S4_{jn} \right) \right\}, \tag{49}$$

where

$$S3_{jn} = \sum_{m=0}^{\infty}\left(\frac{\Upsilon(\sigma_m) L_n(\sigma_m) P_j(\sigma_m)}{(-1)^{m+1}[\gamma^2(\sigma_m) + (j\pi/b)^2]} + \frac{2\theta_m V_{nm}(a) V_{jm}(a)}{a\epsilon_m \sinh(2\theta_m \ell)} \right) \tag{50}$$

and

$$S4_{jn} = \sum_{m=0}^{\infty}\left(-\frac{\Upsilon(\sigma_m) L_n(\sigma_m) P_j(\sigma_m)}{\gamma^2(\sigma_m) + (j\pi/b)^2} + \frac{2\theta_m V_{nm}(a) V_{jm}(a)}{a\epsilon_m \tanh(2\theta_m \ell)} \right) \tag{51}$$

with $\sigma_m = im\pi/(2\ell)$.

It is not difficult to show that $S4_{jn} - S1_{jn} = \frac{1}{2}\{\Lambda_{jn} + \Omega_{jn}\}$ and $S3_{jn} - S2_{jn} = \frac{1}{2}\{\Omega_{jn} - \Lambda_{jn}\}$. It follows from (44), (45), and (49) that the total power absorbed by the silencer lining is given by

$$P_{\mathrm{Abs}} = \Re\left\{ \frac{-i}{2b}\left(\sum_{j=0}^{\infty}\sum_{n=0}^{\infty} [\tilde{A}_j^* \tilde{A}_n + D_j^* D_n](\Lambda_{jn} + \Omega_{jn}) + [D_j^* \tilde{A}_n + \tilde{A}_j^* D_n](\Lambda_{jn} - \Omega_{jn}) \right) \right\}. \tag{52}$$

This expression has been calculated directly from the flux integrals and is independent of the reflected and transmitted

powers. Thus, the power balance can now be stated as

$$P_{\mathrm{Ref}} + P_{\mathrm{Trans}} + P_{\mathrm{Abs}} = P_{\mathrm{Inc}}, \tag{53}$$

where P_{Inc}, P_{Ref}, P_{Trans}, and P_{Abs} are given by (36) and (52), respectively.

For problems involving acoustic transmission in nondissipative conditions, it is well established that the systems of equations derived via mode-matching or similar methods automatically satisfy the power balance *regardless* of the level of truncation used during numerical solution.[8,9] Further, Warren *et al.*[9] demonstrate that even if the system of equations is truncated radically, so that fewer equations are retained than the number of cut-on modes, the power balance is still satisfied. (Although clearly, in this case, the distribution of power between the reflected and transmitted components will be incorrect.) This indicates that, while it is necessary that the various components of power should balance this, in itself, is not sufficient to ensure that the numerical solution obtained via truncation and inversion represents the actual solution to the physical problem under consideration. The authors are not aware of any equivalent results in the literature for a dissipative system and, in view of the importance of this issue, it is now shown that the system of equations, (23)–(25), automatically satisfies the power balance regardless of the level of truncation. It is assumed that the system of equations is truncated to $T+1$ terms where $T>0$. Thus, on adding and subtracting (23) and (24), it is found that

$$A_m \eta_m = F_m \eta_m - \frac{1}{b\epsilon_m} \sum_{j=0}^{T} [\tilde{A}_j(\Lambda_{jm} + \Omega_{jm}) + D_j(\Lambda_{jm} - \Omega_{jm})] \tag{54}$$

and

$$D_m \eta_m = -\frac{1}{b\epsilon_m} \sum_{j=0}^{T} [\tilde{A}_j(\Lambda_{jm} - \Omega_{jm}) + D_j(\Lambda_{jm} + \Omega_{jm})], \tag{55}$$

where $m=0,1,2,\ldots,T$ and $\tilde{A}_j=(A_j+F_j)$. Now the sum of the reflected and transmitted powers can be written as

$$P_{\mathrm{Ref}} + P_{\mathrm{Trans}} = \frac{1}{2}\left\{ -i \sum_{m=0}^{T} [A_m^*(A_m \eta_m)\epsilon_m + D_m^*(D_m \eta_m)\epsilon_m] \right\} \tag{56}$$

and on using (54) and (55), this becomes

$$P_{\mathrm{Ref}} + P_{\mathrm{Trans}} = \frac{1}{2}\Re\left\{ -i\left(\sum_{m=0}^{T} A_m^* F_m \eta_m \epsilon_m - \frac{1}{b}\sum_{m=0}^{T}\sum_{j=0}^{T} [A_m^* \tilde{A}_j + D_m^* D_j](\Lambda_{jm} + \Omega_{jm}) - \frac{1}{b}\sum_{m=0}^{T}\sum_{j=0}^{T} [A_m^* D_j + D_m^* \tilde{A}_j](\Lambda_{jm} - \Omega_{jm}) \right) \right\}. \tag{57}$$

Since only the real part is of interest, the first term on the right-hand side of (57) can be replaced with its conjugate. Then, on using (54) to eliminate A_m from this term, it is found that

$$\begin{aligned} P_{\mathrm{Ref}} + P_{\mathrm{Trans}} &= \frac{1}{2}\Re\left\{ i \sum_{m=0}^{T} \eta_m^* F_m^* F_m \epsilon_m - \frac{i}{b}\sum_{m=0}^{T}\sum_{j=0}^{T} \times \frac{\eta_m^*}{\eta_m}[F_m^* \tilde{A}_j(\Lambda_{jm} + \Omega_{jm}) + D_j F_m^*(\Lambda_{jm} - \Omega_{jm})] \right\} \\ &\quad + \frac{1}{2}\Re\left\{ \frac{i}{b}\sum_{m=0}^{T}\sum_{j=0}^{T} [A_m^* \tilde{A}_j + D_m^* D_j](\Lambda_{jm} + \Omega_{jm}) + \frac{i}{b}\sum_{m=0}^{T}\sum_{j=0}^{T} [A_m^* D_j + D_m^* \tilde{A}_j](\Lambda_{jm} - \Omega_{jm}) \right\}. \end{aligned} \tag{58}$$

On referring to (59) it is seen that $F_j=0$ for $j \geqslant M_I$ where M_I is the number of cut-on modes, while η_j is imaginary for $j < M_I$. Thus, without loss of generality the quantity η_m^*/η_m on the right-hand side of (58) can be replaced by -1. Then, on collecting together the terms involving $\Lambda_{jm} \pm \Omega_{jm}$, it becomes apparent that the right-hand side of (58) comprises the incident power and that absorbed by the silencer, see (36) and (52), where both expressions are truncated to $T+1$ terms. Hence, it is shown that the power balance is an algebraic identity and in no way guarantees that the numerical solution has converged to that representing the physical problem.

IV. NUMERICAL RESULTS

All the graphical results presented in this section are calculated using the root-free expressions given in Sec. III. Two silencer configurations are studied. These are identical in height ($\bar{b}=0.6$ m) and half-length ($\bar{\ell}=1.5$ m) and are lined with the same absorbent material. The difference lies in the depth of the absorbent liner. For silencer 1 the absorbent layer is thick at 0.45 m, thus $\bar{a}=0.15$ m. For silencer 2 the depth of the absorbent layer is much thinner at 0.15 m so that $\bar{a}=0.45$ m. The absorbent material is characterized by the regression formulas of Delany and Bazley[12,7] with flow resistivity of 8000 rayl/m. Note, however, that the formulas of Delany and Bazley are known to be invalid at low frequency and so the semiempirical correction of Kirby and Cummings[12] are then used.

The usual measure of performance for a dissipative silencer is transmission loss: $\mathcal{L}=-10\log_{10}(P_{\mathrm{Trans}}/P_{\mathrm{Inc}})$, where P_{Trans} and P_{Inc} are given by (36). The incident field as defined by (1) may be either plane wave, in which case the modal amplitudes are given by $F_j=\delta_{j0}$, or multimodal. Mechel[14] suggests that "equal modal energy density" (EMED) is the most plausible form of multimodal forcing for this class of system and, under this assumption, the modal amplitudes are given by

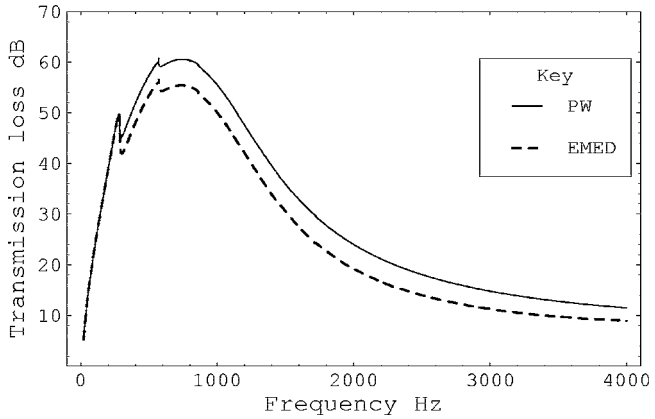


FIG. 2. Transmission loss against frequency for silencer 1. Both plane wave and EMED forcing are depicted.

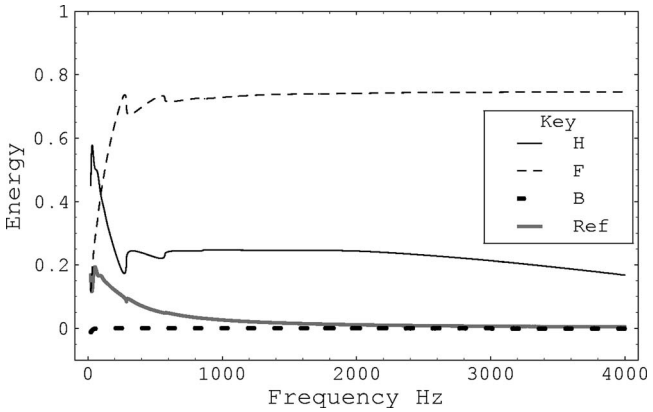


FIG. 3. Proportion of energy reflected and absorbed across each surface of silencer 1 for plane wave forcing. The $x=0$, $y=a$, and $x=2\ell$ surfaces of the silencer are denoted by F, H, and B, respectively. Ref indicates the reflected component of power.

$$F_j^2 = \begin{cases} \dfrac{2i}{\epsilon_j \sum_{m=0}^{M_I-1} \eta_m}, & j < M_I, \\ 0, & j \geqslant M_I, \end{cases} \tag{59}$$

where $\epsilon_j=2$ for $j=0$ and 1 otherwise. Note that, in (59) M_I is the number of waves cut-on in the inlet duct and will depend on $b=k\bar{b}$.

Figure 2 shows transmission loss against frequency for silencer 1. Both EMED and plane wave forcing are shown. At frequencies below 286 Hz the two curves overlie as is to be expected since EMED reduces to plane wave forcing below the first cut-on frequency. Thereafter the transmission loss is slightly higher for plane wave forcing, but the difference is at most 6 decibels. The results presented in Fig. 2 are validated in Table I where, for seven specified frequencies, the value of the transmission loss obtained using the root-free method is compared both with that obtained using conventional "rooty" mode-matching [i.e., solving the characteristic equation in order to obtain sufficient wave numbers, s_m, by which to accurately evaluate the quantities Ω_{jn} and Λ_{jn}] and a finite-element based point-collocation method.[7] All three methods show good agreement.

The power flux across each face of silencer 1, for plane wave and EMED forcing, are shown in Figs. 3 and 4, respectively. It is interesting to note that, although the overall transmission loss is not vastly different between the two types of forcing, the silencer surface primarily involved in the sound absorption depends on the type of forcing. For plane wave forcing (Fig. 3) it is the front face of the lining, i.e., the surface lying along $x=0$, $a \leqslant y \leqslant b$, across which the power flux is the greatest, although the flux across horizontal surface, i.e., that lying along $y=a$, $0 \leqslant x \leqslant 2\ell$, is also significant. Whereas for EMED forcing (Fig. 4), it is clear that the power flux across the horizontal face of the silencer is the greatest. In this case, particularly at frequencies greater than 1000 Hz, the energy absorbed across the front face of the silencer is relatively insignificant. For both forcing mechanisms and indeed both silencer configurations, the flux across the rear face of the silencer, i.e., that lying along $x=2\ell$, $a \leqslant y \leqslant b$, is negligible. Indeed as the frequency tends towards zero, energy actually leaks through this surface which manifests as a negative flux. The amount of power reflected for the two forcing mechanisms is similar, although the characteristic spikes at each cut-on frequency are more apparent in Fig. 4 (EMED forcing).

Figure 5 shows transmission loss against frequency for silencer 2. Again, both EMED and plane wave forcing are shown. As for silencer 1, the two curves overlie for frequencies below 286 Hz but, in this case, they also remain very close for frequencies up to 850 Hz. Thereafter, the two curves diverge but the maximum difference is still in the region of 5 decibels. There are two points to be noted. First, at all frequencies, the transmission loss for silencer 2 is significantly less than that of silencer 1. This is to be expected since silencer 2 has the thinner lining of the two silencers.

TABLE I. Silencer 1, EMED forcing: comparison of transmission loss data for the "root-free," "rooty," and point-collocation methods.

	Silencer 1: EMED forcing		
Frequency	Non-rooty	Rooty	Point collocation
63	16.346 889 36	16.344 785 68	16.569 007 93
125	27.249 628 96	27.246 795 55	27.306 260 97
250	46.213 500 06	46.213 520 61	46.185 140 37
500	53.414 369 69	53.414 101 53	53.404 507 71
1000	50.018 045 44	50.016 207 33	50.009 049 46
2000	19.115 798 43	19.115 733 66	19.113 501 38
4000	8.947 613 426	8.948 377 897	8.946 284 148

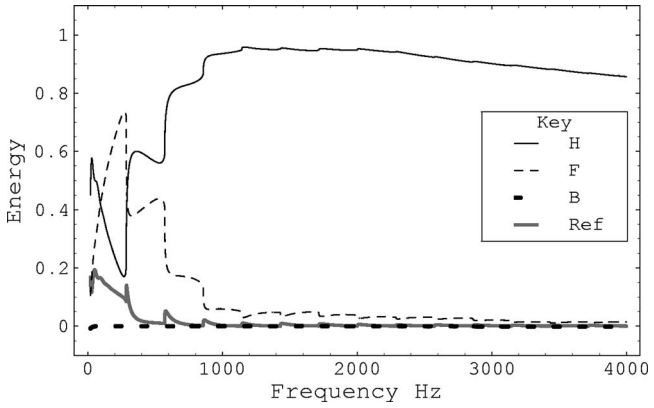


FIG. 4. Proportion of energy reflected and absorbed across each surface of silencer 1 for EMED forcing. The $x=0$, $y=a$, and $x=2\ell$ surfaces of the silencer are denoted by F, H, and B, respectively. Ref indicates the reflected component of power.

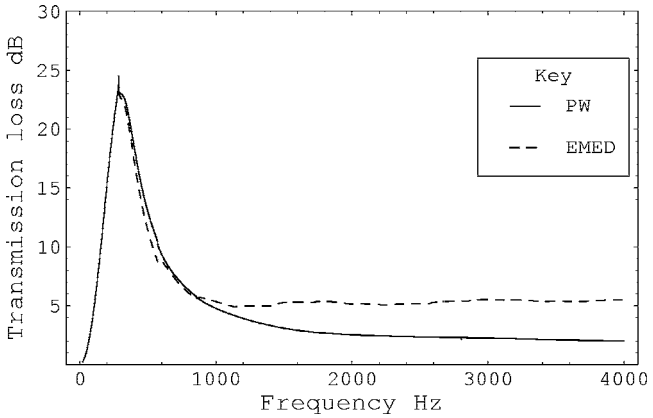


FIG. 5. Transmission loss against frequency for silencer 2. Both plane wave and EMED forcing are depicted.

Second, for this silencer the transmission loss is slightly higher for EMED forcing as opposed to plane wave. Again, the results presented in Fig. 5 are validated in Table II, by comparison with the "rooty" approach and point collocation.

Figures 6 and 7 show the power flux across the component surfaces of the silencer for plane wave and EMED forcing, respectively. For plane wave forcing both the front and the horizontal faces of the silencer are proactive, i.e., the power flux across them is significant. It cannot now be said that the front face is the most proactive. This is to be expected since the silencer lining is comparatively thin and there is, therefore, less sound incident directly onto the front face than with the thicker lining of silencer 1. For EMED forcing, however, it is clear that the horizontal surface accounts for the vast majority of the transmission loss as it does for silencer 1.

V. DISCUSSION

It has been demonstrated that a broad class of problem involving the transmission of sound through a two-dimensional, three-part ducting system can be successfully solved using analytic mode-matching, but without explicit knowledge of any of the roots of the characteristic equation for the "middle" region. Here the method was implemented for a three-part system comprising rigid inlet and outlet ducts with a simple silencer sandwiched between them. The systems of equations obtained via mode-matching were recast, using a contour integral technique, into a form that is independent of the roots to the characteristic equation for the silencer region. Robust and accurate root-free expressions by which to compute all the physical quantities of interest were obtained.

In order to use the method described in this article it is required only that there exists an appropriate orthogonality relation for the eigenfunctions of the "middle" or "component" region. For problems in which the inlet and/or oulet ducts are acoustically hard or soft the systems of equations obtained via mode-matching can be cast into a form that involves no root-finding. For situations in which the inlet and/or oulet ducts are bounded by wave-bearing surfaces such as a membrane or elastic plate, however, the approach is still of value. The underlying eigensystems for such inlet and/or oulet ducts are non-Sturm-Liouville but are well studied.[10,3] Further, although the admissible wave numbers must be determined numerically, this task is not usually onerous since the roots to the characteristic equation are known to lie only on either the real or imaginary axis and can be located with relatively little effort. Thus, for a three-part ducting system comprising inlet and/or outlet ducts with wave-bearing boundaries and a middle component of more complicated structure (possibly comprising both wave-bearing boundaries and layers of porous material) this method will bypass root-finding for the middle region thereby significantly reducing the overall burden of root-finding. Furthermore, it seems likely that this approach can be extended to three-part ducting systems with circular cy-

TABLE II. Silencer 2, EMED forcing: comparison of transmission loss data for the "root-free," "rooty," and point-collocation methods.

	Silencer 2: EMED forcing		
Frequency	Nonrooty	Rooty	Point collocation
63	1.828 077 049	1.829 027 604	1.855 084 98
125	6.784 509 519	6.786 852 05	6.822 066 684
250	20.639 404 65	20.641 849 64	20.693 451 16
500	11.401 092 96	11.401 302 63	11.401 5143
1000	5.355 438 508	5.355 675 135	5.356 555 514
2000	5.125 842 846	5.125 912 909	5.126 296 988
4000	5.504 838 985	5.504 690 444	5.504 658 887

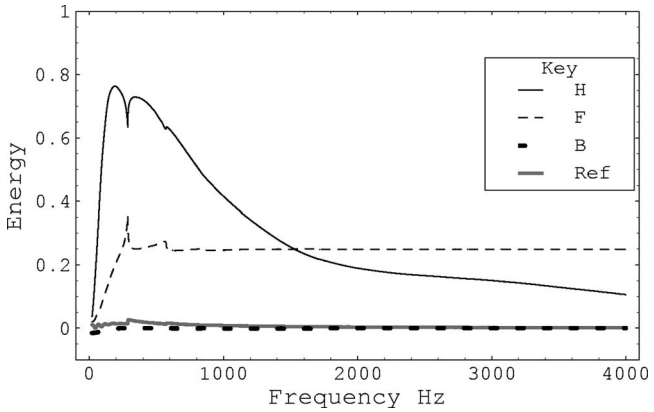


FIG. 6. Proportion of energy reflected and absorbed across each surface of silencer 2 for plane wave forcing. The $x=0$, $y=a$, and $x=2\ell$ surfaces of the silencer are denoted by F, H, and B, respectively. Ref indicates the reflected component of power.

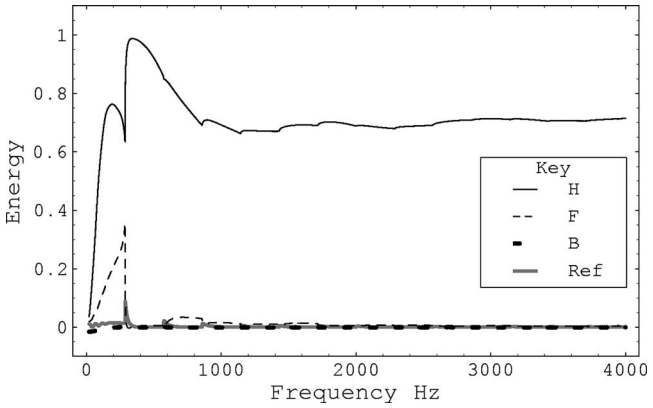


FIG. 7. Proportion of energy reflected and absorbed across each surface of silencer 2 for EMED forcing. The $x=0$, $y=a$, and $x=2\ell$ surfaces of the silencer are denoted by F, H, and B, respectively. Ref indicates the reflected component of power.

lindrical geometry comprising rigid inlet and/or outlet ducts. All that is required is that a suitable OR exists for the component region.

A minor disadvantage is that this approach, although highly accurate, is computationally slower than the root-finding alternative. This disadvantage may be offset, however, against the advantage of eliminating the inaccuracies that can arise due to missing roots. This is of particular importance when plotting physical quantities, such as transmission loss, against frequency.

APPENDIX: CALCULATION OF THE ABSORBED POWER

In Sec. III C the power absorbed across each of the three faces of the silencer lining was given, in terms of the roots s_n, $n=0,1,2,\ldots$ of the dispersion relation, by (41) and (48). These expressions can be recast into forms that depend only on A_n, D_n, $\sigma_n=in\pi/(2\ell)$ and $\nu_n=(\Gamma^2+n^2\pi^2/d^2)^{1/2}$ and which are, therefore, independent of s_n. Consider first the vertical faces, the first step in recasting (41) is to eliminate B_n and C_n using (26), (27), and the following expressions:

$$B_n - C_n = \frac{\coth(2s_n\ell)}{E_n}\sum_{j=0}^{\infty}\tilde{A}_j R_{jn} - \frac{1}{E_n\sinh(2s_n\ell)}\sum_{j=0}^{\infty} D_j R_{jn}, \tag{A1}$$

$$B_n e^{2s_n\ell} - C_n e^{-2s_n\ell} = -\frac{\coth(2s_n\ell)}{E_n}\sum_{j=0}^{\infty} D_j R_{jn} + \frac{1}{E_n\sinh(2s_n\ell)}\sum_{j=0}^{\infty}\tilde{A}_j R_{jn}, \tag{A2}$$

where $\tilde{A}_j=(A_j+F_j)$. Due to the similarity in the structures of (26) and (27) and also (A1) and (A2), it is only necessary to derive the root-free expression for (41). The appropriate result for the power absorbed across the vertical face at $x=2\ell$ can then be obtained by interchanging the coefficients D_j and $\tilde{A}_j$ (and likewise D_q and $\tilde{A}_q$) throughout. On eliminating B_n and C_n from (41) and interchanging the orders of summation, it is found that

$$P_{x=0} = \Re\left\{-\frac{i}{b}\sum_{j=0}^{\infty}\sum_{q=0}^{\infty}\left(\tilde{A}_j\tilde{A}_q^*\sum_{n=0}^{\infty}\frac{s_n^* R_{qn}^* T1_{jn}}{\tanh(2s_n^*\ell)E_n^*} - \tilde{A}_j D_q^*\sum_{n=0}^{\infty}\frac{s_n^* R_{qn}^* T1_{jn}}{\sinh(2s_n^*\ell)E_n^*}\right)\right\}, \tag{A3}$$

where

$$T1_{jn} = \sum_{m=0}^{\infty}\frac{\Delta_{mn}R_{jm}}{[\lambda_m^2-(\lambda_n^*)]E_m} = \frac{\gamma_n^*\sinh(\gamma_n^* a)Q_j(s_n^*)}{[(j\pi/b)^2+(\lambda_n^*)^2]} \tag{A4}$$

with Δ_{mn} defined by (42). Note that the right-hand side of (A4) was obtained using a similar contour integral approach to that described in Sec. III B. It follows that

$$P_{x=0} = \Re\left\{-\frac{i}{b}\sum_{j=0}^{\infty}\sum_{q=0}^{\infty}[\tilde{A}_j\tilde{A}_q^* S1_{jq}^* - \tilde{A}_j D_q^* S2_{jq}^*]\right\}, \tag{A5}$$

where

$$S1_{jq} = \sum_{n=0}^{\infty}\frac{s_n\gamma_n\sinh(\gamma_n a)R_{qn}Q_j(s_n)}{\tanh(2s_n\ell)E_n[(j\pi/b)^2+(\lambda_n)^2]} \tag{A6}$$

and $S2_{jq}$ is identical in structure to $S1_{jq}$ but with the term $\tanh(2s_n\ell)$ in the denominator of the summand replaced by $\sinh(2s_n\ell)$. Again, the method described in Sec. III B can be used to express the quantities $S1_{jq}$ and $S2_{jq}$ in the forms (46) and (47). For $S1_{jq}$, the appropriate integral is

$$J_{jq}^{(S1)} = \lim_{X\to\infty}\int_{-iX}^{iX}\frac{\ell\Upsilon(s)L_q(s)Q_j(s)}{\tanh(2s\ell)[(j\pi/b)^2+\lambda^2]}ds \tag{A7}$$

and for $S2_{jq}$ the integral is the same except the quantity $\tanh(2s\ell)$ in the denominator of the integrand is replaced by $\sinh(2s\ell)$. To set (A5) in exactly the same form as (44), it is necessary only to replace the summand with its conjugate and multiply the whole expression by -1. Finally, as mentioned above, the equivalent expression for the power absorbed by the vertical face at $x=2\ell$, that is

$P_{x=2\ell}$, is obtained simply by interchanging the quantities $\tilde{A}_j$ and D_j (likewise $\tilde{A}_q$ and D_q).

Now consider the power absorbed by the horizontal face of the silencer lining, this is given, in terms of the roots s_n, $n=0,1,2,\ldots$ of the dispersion relation, by (48). On interchanging the counters in the second and fourth sums, (48) may be expressed as

$$P_{y=a} = -\Re\left\{ \frac{i}{b}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\cosh(\gamma_m a)\gamma_n^* \sinh(\gamma_n^* a)}{s_m^2 - (s_n^*)^2} W_{mn} - \frac{i}{b}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\cosh(\gamma_n a)\gamma_m^* \sinh(\gamma_m^* a)}{s_n^2 - (s_m^*)^2} W_{mn}^* \right\}, \tag{A8}$$

where $W_{mn}=G_{mn}(\ell)-G_{mn}(0)$ and $G_{mn}(x)$ is defined in (43).

It is intended to deal only with the first and third terms of (A8), i.e., those containing ℓ. The equivalent results for the second and third terms can be deduced by interchanging $\tilde{A}_j$ and D_j (likewise $\tilde{A}_q$ and D_q) and changing the sign of the expression. Thus, on using (27) and (A2) to re-express the first and third terms of (A8) in terms of the coefficients D_j and $\tilde{A}_j$, and noting that

$$\left(\sum_{n=0}^{\infty} \frac{\gamma_n \sinh(\gamma_n a) R_{jn}}{E_n(s_n^2 - (s_m^*)^2)}\right)^* = -\frac{\gamma_m \sinh(\gamma_m a) L_j^*(s_m)}{K^*(s_m)} \tag{A9}$$

and

$$\sum_{n=0}^{\infty} \frac{\cosh(\gamma_n a) R_{jn}}{E_n(s_n^2 - (s_m^*)^2)} = -\frac{\cosh(\gamma_m^* a) L_j(s_m^*)}{K(s_m^*)} + \frac{P_j(s_m^*)}{(j\pi/b)^2 + (\gamma_m^*)^2}, \tag{A10}$$

where $K(s_m^*) \neq K^*(s_m)$ and $L(s_m^*) \neq L^*(s_m)$, it is found that

$$\Re\left\{ \frac{i}{b}\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{\cosh(\gamma_m a)\gamma_n^* \sinh(\gamma_n^* a) G_{mn}(\ell)}{s_m^2 - (s_n^*)^2} - \frac{\cosh(\gamma_n a)\gamma_m^* \sinh(\gamma_m^* a) G_{mn}^*(\ell)}{s_n^2 - (s_m^*)^2} \right\} = \Re\left\{ \frac{i}{b}\sum_{j=0}^{\infty}\sum_{q=0}^{\infty} [D_j \tilde{A}_q^* S3_{jq}^* - D_j D_q^* S4_{jq}^*] \right\}. \tag{A11}$$

Note that, the results given in (A9) and (A10) are again proven using contour integration. Now, on adding in the contributions from the terms in (A8) that do not include ℓ, it is found that

$$P_{y=a} = -\Re\left\{ \frac{i}{b}\sum_{j=0}^{\infty}\sum_{q=0}^{\infty} [D_j \tilde{A}_q^* + D_q^* \tilde{A}_j] S3_{jq}^* - [D_j D_q^* + \tilde{A}_j \tilde{A}_q^*] S4_{jq}^* \right\}. \tag{A12}$$

On taking the complex conjugate of the summand and changing the sign of the whole expression, this is easily recognized as (49).

Note that, in (A11), $S4_{jq}$ is given by

$$S4_{jq} = \sum_{m=0}^{\infty} \frac{s_m \gamma_m \sinh(\gamma_m a) R_{qm} P_j(s_m)}{\tanh(2 s_m \ell) E_m[(j\pi/b)^2 + (\gamma_m)^2]} \tag{A13}$$

and $S3_{jq}$ is identical in structure to $S4_{jq}$ but with $\tanh(2s_n\ell)$ replaced by $\sinh(2s_n\ell)$. These sums can be cast in the forms given in (50) and (51) using the integral approach described in Sec. III B. The appropriate integrals have a similar form to (A7).

[1] L. Huang, "A theoretical study of duct noise control by flexible panels," J. Acoust. Soc. Am. **106**, 1801–1809 (1999).
[2] L. Huang, "Modal anaylis of a drumlike silencer," J. Acoust. Soc. Am. **112**, 2014–2025 (2002).
[3] J. B. Lawrie and I. D. Abrahams, "An orthogonality condition for a class of problems with high order boundary conditions; applications in sound/structure interaction," Q. J. Mech. Appl. Math. **52**, 161–181 (1999).
[4] H. Besserer and P. G. Malischewsky, "Mode series expansions at vertical boundaries in elastic waveguides," Wave Motion **39**, 41–59 (2004).
[5] A. Selamet, M. B. Xu, I.-J. Lee, and N. T. Huff, "Analytic approach for sound attenuation in perforated dissipative silencers," J. Acoust. Soc. Am. **115**, 2091–2099 (2004).
[6] A. Cummings and N. Sormaz, "Acoustic attenuation in dissipative splitter silencers containing mean fluid flow," J. Sound Vib. **168**, 209–227 (1993).
[7] R. Kirby and J. B. Lawrie, "A point collocation approach to modelling large dissipative silencers," J. Sound Vib. **286**, 313–339 (2005).
[8] G. A. Kriegsmann, "The flanged waveguide antenna: discrete reciprocity and conservation," Wave Motion **29**, 81–95 (1999).
[9] D. P. Warren, J. B. Lawrie, and I. M. Mohamed, "Acoustic scattering in wave-guides with discontinuities in height and material property," Wave Motion **36**, 119–142 (2002).
[10] D. V. Evans and R. Porter, "Wave scattering by narrow cracks in ice sheets floating on water of finite depth," J. Fluid Mech. **484**, 143–165 (2003).
[11] J. B. Lawrie and I. D. Abrahams, "On the propagation and scattering of fluid-structural waves in a three-dimensional duct bounded by thin elastic walls," in *Proceedings of IUTAM 2000/10*, edited by I. D. Abrahams, P. A. Martin, and M. J. Simon (Kluwer, Berlin, 2002).
[12] R. Kirby and A. Cummings, "Prediction of the bulk acoustic properties of fibrous materials at low frequencies," Appl. Acoust. **56**, 101–125 (1999).
[13] R. Glav, "The point-matching method on dissipative silencers of arbitrary cross-section," J. Sound Vib. **189**, 123–135 (1996).
[14] F. P. Mechel, "Theory of baffle-type silencers," Acustica **70**, 93–111 (1990).

Impedance of a sphere oscillating in an elastic medium with and without slip

Andrew N. Norris[a)]
Department of Mechanical and Aerospace Engineering, Rutgers University, 98 Brett Road, Piscataway, New Jersey 08854-8058

(Received 5 December 2005; revised 9 January 2006; accepted 11 January 2006)

The dynamic impedance of a sphere oscillating in an elastic medium is considered. Oestreicher's [J. Acoust. Soc. Am. **23**, 707–714 (1951)] formula for the impedance of a sphere bonded to the surrounding medium can be expressed in a relatively simple form in terms of three lumped impedances associated with the displaced mass and the longitudinal and transverse waves. If the surface of the sphere slips while the normal velocity remains continuous, the impedance formula is modified by adjusting the definition of the transverse impedance to include the interfacial impedance. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2171526]

PACS number(s): 43.20.Rz, 43.40.Yq, 43.80.Ev, 43.20.Tb [DF] Pages: 2062–2066

I. INTRODUCTION

The dynamic impedance of a spherical particle embedded in a medium is important for acoustical measurement and imaging. The impedance is used, for instance, in measurement of the mechanical properties of tissue, e.g., Ref. 1, and is intimately related to the radiation forcing on particles.[2] The latter is the basis for imaging techniques such as vibroacoustography which has considerable potential in mammography for detection of microcalcifications in breast tissue.[3] Oestreicher[4] derived the impedance for a rigid sphere oscillating in a viscoelastic medium over 50 years ago. Although it was derived for a sphere in an infinite medium, Oestreicher's formula is also applicable, with minor modification,[5] to dynamical indentation techniques where the particle is in contact with the surface of the specimen; see Zhang *et al.*[6] for a review of related work. Chen *et al.*[7] recently validated Oestreicher's formula experimentally by measuring the dynamic radiation force on a sphere in a fluid. The impedance formula is based on perfect no slip conditions between the spherical inclusion and its surroundings. This may not always be a valid assumption, e.g., in circumstances where a foreign object is embedded in soft material. This was precisely the situation in recent measurements of the viscosity of DNA cross-linked gels by magnetic forcing on a small steel sphere.[8] This paper generalizes the impedance formula to include the possibility of dynamic slip.

Two related results are derived in this paper. The first is a modified form of Oestreicher's formula which enables it to be interpreted in terms of lumped parameter impedances. This leads to a simple means to consider the more general case of a sphere oscillating in a viscoelastic medium which is permitted to slip relative to its surroundings. The slip is characterized by an interfacial impedance which relates the shear stress to the discontinuity in tangential velocities. This generalization includes Oestreicher's original formula as the limit of infinite interfacial impedance, and agrees with previous results for the static stiffness of a spherical inclusion with and without slip.[9]

II. SUMMARY OF RESULTS

A sphere undergoes time harmonic oscillatory motion of amplitude u_0 in the direction $\hat{\boldsymbol{x}}$,

$$\boldsymbol{u}^{\text{sphere}} = u_0 e^{-i\omega t}\hat{\boldsymbol{x}}. \tag{1}$$

The time harmonic factor $e^{-i\omega t}$ is omitted but understood in future expressions. The sphere, which is assumed to be rigid and of radius a, is embedded in an elastic medium of infinite extent with mass density ρ and Lamé moduli λ and μ. The moduli may be real or complex, corresponding to an elastic or viscoelastic solid. We will later consider complex shear modulus $\mu=\mu_1-i\omega\mu_2$, where the imaginary term dominates in a viscous medium. The force exerted on the sphere by the surrounding medium acts in the $\hat{\boldsymbol{x}}$-direction, and is defined by

$$F\hat{\boldsymbol{x}} = \int_{r=a} \boldsymbol{T}\mathrm{d}s, \tag{2}$$

where $\boldsymbol{T}$ is the traction vector on the surface. The sphere impedance is defined

$$Z = \frac{F}{-i\omega u_0}. \tag{3}$$

Oestreicher's expression for the impedance of a sphere that does not slip relative to the elastic medium is[4] [Equation (4) is Oestreicher's[4] Eq. (18) with i replaced by $-i$ since he used time dependence $e^{i\omega t}$.]

$$Z = \frac{4}{3}\pi a^3\rho i\omega\left[\left(1+\frac{3i}{ah}-\frac{3}{a^2h^2}\right)-2\left(\frac{-i}{ah}+\frac{1}{a^2h^2}\right)\times\left(3-\frac{a^2k^2}{1-iak}\right)\right]\Bigg/\left[\left(\frac{-i}{ah}+\frac{1}{a^2h^2}\right)\frac{a^2k^2}{1-iak}+\left(2-\frac{a^2k^2}{1-iak}\right)\right]. \tag{4}$$

a)Electronic mail: norris@rutgers.edu

 0001-4966/2006/119(4)/2062/5/$22.50

Here k and h are, respectively, the longitudinal and transverse wave numbers, $k=\omega/c_L$, $h=\omega/c_T$ with $c_L=\sqrt{(\lambda+2\mu)/\rho}$ and $c_T=\sqrt{\mu/\rho}$.

Noting that Oestreicher's formula can be rewritten

$$Z=\frac{4}{3}\pi a^3\rho i\omega\left\{-1+\left[\frac{1}{3}\left(1-\frac{3(1-ika)}{k^2a^2}\right)^{-1}+\frac{2}{3}\left(1-\frac{3(1-iha)}{h^2a^2}\right)^{-1}\right]^{-1}\right\}, \tag{5}$$

implies our first result, that the impedance satisfies

$$\frac{3}{Z+Z_m}=\frac{1}{Z_L+Z_m}+\frac{2}{Z_T+Z_m}, \tag{6}$$

where the three additional impedances are defined as

$$Z_m=i\omega\tfrac{4}{3}\pi a^3\rho, \tag{7a}$$

$$Z_L=(i\omega)^{-1}4\pi a(\lambda+2\mu)(1-ika), \tag{7b}$$

$$Z_T=(i\omega)^{-1}4\pi a\mu(1-iha). \tag{7c}$$

The second result is that if the sphere is allowed to slip relative to the elastic medium then the general form of Eq. (6) is preserved with Z_T modified. Specifically, suppose the tangential component of the traction satisfies

$$\boldsymbol{T}\cdot\hat{\boldsymbol{t}}=z_I(\dot{\boldsymbol{u}}^{\text{sphere}}-\boldsymbol{v})\cdot\hat{\boldsymbol{t}},\quad r=a, \tag{8}$$

where $\hat{\boldsymbol{t}}$ is a unit tangent vector, $\boldsymbol{v}$ the velocity of the elastic medium adjacent to the sphere, and z_I is an interfacial impedance, discussed later. (Upper case Z and lower case z are used to distinguish impedances defined by force and stress, respectively.) Equation (8) holds at each point on the interface $r=a$. We find that Z now satisfies

$$\frac{3}{Z+Z_m}=\frac{1}{Z_L+Z_m}+\frac{2}{Z_S+Z_m}, \tag{9}$$

where the new impedance Z_S is given by

$$\frac{1}{Z_S}=\frac{1}{Z_T}+\frac{1}{4\pi a^2 z_I+(i\omega)^{-1}8\pi a\mu}. \tag{10}$$

These results are derived in the next section and discussed in Sec. IV.

III. ANALYSIS

We use Oestreicher's[4] representation for the elastic field outside the sphere,

$$\boldsymbol{u}=-A_1\,\text{grad}\left(\frac{h_1(kr)}{kr}x\right)+B_1\left[2h_0(hr)\text{grad}\,x-h_2(hr)r^3\,\text{grad}\,\frac{x}{r^3}\right],\quad r\geqslant a, \tag{11}$$

where $r=|\boldsymbol{r}|$ is the spherical radius and x is the component of $\boldsymbol{r}$ in the $\hat{\boldsymbol{x}}$-direction, both with origin at the center of the sphere. Also, h_n are spherical Hankel functions of the first kind.[10] Let $\hat{\boldsymbol{r}}=r^{-1}\boldsymbol{r}$ denote the unit radial vector, then

$$\boldsymbol{u}=-A_1\left[\frac{h_1(kr)}{kr}\hat{\boldsymbol{x}}-h_2(kr)\frac{x}{r}\hat{\boldsymbol{r}}\right]+B_1\left[2h_0(hr)\hat{\boldsymbol{x}}-h_2(hr)\left(\hat{\boldsymbol{x}}-3\frac{x}{r}\hat{\boldsymbol{r}}\right)\right]. \tag{12}$$

The surface traction is $\boldsymbol{T}=\boldsymbol{\sigma}\hat{\boldsymbol{r}}$, where $\boldsymbol{\sigma}$ is the stress tensor. The traction can be calculated from (12) and the following identity[4] for an isotropic solid,

$$\boldsymbol{T}=\hat{\boldsymbol{r}}\lambda\,\text{div}\,\boldsymbol{u}+\frac{\mu}{r}\,\text{grad}\,\boldsymbol{r}\cdot\boldsymbol{u}+\mu\left(\frac{\partial}{\partial r}-\frac{1}{r}\right)\boldsymbol{u}. \tag{13}$$

Thus, referring to (2), we have

$$\boldsymbol{T}\cdot\hat{\boldsymbol{x}}=\left[2\mu h_2(kr)\left(1-3\frac{x^2}{r^2}\right)+(\lambda+2\mu)krh_1(kr)\frac{x^2}{r^2}\right]\frac{A_1}{r}+\left[2h_2(hr)\left(1-3\frac{x^2}{r^2}\right)-hrh_1(hr)\left(1-\frac{x^2}{r^2}\right)\right]3\mu\frac{B_1}{r}. \tag{14}$$

Integrating over the sphere surface, the resultant is

$$F=\frac{4}{3}\pi a^3\rho\omega^2\left[A_1\frac{h_1(ka)}{ka}-6B_1\frac{h_1(ha)}{ha}\right]. \tag{15}$$

The coefficients A_1 and B_1 follow from the conditions describing the interaction of the sphere with its surroundings. These are the general slip condition (8) plus the requirement that the normal velocity is continuous. The conditions at the surface of the sphere are

$$\left.\begin{aligned}\boldsymbol{u}\cdot\hat{\boldsymbol{r}}&=u_0\hat{\boldsymbol{x}}\cdot\hat{\boldsymbol{r}}\\ \boldsymbol{T}\cdot\hat{\boldsymbol{t}}&=i\omega z_I(\boldsymbol{u}-u_0\hat{\boldsymbol{x}})\cdot\hat{\boldsymbol{t}}\end{aligned}\right\}\quad r=a. \tag{16}$$

By symmetry, the only nonzero tangential component is in the plane of $\hat{\boldsymbol{r}}$ and $\hat{\boldsymbol{x}}$, and we therefore set $\hat{\boldsymbol{t}}=\hat{\boldsymbol{\theta}}\equiv(\hat{\boldsymbol{r}}\cos\theta-\hat{\boldsymbol{x}})/\sin\theta$, where $\theta=\arccos\hat{\boldsymbol{r}}\cdot\hat{\boldsymbol{x}}$ is the spherical polar angle. Using polar coordinates, $\boldsymbol{u}=u_r\hat{\boldsymbol{r}}+u_\theta\hat{\boldsymbol{\theta}}$ and $\boldsymbol{T}=\sigma_{rr}\hat{\boldsymbol{r}}+\sigma_{r\theta}\hat{\boldsymbol{\theta}}$, and (16) becomes

$$\left.\begin{aligned}u_r&=u_0\cos\theta\\ \sigma_{r\theta}-i\omega z_I u_\theta&=i\omega z_I u_o\sin\theta\end{aligned}\right\}\quad r=a,\quad 0\leqslant\theta\leqslant\pi. \tag{17}$$

The shear stress follows from the identity

$$\sigma_{r\theta}=\mu\left(\frac{\partial u_\theta}{\partial r}+\frac{1}{r}\frac{\partial u_r}{\partial\theta}-\frac{u_\theta}{r}\right), \tag{18}$$

and the interface conditions (17) then imply, respectively,

$$\left[h_2(ka)-\frac{h_1(ka)}{ka}\right]A_1+6\frac{h_1(ha)}{ha}B_1=u_0. \tag{19a}$$

$$-\left[\frac{h_1(ka)}{ka}+\frac{2\mu h_2(ka)}{i\omega a z_I}\right]A_1+\left[\left(2+\frac{\mu h^2a^2}{i\omega a z_I}\right)\frac{h_1(ha)}{ha}-\left(1+\frac{2\mu}{i\omega a z_I}\right)h_2(ha)\right]3B_1=u_0. \tag{19b}$$

Solving for A_1 and B_1, then substituting them into Eqs. (15)

and (3), and using the known forms for the spherical Hankel functions, yields

$$Z = -Z_m + 3/[1/(Z_L + Z_m) + 2/(Z_S + Z_m)]. \tag{20}$$

Equation (20) is identical to (9), which completes the derivation of the generalized impedance formula.

IV. DISCUSSION

It is useful to recall some basic properties of lumped parameter impedances. The impedance of a spring mass damper system of stiffness K, mass M, and damping C is

$$Z = (i\omega)^{-1}K - C + i\omega M. \tag{21}$$

Two impedances Z_1 and Z_2 combined in series have an effective impedance $(Z_1^{-1}+Z_2^{-1})^{-1}$, while the result for the same pair in parallel is (Z_1+Z_2).

Referring to the definitions of Eq. (7), it is clear that Z_m is the impedance of the mass of the volume removed from the elastic medium. The impedance of a longitudinal or transverse plane wave is defined as the ratio of the stress (normal or shear) to particle velocity, and equals z_L, z_T, where

$$z_L = \rho c_L, \quad z_T = \rho c_T. \tag{22}$$

Thus, both Z_L and Z_T have the form

$$Z = \left(\frac{1}{i\kappa a} - 1\right) 4\pi a^2 z, \tag{23}$$

where κ is the wave number (k or h). In particular, the impedances Z_L and Z_T have stiffness and damping, but no mass contribution. The damping can be ascribed to the radiation of longitudinal and transverse waves from the sphere.

The impedance Z_S of Eq. (10) corresponds to Z_T in series with an impedance Z_I, where

$$Z_I = 4\pi a^2 z_I + \frac{8\pi a^2 z_T}{iha}. \tag{24}$$

Thus, Z_I can be interpreted as the total interfacial impedance for the surface area of the sphere in parallel with twice the stiffness part of Z_T.

The limit of a purely acoustic fluid is obtained by letting the shear modulus μ tend to zero with λ finite, while an incompressible elastic or viscous medium is obtained in the limit as the bulk modulus $\lambda + \frac{2}{3}\mu$ becomes infinite with μ finite. The acoustic and incompressible limits follow from (20) as

$$Z = \begin{cases} \left(\dfrac{2}{Z_m} + \dfrac{3}{Z_L}\right)^{-1}, & \text{acoustic medium,} \\ \dfrac{1}{2}Z_m + \dfrac{3}{2}Z_S, & \text{incompressible medium.} \end{cases} \tag{25}$$

Thus, Z for the acoustic fluid comprises $\frac{1}{2}Z_m$ in series with $\frac{1}{3}Z_L$. Note that, as expected, the interfacial impedance z_I is redundant in the acoustic limit. The impedance for the incompressible medium is $\frac{1}{2}Z_m$ and $\frac{3}{2}Z_S$ in parallel, and it depends upon the interfacial impedance.

In order to examine the role of z_I, we first express the impedance Z of Eq. (20) in a form similar to (5),

$$Z = \frac{4}{3}\pi a^3 \rho i\omega \left\{ -1 + \left[\frac{1}{3}\left(1 - \frac{3(1-ika)}{k^2a^2}\right)^{-1} + \frac{2}{3}\left(1 - \frac{3(1-iha)}{h^2a^2\left[1 + \frac{\chi}{2}(1-iha)\right]}\right)^{-1}\right]^{-1}\right\}, \tag{26}$$

where the influence of the interfacial impedance is represented through the nondimensional parameter

$$\chi = \left(1 + \frac{i\omega a z_I}{2\mu}\right)^{-1}. \tag{27}$$

The form of χ is chosen so that it takes on the values zero or unity in the limit that the sphere is perfectly bonded or is perfectly lubricated,

$$\chi = \begin{cases} 0, & \text{no slip,} \quad z_I \to \infty, \\ 1, & \text{slip,} \quad z_I = 0. \end{cases} \tag{28}$$

The acoustic and incompressible limits of (25) are explicitly

$$Z = \begin{cases} \dfrac{4}{3}\pi a^3 \rho i\omega \dfrac{(1-ika)}{2(1-ika) - k^2a^2}, & \text{acoustic,} \\ \dfrac{6\pi a\mu}{i\omega}\left[\dfrac{1-iha}{1 + \frac{\chi}{2}(1-iha)} - \dfrac{h^2a^2}{9}\right], & \text{incompressible.} \end{cases} \tag{29}$$

Oestreicher[4] showed that the original formula (4) provides the acoustic and incompressible limits for perfect bonding (χ=0). Ilinskii *et al.*[11] derived the impedance in the context of incompressible elasticity, also for the case of no slip.

The behavior of Z at low and high frequencies depends upon how z_I and hence χ behaves in these limits. For simplicity, let us consider χ as constant in each limit, equal to χ_0 at low frequency, and χ_∞ at high frequency. Then,

$$Z = \begin{cases} (i\omega)^{-1}\dfrac{12\pi a\mu}{2 + \chi_0 + c_T^2/c_L^2}\left[1 - iha\left(\dfrac{2 + c_T^3/c_L^3}{2 + \chi_0 + c_T^2/c_L^2}\right) + O(h^2a^2)\right], & |ha|,|ka| \ll 1, \\ \dfrac{4}{3}\pi a^2 \rho c_L\left[-\left(1 + 2\dfrac{c_T}{c_L}(1-\chi_\infty)\right) + \dfrac{1}{ika}\left(1 - 4\dfrac{c_T}{c_L}\left[1 + \left(\dfrac{c_T}{c_L} - 1\right)\chi_\infty\right]\right) + O\left(\dfrac{1}{k^2a^2}\right)\right], & |ha|,|ka| \gg 1. \end{cases} \tag{30}$$

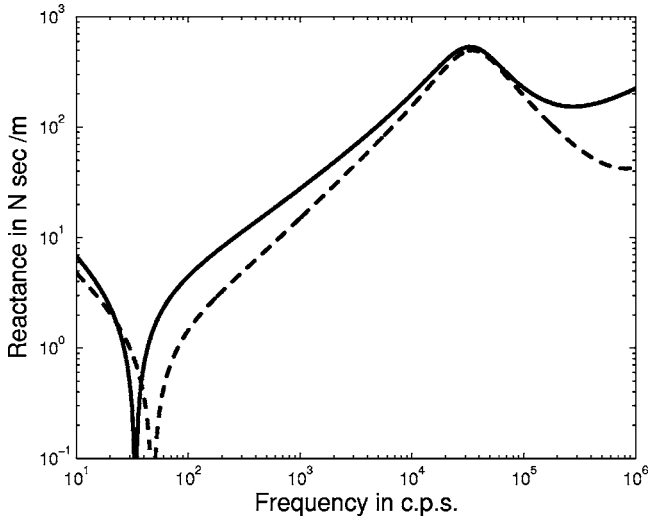


FIG. 1. The reactance (real part of Z) for an oscillating sphere in a tissuelike material (Ref. 4). The solid and dashed curves correspond to a bonded ($\chi=0$) and slipping ($\chi=1$) spherical interface, respectively. The reactance is positive (masslike) except for frequency below 30 c.p.s. (50 c.p.s. for the dashed curve) where it is negative (stiffnesslike).

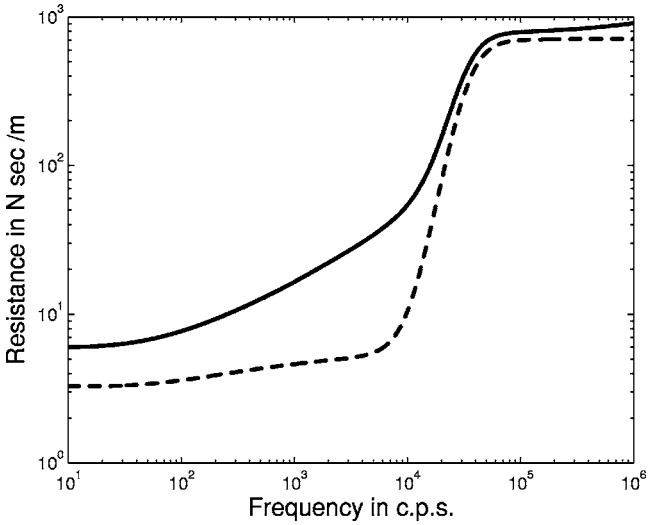


FIG. 2. The resistance (imaginary part of $-Z$) for an oscillating sphere in a tissuelike material (Ref. 4). The solid and dashed curves correspond to a bonded ($\chi=0$) and slipping ($\chi=1$) spherical interface.

The leading order term at high frequency is a damping, associated with radiation from the sphere. The dominant effect at low frequency is, as one might expect, a stiffness, with the second term a damping. The low frequency stiffness is identical to that previously determined by Lin *et al.*[9] who considered the static problem of a sphere in an elastic medium with an applied force. They derived the resulting displacement, and hence stiffness, under slip and no slip conditions. In order to compare with their results, we rewrite the leading order term as

$$Z=(i\omega)^{-1}\frac{24\pi a\mu(1-\nu)}{5-6\nu+2(1-\nu)\chi_0}[1+o(1)], \tag{31}$$

where ν is the Poisson's ratio,

$$\nu=\frac{\frac{1}{2}c_L^2-c_T^2}{c_L^2-c_T^2}. \tag{32}$$

Equation (31) with $\chi_0=0$ and $\chi_0=1$ agrees with Eqs. (40) and (41) of Lin *et al.*,[9] respectively. In an incompressible viscous medium with $\nu\approx 1/2$ and $\mu=-i\omega\mu_2$, (31) becomes

$$Z\approx-\frac{6\pi a\mu_2}{1+\frac{1}{2}\chi_0}, \tag{33}$$

which reduces to the Stokes[12] drag formula $F=-6\pi a\mu_2 v$ for perfect bonding. When there is slip ($\chi_0=1$) the drag is reduced by one third, $F=-4\pi a\mu_2 v$. It is interesting to note that one third of the contribution to the drag in Stokes' formula is from pressure, $2\pi a\mu_2 v$, the remained from shear acting on the sphere. However, under slip conditions, the shear force is absent and the total drag $4\pi a\mu_2 v$ is caused by the pressure.

The simplest example of the interfacial impedance is a constant value, which is necessarily negative and corresponds to a damping, $z_I=-C$. For an elastic medium we have

$$\chi=\frac{1}{1-i\omega/\omega_c},\quad \omega_c=\frac{2\mu}{aC},\quad \text{elastic medium},\quad z_I=-C. \tag{34}$$

Hence, $\chi_0=1$ and $\chi_\infty=0$, corresponding to slip at low frequency and no slip at high frequency. The transition from the low to high frequency regime occurs for frequencies in the range of a characteristic frequency ω_c. Alternatively, if the medium is purely viscous $\mu=-i\omega\mu_2$, again with constant z_I, the parameter χ becomes

$$\chi=\frac{1}{1+\frac{aC}{2\mu_2}},\quad \text{viscous medium},\quad \mu=-i\omega\mu_2,\quad z_I=-C. \tag{35}$$

In this case χ is constant with a value between 0 and 1 that depends upon the ratio of the interfacial to bulk viscous damping coefficients, and also upon a. One can define a characteristic particle size $a_c=\mu_2/C$, such that spheres of radius $a\ll a_C$ ($a\gg a_C$) are effectively bonded (lubricated).

Figures 1 and 2 show the reactance and resistance of a sphere of radius 0.01 m in a medium with the parameters considered by Oestreicher[4] based on measurements of human tissue, $\rho=1100\ \text{kg/m}^3$, $\mu_1=2.5\times10^3$ Pa, $\mu_2=15$ Pa s, $\lambda=2.6\times10^9$ Pa. The perfectly bonded ($\chi=0$) and perfect slip ($\chi=1$) conditions are compared. Figure 1 indicates that the masslike reactance is generally reduced by the slipping, and it also shows that the low frequency stiffness is two-thirds that of the bonded case, Eq. (31). Interfacial slip leads to a significant decrease in the resistance, as evident from Fig. 2 which shows a reduction for all frequencies.

[1]Y. L. Yin, S. F. Ling, and Y. Liu, "A dynamic indentation method for characterizing soft incompressible viscoelastic materials," Mater. Sci. Eng., A **379**, 334–340 (2004).
[2]S. Chen, M. Fatemi, and J. F. Greenleaf, "Remote measurement of material properties from radiation force induced vibration of an embedded sphere," J. Acoust. Soc. Am. **112**, 884–889 (2002).
[3]M. Fatemi, L. E. Wold, A. Alizad, and J. F. Greenleaf, "Vibro-acoustic

tissue mammography," IEEE Trans. Med. Imaging **21**, 1–8 (2002).
[4]H. L. Oestreicher, "Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics," J. Acoust. Soc. Am. **23**, 707–714 (1951).
[5]H. E. von Gierke, H. L. Oestreicher, E. K. Franke, H. O. Parrack, and W. W. von Wittern, "Physics of vibrations in living tissues," J. Appl. Physiol. **4**, 886–900 (1952).
[6]X. Zhang, T. J. Royston, H. A. Mansy, and R. H. Sandler, "Radiation impedance of a finite circular piston on a viscoelastic half-space with application to medical diagnosis," J. Acoust. Soc. Am. **109**, 795–802 (2002).
[7]S. Chen, G. T. Silva, R. R. Kinnick, M. Fatemi, and J. F. Greenleaf, "Measurement of dynamic and static radiation force on a sphere," Phys. Rev. E **71**, 056618(4) (2005).
[8]D. C. Lin, B. Yurke, and N. A. Langrana, "Mechanical properties of a reversible, DNA-crosslinked polyacrylamide hydrogel," J. Biomech. Eng. **126**, 104–110 (2004).
[9]D. C. Lin, N. A. Langrana, and B. Yurke, "Force-displacement relationships for spherical inclusions in finite elastic media," J. Appl. Phys. **97**, 043510(4) (2005).
[10]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York, 1974).
[11]Y. A. Ilinskii, G. D. Meegan, E. A. Zabolotskaya, and S. Y. Emelianov, "Gas bubble and solid sphere motion in elastic media in response to acoustic radiation force," J. Acoust. Soc. Am. **117**, 2338–2346 (2005).
[12]G. G. Stokes, "On the effect of the internal friction of fluids on the motion of pendulums," Proc. Cambridge Philos. Soc. **9**, 8–106 (1851).

Bubble pulsations between parallel plates

Jianying Cui, Mark F. Hamilton, Preston S. Wilson, and Evgenia A. Zabolotskaya
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029, and Department of Mechanical Engineering, The University of Texas at Austin, 1 University Station C2200, Austin, Texas 78712-0292

(Received 1 April 2005; revised 4 September 2005; accepted 14 January 2006)

The dynamic response of an acoustically driven spherical bubble between parallel plates is investigated in the linear approximation. For the case of rigid plates, explicit expressions are provided for the resonance frequency and damping, and the importance of including compressibility of the liquid is discussed. For wide channels, approximate results are presented that account for finite acoustic impedance of the plates. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2172545]

PACS number(s): 43.25.Yw [AJS] Pages: 2067–2072

I. INTRODUCTION

The use of microbubbles as contrast agents in medical ultrasound imaging, their development for drug and gene delivery, and their role in shock wave lithotripsy all involve bubbles in constrained media. Such applications have led to recent interest in modeling bubble dynamics in blood vessels[1–3] and other narrow biological channels.[4] Our work is motivated by these same applications and is intended to provide solutions for a simple geometry that are free of some limitations of earlier work on bubble dynamics in channels, and to provide benchmarks for more advanced analyses.

The analyses of Sassaroli and Hynynen[1,2] are based on theoretical models developed by Prosperetti and co-workers.[5,6] In one, Sassaroli and Hynynen[1] use the model of Chen and Prosperetti[5] for a bubble that occupies the entire cross section of a tube. The columns of liquid on either side of the bubble have finite length and are modeled as lumped, finite masses. In their other work[2] they use the model of Oğuz and Prosperetti[6] for a bubble in a tube that has at least one end open, either connected with an unbounded reservoir or exposed to the atmosphere. The finite effective tube lengths in these cases ensure that the mechanical impedance at the bubble wall due to inertia of the liquid is finite, even though the liquid in the channel is assumed to be incompressible.

Krasovitski and Kimmel[4] and Hu *et al.*[3] analyzed bubble dynamics in a tube of infinite length, with a focus on deformation of the bubble shape. Krasovitski and Kimmel did so numerically using a boundary element approach, while Hu *et al.* used a perturbation approach. However, as pointed out by Oğuz and Prosperetti[6] in connection with modeling bubble dynamics in rigid tubes of finite length, "if the length of the tube were infinite, volume changes of the bubble would only be possible in a compressible fluid." Our analysis supports this statement.

Using a Green's function approach to derive the acoustic radiation impedance of a pulsating sphere, and thus accounting for compressibility of the liquid, Leighton *et al.*[7] developed an approximate expression for the resonance frequency shift of a bubble in an infinite cylindrical pipe. Calculations are presented for a pipe with diameter so large, more than 100 times the bubble radius, that the shift in resonance frequency due to constraint of the flow by the pipe is negligible. In contrast, our emphasis is on relatively narrow channels, with wall separations of order 10 times the bubble radius, for which the corresponding resonance frequency shift can be substantial.

A subsequent paper by Leighton *et al.*[8] extends the earlier analysis of a bubble in a cylindrical pipe[7] to a bubble in a rectangular tank. Using a Green's function solution and the method of images, they obtained a result for the acoustic radiation impedance of the bubble, and thus for the resonance frequency and damping factor. While the solution developed below for the resonance frequency may be obtained from their theoretical framework, their focus is on radiation damping of a bubble in a large tank, and as such it is entirely different from our application. Their approach is also appropriate for studying bubble dynamics in microfluidic devices in which a liquid is constrained within a sealed cavity.

In the present paper we analyze the solution for an acoustically driven spherical bubble that pulsates at small amplitude in a channel formed by rigid parallel plates. Explicit expressions are provided for the resonance frequency and quality factor as functions of channel width. For sufficiently wide channels, an approximate method is described for including effects of finite channel wall impedance.

The decision to study a channel formed by rigid parallel plates is motivated by the opportunity to use the method of images to satisfy the boundary conditions and ultimately obtain a closed-form solution. The problem is thus equivalent to calculating the response to acoustic excitation of an infinite line array of equally spaced bubbles in unbounded liquid. An analogous problem has been the subject of investigations in underwater acoustics.

Weston[9] was the first to derive a mathematical model for acoustic scattering from an infinite line array of bubbles that accounts for multiple scattering between the bubbles. In other words, the bubbles were considered to be dynamically coupled. His analysis is restricted to bubble separations that are small on the scale of a wavelength, and therefore free-field results are not recovered for large bubble separations.

An expression for the resonance frequency corresponding to the scattering is presented, but it is different from the result derived below for the bubble resonance.

A mathematical model of acoustic scattering from an infinite line array of bubbles, unrestricted to small bubble separations, was developed by Tolstoy and Tolstoy.[10] Their results and discussion describe scattering amplitude as a function of the propagation direction of a plane sound wave incident on the line array. They do not investigate bubble resonance frequency or damping. In general, while the papers by Weston and subsequently Tolstoy and Tolstoy provide models that are relevant to our problem at hand, their focus on target strength entails theoretical formulations that are not well suited to the study of bubble dynamics.

Feuillade[11] performed an exhaustive analysis of resonances and damping associated with pairs and triplets of acoustically excited bubbles. Both in-phase and antiphase pulsations are considered. The results for in-phase excitation display the trends shown below for a bubble between parallel plates. While the paper concludes with the solution for excitation of an infinite line array, a discussion of this case is brief and focused on acoustic scattering. The solution is used to illustrate the limiting target strength as the number of bubbles in a line array is increased. Solutions are not presented for the resonance frequency or damping.

We begin by developing the linear dynamical equation for a spherical bubble between rigid parallel plates. This development is presented to clarify the ordering of terms and to illustrate the essential role of compressibility of the liquid. Expressions for the amplitude response, resonance frequency, and damping factor are obtained from the solution presented by Feuillade[11] for an infinite line array of bubbles. An approximate method is described for including effects of finite plate impedance for wide channels.

II. DYNAMICAL EQUATION

We consider an acoustically driven spherical bubble of equilibrium radius R_0 located midway between two parallel rigid plates separated by distance d, as depicted by the shaded circle in Fig. 1. The effects of viscosity, heat conduction, and surface tension are not considered here, as attention is focused on the influence of compressibility of the liquid. Frequencies under consideration are presumed to be such that $\lambda \gg R_0$, where λ is the characteristic wavelength of sound in the liquid.

The gas inside the bubble is assumed to behave adiabatically, such that its pressure is $P_g = P_0(R_0/R)^{3\gamma}$, where P_0 is atmospheric pressure, $R(t)$ the instantaneous bubble radius, and γ the ratio of specific heats. In the linear approximation this becomes, with $R(t) = R_0 + \xi(t)$,

$$P_g(t) = P_0 - (3\gamma P_0/R_0)\xi(t). \tag{1}$$

The gas pressure P_g must equal the pressure P_l in the liquid at the bubble wall. The latter is expressed as

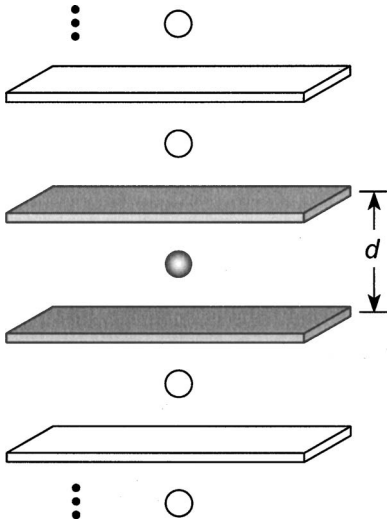


FIG. 1. Geometry of a bubble located midway between parallel plates. Mirror images (open circles) separated by distance d are shown extending indefinitely in both directions.

$$P_l(t) = P_0 + p_{ext}(t) + p_{\mathrm{rad}}(t), \tag{2}$$

where p_{ext} is the externally applied acoustic pressure, and p_{rad} is the acoustic pressure due to radiation from the bubble.

There are two contributions to p_{rad}. One is the reaction to the acoustic wave as it radiates away from the bubble wall, and the other is the sound incident on the bubble due to reflections from the plates. To identify the appropriate form of p_{rad} it is easiest to consider harmonic excitation of the bubble and express the bubble wall velocity as $\dot{\xi}(t) = u_0 e^{j\omega t}$, where the overdot indicates a time derivative. Employing the solution for acoustic radiation from a pulsating spherical source,[12] we write

$$p_{\mathrm{rad}}(t) = \frac{jkR_0}{1 + jkR_0}\rho_0 c_0 u_0 e^{j\omega t} + 2\frac{jkR_0}{1 + jkR_0}\sum_{n=1}^{\infty}\frac{R_0}{nd}\rho_0 c_0 u_0 e^{j[\omega t - k(nd - R_0)]}, \tag{3}$$

where ρ_0 and c_0 are the equilibrium density and sound speed of the liquid, respectively, and $k = \omega/c_0$ is the acoustic wave number. The first term in Eq. (3) is the pressure at the bubble wall in an unbounded liquid. The summation accounts for sound reflected from the plates. It represents a doubly infinite sequence of image bubbles (the open circles in Fig. 1) separated from one another by distance d, extending in opposite directions, that satisfies the condition of zero normal fluid velocity on the plates. The field produced by the images accounts for all reflected sound at the point occupied by the center of the bubble.

Following division by $u_0 e^{j\omega t}$, the right-hand side of Eq. (3) is seen to be the specific acoustic radiation impedance for the bubble. The expansion of Eq. (3) in powers of the small parameter jkR_0 yields

$$p_{\mathrm{rad}}(t) = jkR_0\rho_0 c_0 u_0 e^{j\omega t} - (jkR_0)^2 \rho_0 c_0 u_0 e^{j\omega t} + 2jkR_0 \sum_{n=1}^{\infty} \frac{R_0}{nd} \rho_0 c_0 u_0 e^{j\omega(t-nd/c_0)}, \quad kR_0 \ll 1. \quad (4)$$

Suppressing the time dependence $e^{j\omega t}$ and using the relation $e^{-jnkd} = \cos nkd - j \sin nkd$ in the summation, we observe that the first term and the imaginary part of the third term account for mass loading at the bubble wall, while the second term and the real part of the third term account for radiation damping. The expansion thus includes the leading-order terms accounting for mass loading and radiation damping due to both the radiated and reflected sound fields. Reinstating $\dot{\xi}(t)$ for $u_0 e^{j\omega t}$ in Eq. (4), associating factors of $jkR_0 = j\omega R_0/c_0$ with time derivatives, and then setting $P_g = P_l$ by combining Eqs. (1), (2), and (4) yields

$$\ddot{\xi}(t) + \omega_0^2 \xi(t) = \frac{R_0}{c_0} \dddot{\xi}(t) - 2\frac{R_0}{d} \sum_{n=1}^{\infty} \frac{1}{n} \ddot{\xi}(t - nd/c_0) - \frac{p_{\mathrm{ext}}(t)}{\rho_0 R_0}, \quad (5)$$

where $\omega_0^2 = 3\gamma P_0/\rho_0 R_0^2$. This is the desired model equation for acoustically driven, infinitesimal pulsations of a bubble located midway between rigid parallel plates.

Equation (5) becomes similar to Eqs. (2) of Doinikov *et al.*[13] when the latter are applied to the geometry depicted in Fig. 1 and the external sound field is taken into account. It may also be obtained from Eqs. (7) of Feuillade[11] under the same conditions. Doinikov *et al.* used their model to investigate the effect of time delays, corresponding to the quantities nd/c_0 in Eq. (5), on the free oscillations of linear arrays containing up to 16 equispaced bubbles. Calculations with and without the time delays taken into account were compared to determine the influence on the natural modes of the arrays.

Compressibility of the liquid is responsible for the time delays, because in an incompressible liquid $c_0 = \infty$ and the time delays vanish. Ignoring the effect of compressibility in Eq. (5) does not merely alter the dynamic response; it eliminates it altogether. For infinite sound speed, Eq. (5) reduces to

$$\left(1 + 2\frac{R_0}{d} \sum_{n=1}^{\infty} \frac{1}{n}\right) \ddot{\xi}(t) + \omega_0^2 \xi(t) = -\frac{p_{\mathrm{ext}}(t)}{\rho_0 R_0}, \quad c_0 \to \infty. \quad (6)$$

The expression multiplying $\ddot{\xi}$ is proportional to the effective inertia of the liquid in contact with the bubble wall. Since the summation of $1/n$ diverges, the effective inertia is infinite, and there is no dynamic solution of Eq. (6). The infinite mechanical impedance at the bubble wall is produced by pressure waves arriving simultaneously, and therefore in phase, from the infinity of reflections between the plates (or from scattering by the infinite sequence of bubbles, when placed in the context of a line array).

In the context of incompressible flow in a rigid channel, a physical explanation of the aforementioned infinite impedance is readily appreciated on the basis of energy considerations. The total kinetic energy in an incompressible liquid surrounding a spherical bubble whose wall moves radially with velocity $\dot{\xi}(t)$ is $2\pi\rho_0 R_0^3 \dot{\xi}^2(t)$, a finite quantity.[14] Now consider the same bubble in the channel depicted in Fig. 1. Far from the bubble, where the streamlines are parallel to the walls and the flow is uniform across the channel, the velocity of the liquid at distance r from the bubble is proportional to $\dot{\xi}(t)/r$, such that the kinetic energy per unit volume is proportional to $1/r^2$ (hereafter neglecting time dependence). The kinetic energy in a thin cylindrical shell of volume $d \times 2\pi r\, dr$ is thus proportional to dr/r, the integral of which between inner radius r_i and outer radius r_o yields $\ln(r_o/r_i)$ for a shell of finite thickness. The total kinetic energy in a liquid constrained by rigid plates of infinite extent, $r_o = \infty$, is therefore infinite. Unlike spherical divergence of the flow in an unconstrained liquid, cylindrical divergence is insufficient for the kinetic energy to be finite. The bubble must perform infinite work to change its volume in the channel, however small that change may be, and therefore radial motion of the bubble wall is prohibited. The same argument applies to a bubble in an infinite rigid tube with constant cross section, for which there is no divergence of the flow far from the bubble, and the kinetic energy increases linearly with distance, rather than logrithmically. Equation (5) is free of this shortcoming associated with incompressible flow.

We emphasize that the essential role of compressibility discussed in the previous paragraph is a consequence of assuming the channel has infinite length and rigid walls. Relaxing either of these assumptions permits dynamic solutions to be obtained without accounting for compressibility. As discussed in the Introduction, models of bubble dynamics in incompressible liquid constrained by a rigid but finite tube have been developed by Oğuz and Prosperetti.[6] An approximate approach to modeling bubble dynamics in incompressible liquid constrained by parallel plates with walls having finite impedance is described in Sec. IV. Of course, for tubes of finite length but that are sufficiently long, or having walls with finite but not sufficiently large impedance, compressibility of the liquid competes with and eventually dominates these other mechanisms for reducing the effective inertia of the flow.

III. HARMONIC EXCITATION

For harmonic excitation by an externally applied sound pressure $p_{\mathrm{ext}}(t) = p_0 e^{j\omega t}$, the response takes the form $\xi(t) = \Xi(\omega) e^{j\omega t}$, and substitution in Eq. (5) yields

$$\frac{\Xi(\omega)}{\Xi(0)} = \left[1 - \left(1 + 2\frac{R_0}{d} \sum_{n=1}^{\infty} \frac{e^{-jnkd}}{n}\right) \frac{\omega^2}{\omega_0^2} + jk_0 R_0 \frac{\omega^3}{\omega_0^3}\right]^{-1}, \quad (7)$$

where $\Xi(0) = -(p_0/3\gamma P_0) R_0$ is the response at zero frequency, and $k_0 = \omega_0/c_0$. This solution is equivalent to Eq. (22) in Feuillade's paper[11] after bubble radius and volume are related in the linear approximation. Evaluation and interpretation of this solution are facilitated by the following two exact relations for the summation:[15]

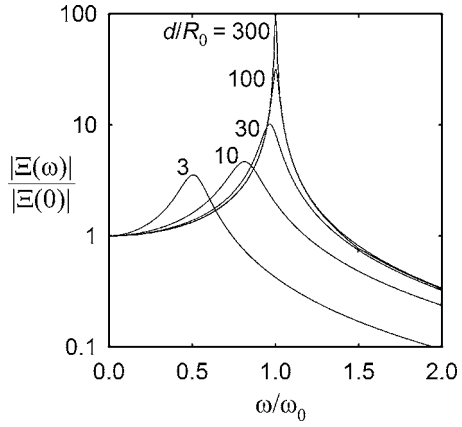


FIG. 2. Amplitude response versus acoustic excitation frequency for different plate separations.

$$\sum_{n=1}^{\infty} \frac{e^{-jnkd}}{n} = -\ln(1-e^{-jkd}), \quad \text{for all } kd, \tag{8}$$

$$= -\ln\left(2\sin\frac{kd}{2}\right) - \frac{j}{2}(\pi - kd), \quad 0 < kd < 2\pi. \tag{9}$$

Equation (9) is the relation employed by Tolstoy and Tolstoy.[10]

Use of Eq. (8) yields

$$\frac{\Xi(\omega)}{\Xi(0)} = \left[1 - \left(1 - 2\frac{R_0}{d}\ln(1-e^{-jkd})\right)\frac{\omega^2}{\omega_0^2} + jk_0R_0\frac{\omega^3}{\omega_0^3}\right]^{-1}. \tag{10}$$

Dependence of the magnitude on d/R_0 is shown in Fig. 2 for an air bubble in water, for which $k_0R_0=0.014$. It is seen that as the plate separation is reduced, the resonance frequency decreases from its value of ω_0 in a free field, and the damping increases, as observed by others in connection with line arrays of bubbles.[9–11]

Explicit expressions for the resonance frequency and damping factor may be obtained using the identity given by Eq. (9). Although restricted to $0<kd<2\pi$, it permits the separation of the real and imaginary parts of the solution. This relation may be used for all values of kd by understanding that it represents just one cycle of a function that is periodic in 2π. When Eq. (9) is substituted in (7), the term $jkd/2$ in the former cancels the term $jk_0R_0\omega^3/\omega_0^3$ in the latter, and the result is

$$\frac{\Xi(\omega)}{\Xi(0)} = \left\{1 - \left[1 - 2\frac{R_0}{d}\ln\left(2\sin\frac{kd}{2}\right)\right]\frac{\omega^2}{\omega_0^2} + j\pi\frac{R_0}{d}\frac{\omega^2}{\omega_0^2}\right\}^{-1}, \quad 0 < kd < 2\pi. \tag{11}$$

The restriction $kd<2\pi$ is equivalent to $d<\lambda$. For $kd\to 0$ (incompressible liquid) and $kd\to 2\pi$ (for which arrivals from all images are in phase with radiation from the bubble), it is seen that $\Xi(\omega)\to 0$, corresponding to the infinite effective inertia discussed in connection with Eq. (6).

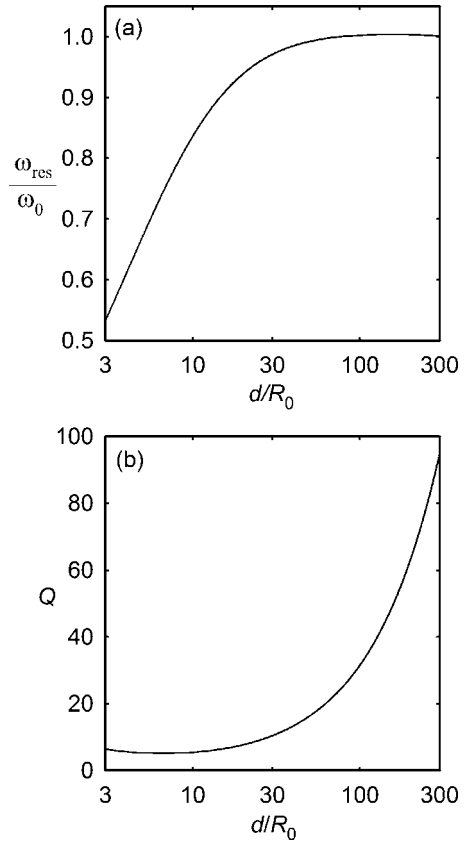


FIG. 3. (a) Resonance frequency and (b) quality factor as functions of plate separation for an air bubble in water.

The resonance frequency ω_r is defined by setting the real part of the expression within the braces in Eq. (11) to zero, which yields the following transcendental relation:

$$\omega_r = \frac{\omega_0}{\sqrt{1 - 2(R_0/d)\ln|2\sin(\omega_r d/2c_0)|}}. \tag{12}$$

While Eq. (11) is restricted to $kd<2\pi$, Eq. (12) has been rendered valid for all kd by taking the absolute value of the sine function to account for the periodicity of Eqs. (8) and (9). Equation (12) also follows from the expression for the resonance frequency in a tank given by Eq. (42) of Leighton *et al.*,[8] when the latter is evaluated for the geometry in Fig. 1 together with rigid boundary conditions, and the relation in Eq. (9) is used. [We also observe that the subscript 0 was omitted from the wave number k in their Eq. (42), corresponding to omitting the subscript r on the right side of our Eq. (12).] Leighton *et al.* do not present solutions of their Eq. (42).

The solution of Eq. (12) is shown in Fig. 3(a), where the resonance frequency shift is seen to be significant for plate separations less than about 10 bubble diameters, $d/R_0 \lesssim 20$. Resonance frequency is decreased because the plates prohibit the ideal radial expansion of flow that occurs in an unbounded liquid, thus increasing the effective fluid inertia. Figure 3(a) is qualitatively the same as Fig. 3 of Feuillade[11] for arrays of two and three bubbles. Quantitatively, the resonance frequency shift depicted in Fig. 3(a) is about twice that which is predicted for three bubbles separated by distance d.

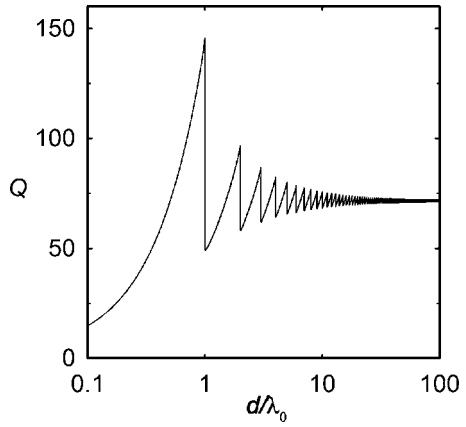


FIG. 4. Convergence of the quality factor to its value for a bubble in an unbounded liquid as plate separation is increased indefinitely.

The quality factor is defined by the relation $Q=|\Xi(\omega_r)/\Xi(0)|$, such that

$$Q=\frac{\omega_0^2}{\omega_r^2}\frac{d}{\pi R_0},\quad k_0 d<2\pi, \tag{13}$$

$$\simeq\frac{d}{\pi R_0},\quad \frac{d}{R_0}\gg 1,\quad k_0 d<2\pi, \tag{14}$$

$$=\frac{1}{k_0 R_0},\quad k_0 d\to\infty. \tag{15}$$

Equations (13) and (14) follow from (11), and Eq. (15) follows from (10). The dependence of Q on d/R_0 is shown in Fig. 3(b). Since losses due to viscosity and heat conduction are not included in the analysis, damping associated with the quality factor is due entirely to compressibility. The limiting value in Eq. (15) for a channel of infinite width, $Q\sim 1/k_0R_0$, corresponds to radiation damping in an unbounded fluid. The transition to this limiting value is illustrated in Fig. 4, where $\lambda_0=2\pi/k_0$, and the periodic structure is due to the sawtooth behavior connected with the term $j(\pi-kd)/2$ in Eq. (9). For comparison, the value $d/\lambda_0=1$ in Fig. 4 corresponds to $d/R_0=450$ in Fig. 3.

IV. FINITE WALL IMPEDANCE

We consider here the effect of channel walls having finite acoustic impedance, whether due to the transmission of sound outside the channel, or due to wall compliance or inertia. A solution of this problem for arbitrary wall impedance is difficult, but solutions under certain restricted conditions suggest the dynamic response of a bubble that may be expected under more general conditions.

In the limit of geometrical acoustics (high frequencies), waves radiated from the bubble and reflected from the channel walls back toward the bubble follow ray paths that are approximately perpendicular to the channel walls. The specific acoustic impedance of the wave is $\rho_0 c_0$ in this limit, and use of the plane-wave reflection coefficient for normal incidence on the channel wall is then appropriate. The necessary condition for these assumptions to be reasonable is that wave number times the radius of curvature of the wave front incident on the wall is large compared to unity. In our case this corresponds to $kd\gg 1$, or $d\gtrsim\lambda$.

Thus, for $kd\gg 1$, if we assume that the surface of the wall is locally reacting, finite wall impedance can be taken into account by introducing a pressure reflection coefficient β in the summation in Eq. (7) as follows:

$$\sum_{n=1}^{\infty}\frac{e^{-jnkd}}{n}\to\sum_{n=1}^{\infty}\beta^n\frac{e^{-jnkd}}{n},\quad kd\gg 1. \tag{16}$$

The quantity β is applied once for every reflection. Writing $\beta^n=e^{n\ln\beta}$, we can achieve the same result by replacing the real wave number $k=\omega/c_0$ in either Eq. (7) or (10) by the complex wave number,

$$\tilde{k}=\frac{\omega}{c_0}+\frac{j}{d}\ln\beta. \tag{17}$$

This substitution cannot be made directly Eq. (11); instead it must be made one step earlier, in Eq. (9), because of how terms were combined to obtain Eq. (11).

For example, in Eq. (10) one has

$$\ln(1-e^{-j\tilde{k}d})=\ln(1-|\beta|e^{-j(kd-\phi)}), \tag{18}$$

$$=\ln(1-|\beta|e^{j\phi}),\quad c_0\to\infty, \tag{19}$$

where the reflection coefficient has been expressed in polar form as $\beta=|\beta|e^{j\phi}$. For an arbitrary specific acoustic wall impedance z_w the reflection coefficient is

$$\beta=\frac{z_w-\rho_0 c_0}{z_w+\rho_0 c_0}. \tag{20}$$

The wall impedance may be expressed as $z_w=r_w+jx_w$, where the real part r_w accounts for the loss of acoustic energy in the channel due to transmission through the walls, and the imaginary part x_w is the reactance associated with inertia and compliance of the walls. Since $|\beta|<1$ for any finite value of r_w, and $\phi\neq 0$ for any finite value of x_w, the expression in Eq. (19) is bounded in either case. This is to say, the problem with the prediction of infinite inertia resulting from the assumption of an incompressible liquid in an infinite channel with rigid walls is avoided by allowing for finite wall impedance.

As indicated in Eq. (16), our analysis of finite wall impedance is restricted to channels having widths greater than about one wavelength. However, the implication of the underlying physics is clear. Change in bubble volume is possible in a constrained incompressible liquid only if the channel walls have finite impedance, allowing them to move in order to accommodate the liquid displaced by motion of the bubble wall.

We note that an alternative use of a complex wave number is to account for the attenuation of sound in the liquid.

V. BUBBLE OFF CENTER

When the bubble is off center, say an arbitrary distance b from one plate, simplicity is lost because of the asymmetric

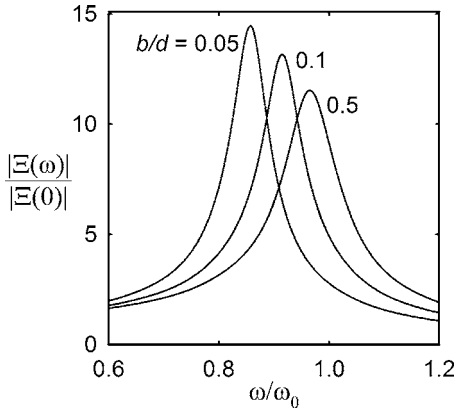


FIG. 5. Dependence of amplitude response on distance b of bubble from channel wall, for $d/R_0=30$.

distribution of the images required to satisfy the boundary conditions. The solution in this case is given by Eq. (7) after replacing the summation as follows:

$$\sum_{n=1}^{\infty}\frac{e^{-jnkd}}{n}\rightarrow\frac{1}{4}\sum_{n=1}^{\infty}\left(\frac{2}{n}+\frac{e^{j2kb}}{n-b/d}+\frac{e^{j2k(d-b)}}{n-1+b/d}\right)e^{-j2nkd}. \tag{21}$$

The two summations are equivalent for a bubble in the midplane, $b=d/2$. For a bubble off center, the summation no longer admits a closed form as in Eqs. (8) and (9). Dependence of the vibration amplitude on the position of the bubble in the channel is shown in Fig. 5 for $d/R_0=30$. Resonance frequency is decreased, and amplitude response is slightly increased, as the bubble is moved closer to the wall.

VI. SUMMARY

A simple mathematical model has been used to illustrate the response of a bubble to acoustic excitation in a channel formed by parallel plates. The essential role of liquid compressibility in channels with rigid walls is discussed. Explicit expressions are provided for the resonance frequency and damping factor. Decreasing the channel width below about ten bubble diameters decreases both the resonance frequency and maximum response amplitude. For wide channels, an approximate solution is provided for plates with finite impedance.

ACKNOWLEDGMENTS

The authors are grateful to Yurii Ilinskii for many helpful discussions. This work was supported by the IR&D Program at ARL:UT, the National Institutes of Health Grant No. EB004047, and the Office of Naval Research.

[1]E. Sassaroli and K. Hynynen, "Forced linear oscillations of microbubbles in blood capillaries," J. Acoust. Soc. Am. **115**, 3235–3243 (2004).
[2]E. Sassaroli and K. Hynynen, "Resonance frequency of microbubbles in small blood vessels," *Proceedings of the 2004 IEEE International UFFC Joint 50th Anniversary Conference*, pp. 1465–1468.
[3]Y. T. Hu, S. Qin, T. Hu, K. W. Ferrara, and Q. Jiang, "Asymmetric oscillation of cavitation bubbles in a microvessel and its implications upon mechanisms of clinical vessel injury in shock-wave lithotripsy," Int. J. Non-Linear Mech. **40**, 341–350 (2005).
[4]B. Krasovitski and E. Kimmel, "Gas bubble pulsation in a semiconfined space subjected to ultrasound," J. Acoust. Soc. Am. **109**, 891–898 (2001).
[5]X. M. Chen and A. Prosperetti, "Thermal processes in the oscillations of gas bubbles in tubes," J. Acoust. Soc. Am. **104**, 1389–1398 (1998).
[6]H. N. Oğuz and A. Prosperetti, "The natural frequency of oscillation of gas bubbles in tubes," J. Acoust. Soc. Am. **103**, 3301–3308 (1998).
[7]T. G. Leighton, D. G. Ramble, A. D. Phelps, C. L. Morfey, and P. P. Harris, "Acoustic detection of gas bubbles in a pipe," Acta Acust. **84**, 801–814 (1998).
[8]T. G. Leighton, P. R. White, C. L. Morfey, J. W. L. Clarke, G. J. Heald, H. A. Dumbrell, and K. R. Holland, "The effect of reverberation on the damping of bubbles," J. Acoust. Soc. Am. **112**, 1366–1376 (2002).
[9]D. E. Weston, "Acoustic interaction effects in arrays of small spheres," J. Acoust. Soc. Am. **39**, 316–322 (1966).
[10]I. Tolstoy and A. Tolstoy, "Line and plane arrays of resonant monopole scatterers," J. Acoust. Soc. Am. **87**, 1038–1043 (1990).
[11]C. Feuillade, "Scattering from collective modes of air bubbles in water and the physical mechanism of superresonances," J. Acoust. Soc. Am. **98**, 1178–1190 (1995).
[12]D. T. Blackstock, *Fundamentals of Physical Acoustics* (Wiley, New York, 2000), pp. 361–363.
[13]A. A. Doinikov, R. Manasseh, and A. Ooi, "Time delays in coupled multibubble systems (L)," J. Acoust. Soc. Am. **117**, 47–50 (2005).
[14]T. G. Leighton, *The Acoustic Bubble* (Academic, New York, 1994), pp. 136–137.
[15]M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970), Secs. 4.1.24 and 27.8.6.

Acoustic pulse propagation near a right-angle wall

Lanbo Liu[a)]
USA ERDC Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755-1290 and Department of Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, U-2037, Storrs, Connecticut 06269-2037

Donald G. Albert[b)]
USA ERDC Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755-1290

(Received 14 July 2005; revised 3 February 2006; accepted 6 February 2006)

Experimental measurements were conducted around a right-angle wall to investigate the effect of this obstacle on sound propagation outdoors. Using small explosions as the source of the acoustic waves allowed reflected and diffracted arrivals to be discerned and investigated in detail. The measurements confirm that diffraction acts as a low-pass filter on acoustic waveforms in agreement with simple diffraction theory, reducing the peak pressure and broadening the waveform shape received by a sensor in the shadow zone. In addition, sensors mounted directly on the wall registered pressure doubling for nongrazing angles of incidence in line-of-sight conditions. A fast two-dimensional finite difference time domain (FDTD) model was developed and provided additional insight into the propagation around the wall. Calculated waveforms show good agreement with the measured waveforms. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2180530]

PACS number(s): 43.28.En, 43.28.Js, 43.20.El, 43.50.Vt [LCS] Pages: 2073–2083

I. INTRODUCTION

Sound propagation in an urban environment is currently an active research topic in both civil and military applications. Understanding acoustic propagation in this environment is important for noise reduction and for designing and predicting the performance of various sensor systems. The presence of buildings in an urban environment introduces reflections, diffractions, and multiple propagation paths that are not present in simpler situations. As the foundation for studying the effects of multiple buildings, which is the theme for a future paper, the perturbations caused by the presence of a single obstacle are the focus of this study.

The effect of barriers on acoustic propagation has been studied extensively, especially to reduce traffic noise. Ray theory treatments are often discussed in textbooks (e.g., Refs. 1 and 2) and Pierce[3,4] has developed a theoretical treatment that is often used in such applications. In recent years the boundary element method[5–7] has proved useful for solving many of these problems, and a number of analytical or semianalytical methods have been developed to investigate the urban environment and street canyons (e.g., Refs. 8–10).

In part because of the nature of traffic noise, most previous experimental work has focused on the frequency domain, and acoustic pulses have only occasionally been used in the past to investigate simple barriers.[11–16] Nonplanar obstacles are not often treated, although Pierce (Ref. 4, p. 498) discusses a single-frequency calculation for propagation around a building corner.

To examine the nature and properties of various waves involved in the interaction with an obstacle, experimental measurements were conducted outdoors on a full-scale, right-angle wall using impulsive sources. A concrete block wall was constructed in an open field and instrumented with pressure sensors. Small explosive charges were detonated at various locations near the wall and the resulting pressure waveforms were recorded using a digital seismograph. These measurements investigated both line-of-sight (LOS) and non-line-of-sight (NLOS) situations.

In recent years, the finite-difference time domain (FDTD) method has gained popularity in simulating different wave propagation phenomena (e.g., Refs. 17–19). Here, the two-dimensional FDTD method is applied to model the experimental measurements on a personal computer. Some of the advantages of the finite-difference method are the ability to include a variety of acoustic pulses, complex barrier or building geometries, and spatially varying sound speeds. In addition, viewing the computational results in an animated movie format can provide insight to complex wave propagation phenomena arising from the interaction of the waves with obstacles. This paper demonstrates that the FDTD technique is a useful tool to understand sound propagation physics in a complex environment.

The next section discusses experimental measurements that were made on an isolated concrete block wall to investigate the simplest case of building effects, reflection, and diffraction. Next, a two-dimensional finite-difference time domain prediction method is developed and validated by comparison with the experimentally measured waveforms. This method is shown to be an accurate tool for the study of outdoor sound propagation in complex situations.

[a)]Electronic mail: lanbo.liu@erdc.usace.army.mil
[b)]Electronic mail: donald.g.albert@erdc.usace.army.mil

TABLE I. Geometry of the measurements. Grid coordinates in meters of walls, sensors, and source locations corresponding to Fig. 1.

Sensor	*X*	*Y*	*Z*	Description	Line-of-sight (LOS)
	−9.4	0.0		Left end of front wall	
A	−4.6	6.0	0.0	Near front wall	All source locations
B	−4.6	3.0	1.5	Near front wall	All source locations
C	−4.6	1.5	1.5	Near front wall	All source locations
D	−4.6	0.0	1.5	On front wall	All source locations
	0.0	0.0		Corner of both walls	
E	0.0	−6.0	1.5	On side wall	S5
F	1.5	−6.0	1.5	Near side wall	S4–S5
G	3.0	−6.0	1.5	Near side wall	S4–S5
H	6.0	−6.0	1.5	Near side wall	S3–S5
	0.0	−9.4		Lower end of side wall	
I	−4.6	−0.2	1.5	Behind the front wall	S1
J	−4.6	−3.0	0.0	Behind the front wall	S1
K	−4.6	−30.0	0.0	30 m behind front wall	S1–S3, S5
SP1	−34.6	0.0	1.5	Source parallel with front wall	
SP2	−30.6	15.0	1.5	Source 30° from front wall	
SP3	−19.6	26.0	1.5	Source 60° from front wall	
SP4	−4.6	30.0	1.5	Source perpendicular to front wall	
SP5	10.4	26.0	1.5	Source 60° from front wall	

II. EXPERIMENTAL MEASUREMENTS

For the experimental measurements, a right-angle wall with equal sides, each 9.4 m long and 3.5 m high, was constructed in an open field using concrete blocks. Pressure sensors were installed on and near the walls (most at a height of 1.5 m above ground level, see Table I), and 0.28-kg charges of C4 explosive, 1.5 m above ground level, were detonated at various locations 30 m from the center of the front wall (see Fig. 1). The resulting pressure pulses were recorded using a Bison Model 9048 digital seismograph at a sampling rate of 5 kHz and with a response bandwidth of 3–2500 Hz. This experimental geometry allowed the reflection and diffraction of the low-amplitude blast waves by the right-angle wall to be studied in detail.

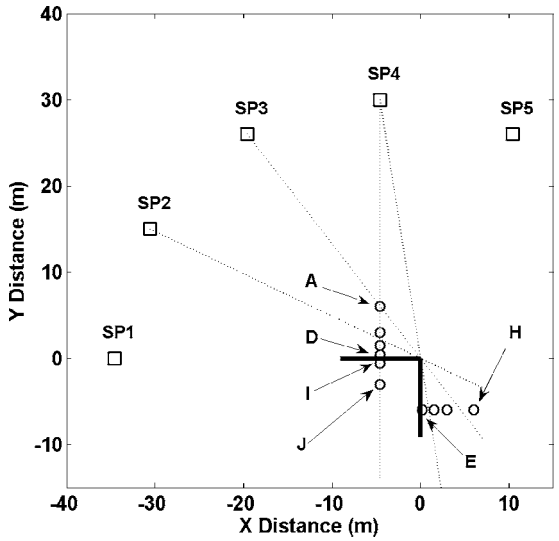


FIG. 1. Plan view of the experimental geometry used for the wall measurements. The horizontal wall segment along the *x* axis is referred to as the front wall, and the vertical wall segment parallel to the *y* axis is called the side wall. Circles indicate the location of pressure sensors and squares the source locations. Dotted lines indicate line-of-sight (LOS) ray paths. A sensor located at (−4.6, −30) is omitted from the plot. The coordinates of all of the sources, sensors, and walls are given in Table I.

The measured pressure waveforms for sensors located near the center of the front and side walls for all of the source positions are shown in Fig. 2. In this figure, the top four waveforms are for sensors A–D near the front wall, while the lower four waveforms are for sensors E–H near the side wall. These measured waveforms reveal a number of characteristics of acoustic pulse interaction with the walls.

The front wall sensors A–D are all within the "line-of-sight" (LOS) of the source positions and have a direct wave arrival in every case. Examining the waveforms in Fig. 2 for these front wall sensors for source position 1 (SP 1) shows that grazing incidence does not have a visible effect on the sensor waveforms. All the waveforms are nearly the same in appearance and have about the same peak pressure amplitude. The positive pulse duration is typically around 5 ms, roughly indicating a peak frequency of about 100 Hz and a wavelength of about 3.5 m. The waveform for sensor A, farthest from the front wall (the top row in Fig. 2), looks slightly different from the others. These differences might occur because this sensor is at ground level, while the other sensors B–D are located 1.5 m above ground level.

At all of the other source positions, reflected waves are visible in the sensors located some distance from the front wall. The interference of the reflected wave with the direct wave depends on the geometry and travel time of the reflected wave. In all cases, the peak pressure of the reflected wave is smaller than the direct wave. The amplitudes of the direct and reflected waves at normal incidence lead to an

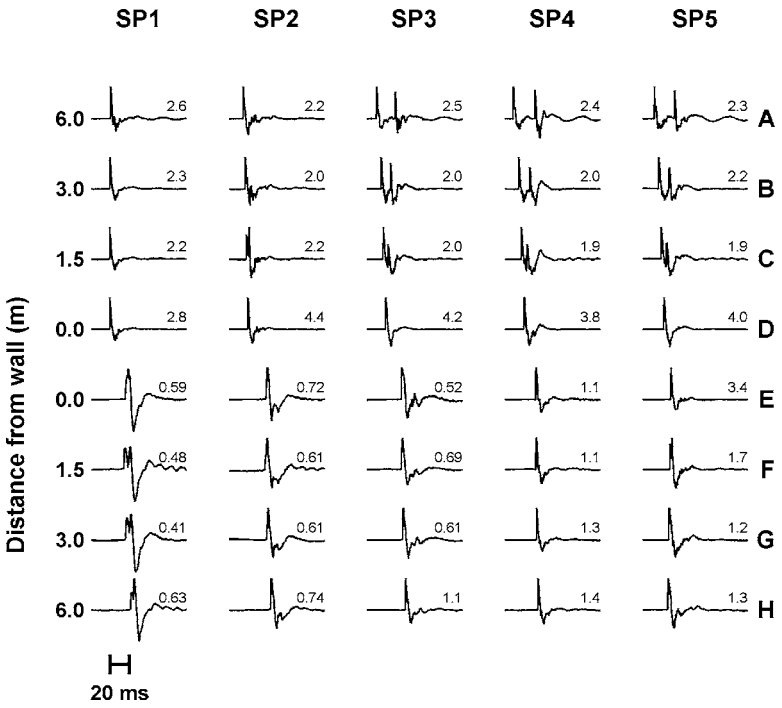


FIG. 2. Normalized pressure waveforms measured for the geometry shown in Fig. 1. Each column contains waveforms recorded for a particular source point and is 100 ms in duration. In each column, the top four waveforms are for sensors A–D located 6, 3, 1.5, and 0 m away from the front wall. The next four waveforms were recorded by sensors E–H located 0, 1.5, 3, and 6 m from the side wall. Each waveform is 100 ms in duration and labeled with the peak absolute pressure in kPa. See Fig. 1 and Table I for the experimental geometry.

estimate for the reflection coefficient of the wall as approximately 0.93. The high value of the reflection coefficient indicates that the finite size of the wall has little effect on the reflected energy.

All of the sensors located near the side wall (sensors E–H) are in the shadow zone for source positions 1 and 2, and some remain in the shadow for source positions 3 (sensors E–G at 0, 1.5, and 3 m) and 4 (sensor E at 0 m). For all of the non-line-of-sight (NLOS) sensors at source positions 1–3, the peak pressure amplitude drops from about 2.2 kPa to about 0.6 kPa, a reduction by a factor of nearly 4. When a sensor is near the shadow boundary, the peak pressure is reduced by about a factor of 2 to about 1.1 kPa. In addition, the NLOS waveforms have longer time durations and broader pulse widths than the LOS waveforms, with the positive duration typically increasing from 5 ms to 6–8 ms. For nongrazing LOS conditions, the wall-mounted sensors have approximately twice the peak pressure of the nearby sensors that are not on the wall. This "pressure doubling" is caused by the constructive interference of the direct and reflected waves.

Figure 3 compares a LOS sensor waveform with a NLOS sensor waveform, recorded with the source at SP 1. The LOS sensor B was located 3 m from the front wall and the NLOS sensor E was located 0 m from the side wall. The LOS waveform is identical to waveforms measured in an open field, with no visible effect from the front wall. The direct wave has peak positive and negative pressures of 2300 and 890 Pa, respectively. For the NLOS sensor, three different diffracted arrivals are possible: a ray around the right-angle corner, a ray over the top of the wall, and a ray around the lower end of the side wall (see Fig. 1). All of these rays should arrive within a few ms of each other, and the small fluctuations near the peak pressure of the waveform indicate that interference is occurring.

Comparison of the pressure waveforms in Fig. 3 shows clearly that the diffracted waveform is a low-pass filtered version of the line-of-sight waveform. The peak pressure in the shadow zone is reduced by a factor of about 3 above

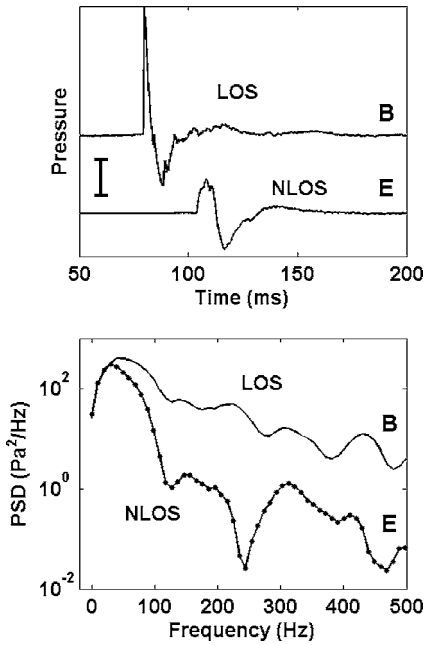


FIG. 3. Measurements illustrating the low-pass filtering effect of diffraction around a corner. The source of the waves was located at SP1 in Fig. 1. The positions of the line-of-sight (LOS) sensor B, 3 m in front of the front wall, and the non-line-of-sight (NLOS) sensor E, mounted on the side wall, are shown in Fig. 1. (Top) Pressure waveforms. The vertical bar indicates 1 kPa, and the measured peak positive pressures were 2300 for the LOS waveform and 590 Pa for the NLOS waveform. For the LOS wave, the positive duration was 4.0 ms and the entire waveform duration was 20.6 ms, while the durations increased to 9.4 and 29.1 ms for the NLOS wave. (Bottom) Power spectral densities of the two waveforms. The NLOS waveform has far less high-frequency content compared to the LOS waveform.

geometrical spreading, and the signal duration is longer. Both of these changes are typical time domain characteristics of low-pass filtering. The spectra calculated from these waveforms (also presented in Fig. 3) show that frequency components of 100 Hz and higher are reduced by two orders of magnitude in the shadow while the lower frequency content is not changed. This low-pass filtering effect is visible in previous higher frequency measurements[11,12] and agrees with insertion loss predictions from simple barrier diffraction theories (e.g. Refs. 1 and 4).

III. FINITE-DIFFERENCE TIME DOMAIN NUMERICAL MODELING

Because barriers are often used in practical situations to reduce traffic or industrial noise, theoretical methods of predicting barrier effects using continuous wave or single-frequency sources have received a great deal of attention. However, these methods may predict constructive or destructive interference for waves with different propagation distances that does not occur when pulses are used instead of continuous sources. In addition, these methods are often limited to simplified geometries. The finite-difference method, a numerical method that overcomes these limitations, is used for the calculations presented here.

A. The finite-difference method

The finite-difference method is an important computational tool that has been used on many different wave propagation problems, including electromagnetics,[20] seismology,[21] and underwater acoustics.[22] In this section the general characteristics of the method are discussed and applied to outdoor sound propagation in the following sections. The basic method is to divide the spatial domain into a discrete grid of nodes and approximate the derivatives appearing in the equations of motion using finite differences between adjacent grid values. Similarly, the time variable is also divided into discrete steps and evaluated using finite differences. This method is commonly called the finite-difference time domain (FDTD) method.

In this case the acoustic propagation is expressed as a set of first-order, velocity-pressure coupled differential equations, similar to the motion and continuity expressions.[23,24] In a two-dimensional (x,y) plane, these equations can be expressed as

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial u}{\partial t},$$

$$\frac{\partial p}{\partial y} = -\rho \frac{\partial v}{\partial t}, \tag{1}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = -\frac{1}{\rho c^2}\frac{\partial p}{\partial t},$$

where u and v are the x and y components of the particle velocity, p is the pressure, ρ is the density, and c is the sound speed of the medium. [These are Eqs. (1)–(3) of Ref. 24.] For a spatial grid size of h, the first-order spatial partial derivatives appearing in the above equations can be approximated by

$$\left.\frac{\partial u}{\partial x}\right|_{t=n\Delta t} \approx \frac{1}{2h}[u(i+1,j;n) - u(i-1,j;n)] \tag{2}$$

with similar approximations for the y and t derivatives. In the above equation, $u(i,j;n)$ represents the x component of the particle velocity at (x,y) grid point (i,j) and at time step n. Using these approximations the equations of motion can be rewritten in finite-difference form (see the Appendix).

There are two steps to determine the size of the grid and the length of the time step to insure numerical stability when implementing the FDTD algorithm. First, for a second-order finite difference, as adapted in this study, 20 spatial nodes per major wavelength are needed to suppress unwanted numerical dispersion. Next, for a given grid size h, the time step must meet the theoretical Courant stability criterion[24]

$$\Delta t \leqslant h/(c\sqrt{2}) \tag{3}$$

for the two-dimensional case. (In three-dimensional cases the square root of 2 is replaced by the square root of 3 in the denominator.) However, for problems involving materials having a high impedance contrast, the Courant condition of Eq. (3) may not be sufficient to insure numerical stability. A more stringent stability criterion with a much smaller time step has been imposed on the calculations in this paper of

$$\Delta t = h/(16c) \tag{4}$$

and is discussed further below.

The finite-difference equations are solved using a staggered difference algorithm proposed by Yee[20] in a two-dimensional spatial domain (see the Appendix). The computations are also staggered between the air pressure p and the particle velocity v in the time domain. Yee's staggered grid algorithm remains the most economical and robust way to carry out finite-difference time domain calculations.[19]

The perfectly matched layer (PML) technique[25] was adapted for the absorption boundary condition and achieved highly effective suppression of reflections from the domain boundaries. Numerical experiments demonstrated that an eight-layer PML boundary condition reduced the reflected error by 30–40 dB over any previously proposed absorption boundary conditions.[17] This method has also been successfully implemented in modeling seismic waves[26] and radar waves.[27]

B. Two-dimensional approximation

While the solution of the finite-difference equations discussed in the previous section is straightforward, a very fine spatial grid is required for an accurate solution as discussed below. This requirement normally limits finite-difference calculations for outdoor sound propagation problems to supercomputers or specialized multiprocessor networks. To reduce the computational effort and make the problem tractable on a desktop computer, a simplified two-dimensional model is used to represent the real three-dimensional world. This reduction of the calculations to two dimensions has substantial benefits but also introduces a number of approximations and

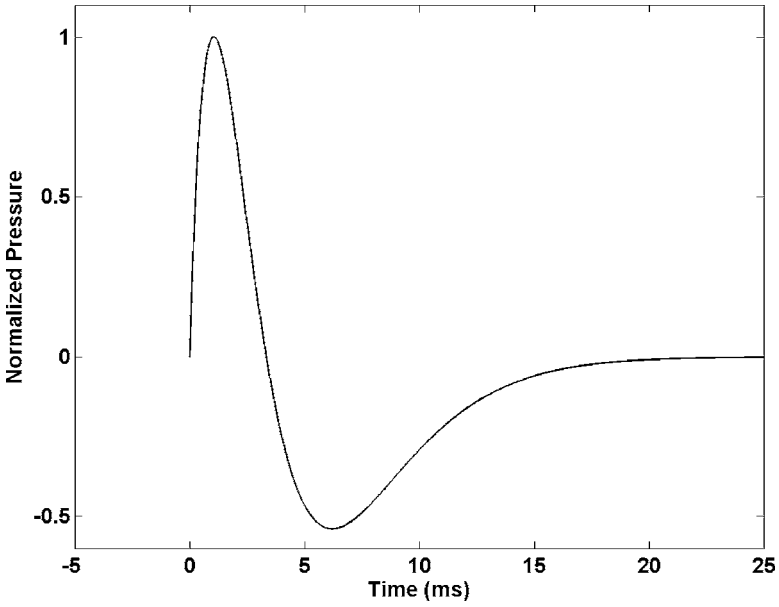


FIG. 4. The source pulse used in the finite-difference calculations. This pulse waveform was calculated from Eq. (5) with the parameter values of $f=150$ Hz, $a=16$, $b=\frac{1}{4}$, and $c=\frac{1}{3}$.

limitations to the modeling capability. While the kinematics of the computations (e.g., the speed and shape of the wavefronts and the arrival times) remain accurate, the dynamics (pressure amplitudes) differ from the three-dimensional situation and must be corrected if the calculations are to be compared to real measurements.

In the real physical world and in actual measurements, a point source generates a spherical acoustic wave with a geometric spreading factor of $1/ct=1/r$. However, the two-dimensional model uses a line source, which extends to infinity in the direction perpendicular to the model plane. This source generates a cylindrical wave with the geometric spreading of $1/\sqrt{ct}=1/\sqrt{r}$. Thus the two-dimensional modeling results are renormalized by an additional factor of $1/\sqrt{r}$ to account for the actual three-dimensional geometric spreading present in the measured data.

The two-dimensional model plane coincides with the horizontal ground plane, so the model does not intrinsically include the effect of the ground surface. To compare the calculated waveforms with the measured data, the ground is assumed to be rigid (finite ground impedance effects are neglected) and the direct and reflected path lengths equal; these assumptions imply that the calculated model pressures be multiplied by a factor of 2 to account for ground reflection. The factor of 2 is included in the modeling results when compared with the recorded data. This approach is accurate for frequencies up to about 600 Hz, but would have to be modified for higher frequency propagation or for propagation over more porous grounds.

Memory constraints in the 1-GB personal computer used in the calculations limit the problem size to a total of about 1.2 million spatial grid points. With a spatial grid interval of 0.1 m, a two-dimensional problem of 110 m^2 can be studied, compared to a volume of only 10.6 m^3 that could be computed in the full three-dimensional case. The reader should also remain aware that energy outside the plane of the propagation model is ignored, i.e., propagation over the top of a berm, wall, or building is not included in the model. As the results below will show, this limitation introduces only a surprisingly small error.

C. Source waveform

The source waveform used for the calculations (Fig. 4) is based on measurement close to the source and is modified from the theoretical blast wave source pulse presented by Reed.[28] The analytical expression for the source pressure as a function of time is

$$p(t) = A(1 - a[ft - b]^2) \exp(-cft), \tag{5}$$

where $p(t)$ is the source pressure at time t, A is the source strength, a, b, and c are constants that determine the delay, rise time, and the amplitude of the negative trough of the source waveform, and f is the central frequency of the pulse. A value for f of 150 Hz was found to produce a source waveform similar to that measured for the type and size of explosive sources used in this study. When used to construct the starting pulse for the simulations, this frequency value also produced good agreement with the measured waveforms as will be discussed below. Other source waveforms can easily be used in the FDTD simulations if desired.

D. Material parameters used in the simulations

A sound speed of 353 m s^{-1} was measured using a blank pistol as the source of the waves, slightly higher than the expected sound speed of about 350 m s^{-1} for the air temperature of about 32 °C. However, for the small explosions a sound speed of 370 m s^{-1} was measured. This higher speed is a weak nonlinear effect caused by higher pressures of the blast waves, and corresponds to a Mach number of about 1.05.[29] Thus in the finite difference calculations a sound speed of 370 m s^{-1} and a density of 1.2 kg m^{-3} are used to represent air, giving an acoustic impedance of 440 kg m^{-2} s^{-1}.

Concrete has a nominal density of 2300 kg m^{-3} and an acoustic wave speed of 2950 m s^{-1} for an acoustic impedance of 6.8×10^6 kg m^{-2} s^{-1}. The very large impedance contrast with air imposes a more strict stability condition than the classic Courant condition[30] so that the final selections of the grid size and time step were determined by the material contrast. However, using realistic values of the material properties for air and concrete makes the simulation unrealistically expensive in both memory and CPU time. To avoid these problems, the impedance contrast must be reduced to a manageable level. In these simulations, the full velocity contrast (a ratio of 1/8) is implemented to capture the correct propagation phenomenology, but the density contrast is limited to about 1/20 instead of the true ratio of 1/2000 by reducing the density of the concrete wall from 2300 kg m^{-3} to a value of 23 kg m^{-3}. Physically, this change does not affect the arrival times of the waves or the propagation behavior. Using a lower density for concrete reduces the impedance ratio from approximately 15 000 to about 150, and also makes a slight change in the reflection and transmission coefficients. For example, the true normal reflection coefficient value of 0.999 94 will be replaced by a value of 0.989 74, about a 1% change in reflection coefficient.

As shown by Schröder and Scott,[30] a high impedance contrast may introduce numerical instabilities in the finite-difference solution. Three steps have been taken in the calculations presented here to eliminate these instabilities. First, the impedance contrast between air and concrete is reduced by lowering the density of concrete used in the calculations as discussed in the preceding paragraph. This adjustment reduces the impedance ratio from $15k$ to 150, while the reflection coefficient and wave velocities remain unaffected. Second, following Schröder and Scott,[30] the first grid layer at the air-wall boundary uses the average impedance between air and concrete instead of the full impedance contrast. Finally, the time step is reduced from the classical Courant condition given in Eq. (3) to the much more stringent condition of Eq. (4). Imposing these conditions on the calculations insured their stability and also gave good accuracy, as demonstrated by comparison with field measurements in the following section.

Although rigid boundaries can be used in the simulations, they may introduce complications in the discretization of complex geometries and cause further numerical instability and so were not included in the calculations presented here. The shear properties of the wall are also ignored in these acoustic wave simulations.

1. Analytical comparison

To validate the approximations made to reduce the impedance contrast between the concrete walls and air, the analytical reflection coefficient for a planar surface is compared to the reflection coefficient calculated using the FDTD method. A symmetric Ricker wavelet with a frequency of 100 Hz was used in this finite-difference calculation. The analytical reflection coefficient is given by

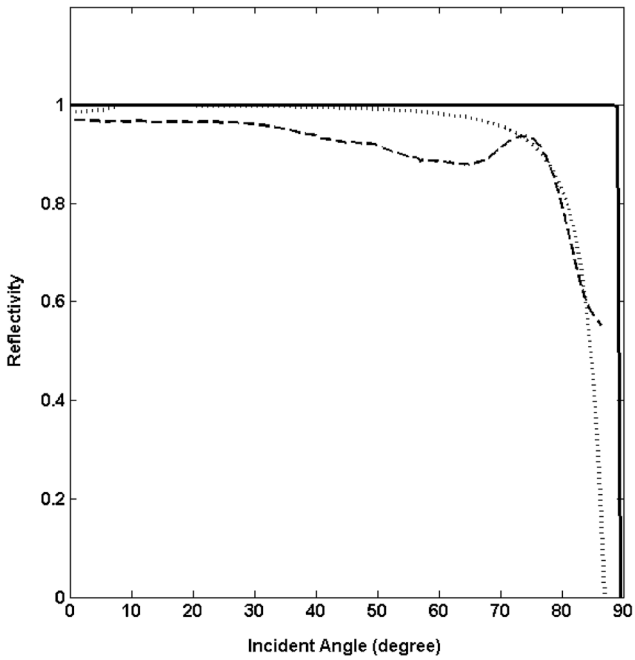


FIG. 5. Comparison of the analytical reflection coefficient with the FDTD reflection coefficient for an air-concrete surface as a function of angle of incidence. The solid line is the analytical result for the true concrete impedance ($Z=15k$) and the dotted line is the analytical result when the concrete density is reduced as discussed in the text. These coefficients were calculated using Eq. (6). The dashed line is the finite difference result for a 100-Hz Ricker wavelet.

$$R=\frac{Z_1\cos\theta-Z_0\cos\theta}{Z_1\cos\theta+Z_0\cos\theta}, \tag{6}$$

where $Z_i=\rho_i c_i$ are the concrete and air impedances and θ is the angle of incidence (with $\theta=0$ at normal incidence). Figure 5 compares two analytical reflection coefficients, one with the full impedance contrast and one calculated with a lower concrete density as discussed previously. All of the calculation methods give a reflection coefficient near one for all angles of incidence below 80°. The finite-difference method matches the reduced density concrete coefficient at grazing angles above 80°. This approximation is acceptable for the propagation environment discussed in this paper.

The major advantage of simulating the building effects with a two-dimensional FDTD code is that fewer computational resources are needed compared to a full three-dimensional simulation. For example, each two-dimensional simulation in this paper required about 155 min to complete on a 1.6-GHz personal computer. A substantial part of this time was used to write 250 movie frames for each computation. Because the finite-difference code is written in MATLAB, converting to a compiled language like FORTRAN is likely to provide substantial improvements in the speed of the computations in the future.

IV. SIMULATION RESULTS FOR A RIGHT-ANGLE WALL

In this section, the finite-difference simulation for the experimental geometry shown in Fig. 1 is discussed and

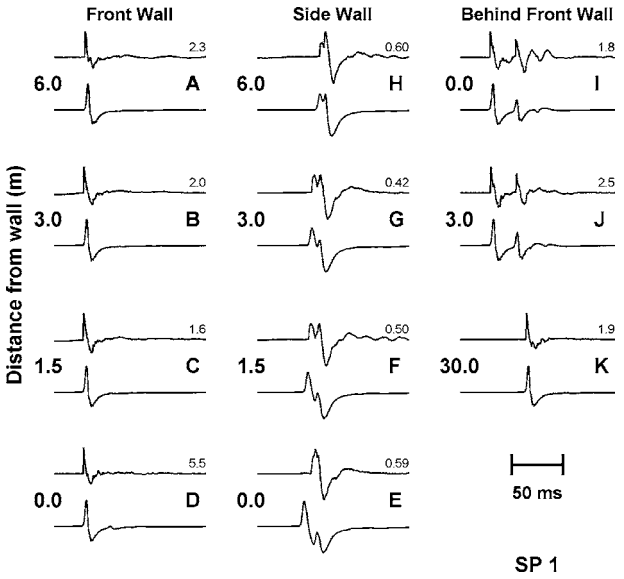


FIG. 6. Comparison of measured and calculated pressure waveforms for the geometry shown in Fig. 1 with the source at SP 1. The left and center columns show waveforms for sensors A–D located near the front wall and sensors E–H located near the side wall. The right column shows the sensors I and J located behind the front wall. For each sensor location, the top trace is the measured signal (with the absolute peak pressure in kPa indicated) and the bottom trace is the calculated signal. Each waveform is 150 ms in duration and has been normalized. (The upper left and bottom center measured waveforms were shown in Fig. 3.)

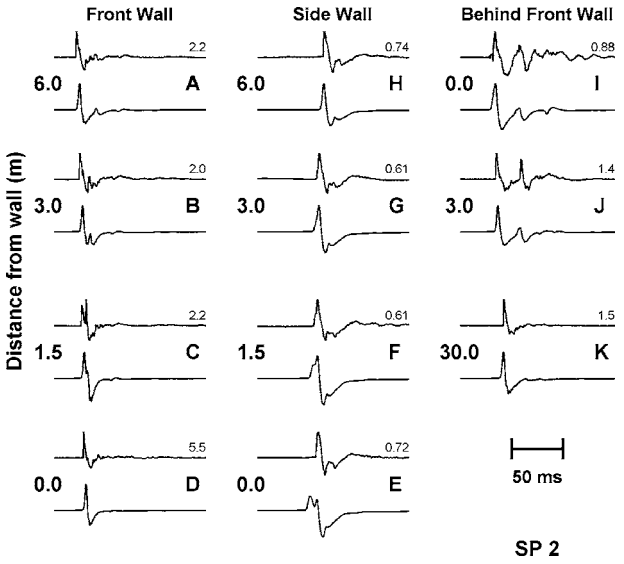


FIG. 7. Comparison of measured and calculated pressure waveforms for the geometry shown in Fig. 1 with the source at SP 2. The left and center columns show waveforms for sensors A–D located near the front wall and sensors E–H located near the side wall. The right column shows the sensors I and J located behind the front wall. For each sensor location, the top trace is the measured signal (with the absolute peak pressure in kPa indicated) and the bottom trace is the calculated signal. Each waveform is 150 ms in duration and has been normalized. The bottom right waveforms for the location 30 m behind the front wall have been advanced by 50 ms for this plot.

compared to the measured waveforms. The two-dimensional finite-difference simulation was run separately for sources located in the five source positions (SPs) shown in Fig. 1. For these simulations, a 600×700 grid of nodes was used with a spatial sampling of 0.1 m; thus, the simulated area was 60×70 m with absorbing boundaries at the edges. The time step was 16.67 μs, and the simulation ran for 15 000 time steps, providing a total time window of 0.25 s.

In the calculations, the concrete wall was assigned a density of 23 kg m^{-3} and an acoustic wave speed of 2950 m s^{-1}, while the air surrounding the wall was assigned a density of 1.2 kg m^{-3} and a sound speed of 370 m s^{-1} as discussed earlier. For the frequency band (less than 200 Hz) and propagation distances (less than 100 m) used here, the intrinsic absorption of the atmosphere is negligible and has been ignored in the calculations.

A. Comparison with measurements in the time domain

Figures 6–10 compare the measured waveforms with the waveforms calculated using the finite-difference method for each source location. For each waveform pair shown in the figures, the upper waveform has been measured experimentally while the lower waveform has been calculated using the two-dimensional FDTD method. The left column of each of these figures shows the sensors (A–D) located near the front wall. These sensors are line-of-sight (LOS) in all cases. The center column of each figure shows the results from the sensors (E–H) located near the side wall, and the right column of each figure shows the results from the sensors (I–K) located behind the front wall. Sensors in the center and right columns are sometimes NLOS as discussed previously; see Fig. 1 and Table I.

In Fig. 6 (for SP 1), the source is located at a grazing angle with respect to the front wall, and no reflected waves appear in the measurements or in the simulations. As the

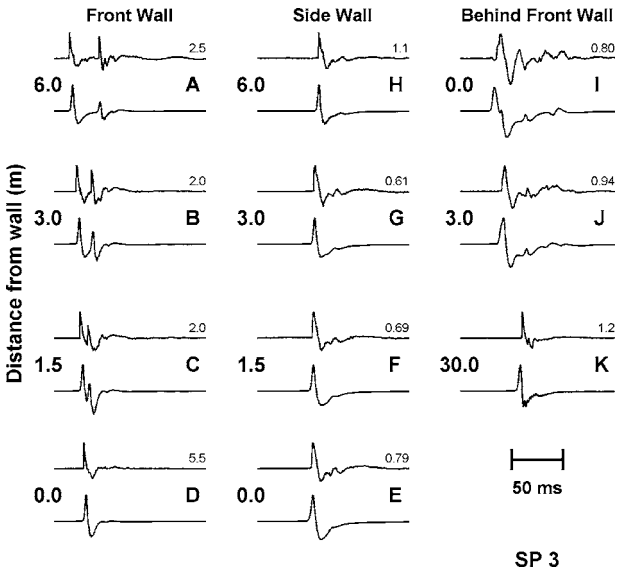


FIG. 8. Comparison of measured and calculated pressure waveforms for the geometry shown in Fig. 1 with the source at SP 3. The left and center columns show waveforms for sensors A–D located near the front wall and sensors E–H located near the side wall. The right column shows the sensors I and J located behind the front wall. For each sensor location, the top trace is the measured signal (with the absolute peak pressure in kPa indicated) and the bottom trace is the calculated signal. Each waveform is 150 ms in duration and has been normalized. The bottom right waveforms for the location 30 m behind the front wall have been advanced by 50 ms for this plot.

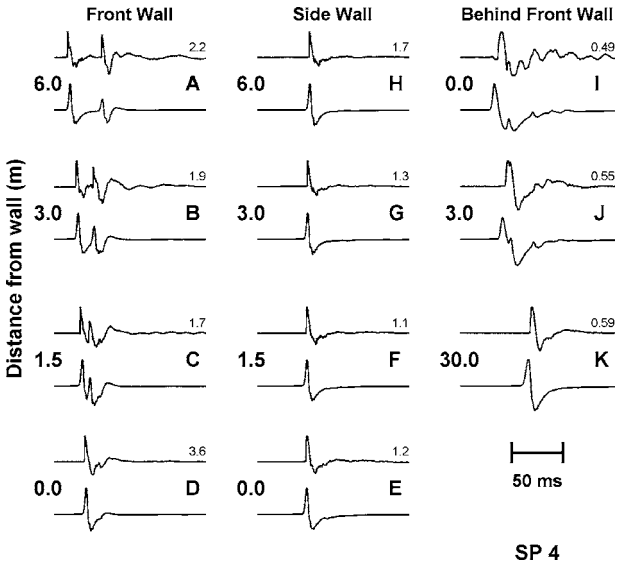


FIG. 9. Comparison of measured and calculated pressure waveforms for the geometry shown in Fig. 1 with the source at SP 4. The left and center columns show waveforms for sensors A–D located near the front wall and sensors E–H located near the side wall. The right column shows the sensors I and J located behind the front wall. For each sensor location, the top trace is the measured signal (with the absolute peak pressure in kPa indicated) and the bottom trace is the calculated signal. Each waveform is 150 ms in duration and has been normalized. The bottom right waveforms for the location 30 m behind the front wall have been advanced by 50 ms for this plot.

source position is moved to produce the following figures, the waveforms in the left column show the reflected wave interfering with the direct wave. In all cases, the agreement between the observed and calculated waveforms is very good.

The center column of Figs. 6–10 shows the waveforms

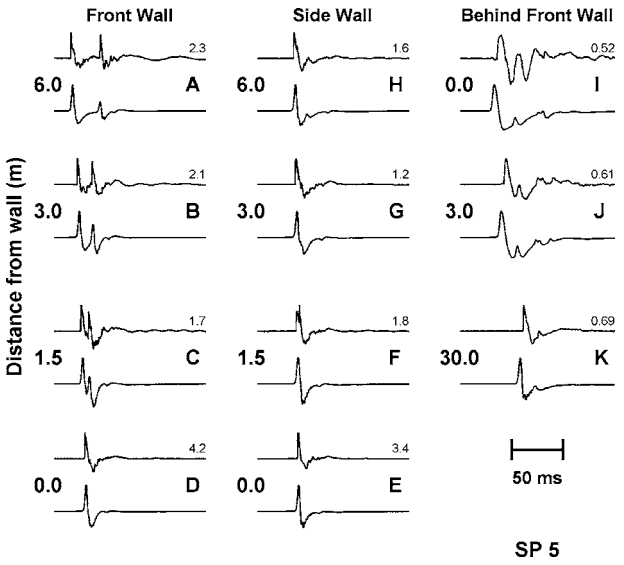


FIG. 10. Comparison of measured and calculated pressure waveforms for the geometry shown in Fig. 1 with the source at SP 5. The left and center columns show waveforms for sensors A–D located near the front wall and sensors E–H located near the side wall. The right column shows the sensors I and J located behind the front wall. For each sensor location, the top trace is the measured signal (with the absolute peak pressure in kPa indicated) and the bottom trace is the calculated signal. Each waveform is 150 ms in duration and has been normalized. The bottom right waveforms for the location 30 m behind the front wall have been advanced by 50 ms for this plot.

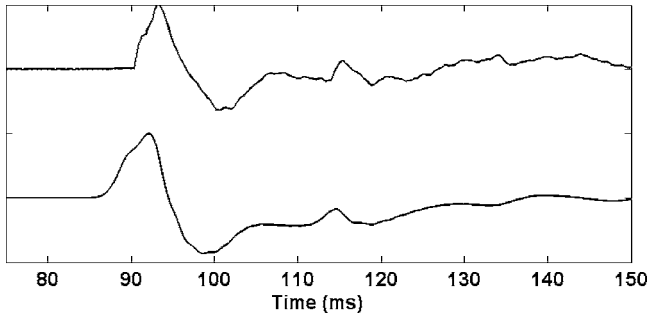


FIG. 11. Comparison of measured and calculated pressure waveforms for sensor J located 3 m behind the front wall and with the source located at SP 3. The top trace is the measured signal and the bottom trace is the FDTD-calculated signal. Each waveform is 75 ms in duration and has been normalized. Based on comparison with the calculated waveform and the wavefront movie (Fig. 12), the arrival at 93 ms is identified as a diffraction around the left edge of the front wall (−9.4,0), the arrival at 115 ms is this same wave reflected off the rear (the left side in Fig. 1) of the side wall, and the arrival at 135 ms is a diffraction from the bottom corner of the side wall (0.0, −9.4). Refer to Fig. 1 for the wall geometry.

for sensors E–H on the side wall. With the source at SP 1 (Fig. 6), all of these sensors are directly behind the side wall in the deep shadow zone. All of the observed and calculated waveforms have longer durations than the LOS sensors. The waveform agreement is generally good, although there are some visible differences. The observed waveforms often show a second arrival peak that is larger in amplitude than the second arrival in the calculated waveforms. This difference may arise because the calculations are ignoring any diffractions over the top of the wall due to the two-dimensional nature of the model. As the source moves in Figs. 6–10, the diffraction angle decreases and the agreement between the observed and calculated waveforms improves. The transition from NLOS to LOS is accurately depicted in the calculated waveforms.

The right column in Figs. 6–10 compares the sensors (I–K) located behind the front wall. Here, the results for SP 1 (Fig. 6) show direct waves and waves reflected from the left surface of the side wall (Fig. 1) for the first two sensors, while the sensor located 30 m away is unaffected by the wall's presence. The agreement between the observed and calculated waveforms is very good in this and in subsequent figures. Again, the NLOS waveforms exhibit longer durations and lower amplitudes than the LOS waveforms. The measured waveforms for the sensor located directly on the back of the front wall show more oscillations than the calculated waveforms for this position and more oscillations than any of the other measurements in all cases. This excessive oscillation is believed to be caused by a loose sensor mounting or poor sensor performance ("ringing"). Despite these discrepancies, the main characteristics of the calculated waveforms are visible in the measured waveforms.

B. Comparison with measurements in the spatial domain

One very useful advantage of the finite-difference method is the ability to save snapshots of the pressure wave field at any time step in the calculation. By assembling such snapshots in sequence, a movie of the wave-field evolution is

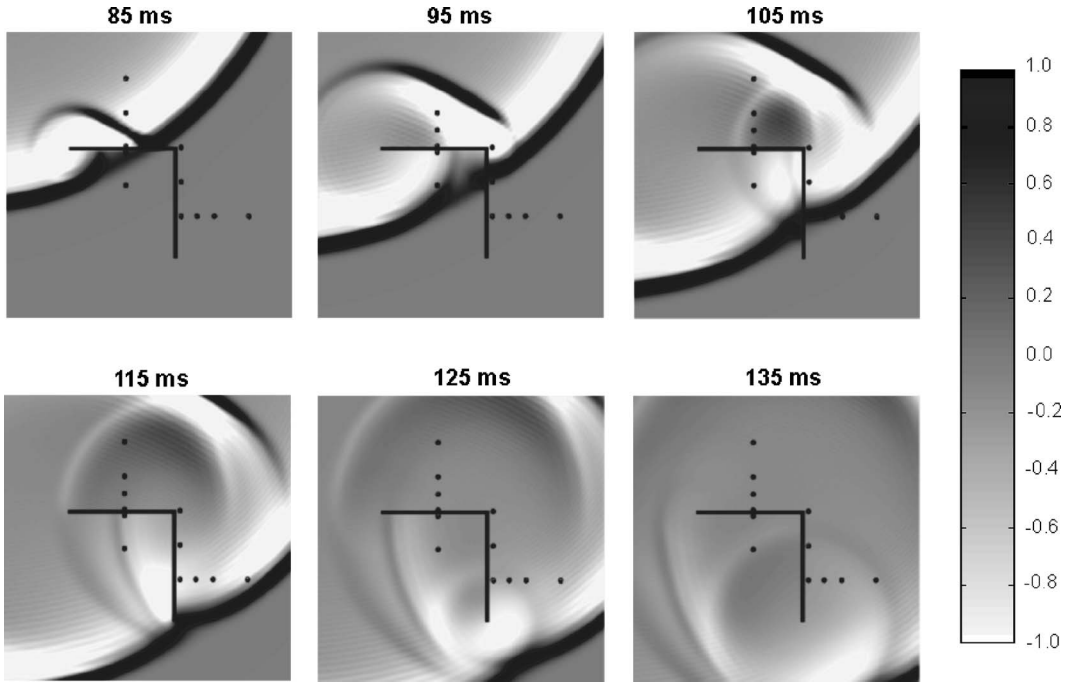


FIG. 12. Snapshots of an acoustic pulse produced by a source at SP 3 interacting with the concrete wall. These movie frames were generated using the finite-difference time domain method discussed in the text.

produced that is very useful for understanding details of the individual sensor waveforms. As an example, the measured and calculated waveforms for sensor J, located 3.0 m behind the front wall, when the source was at SP 3 is shown in Fig. 11. For the measured waveform, aside from the largest arrival at 93 ms, other positive pressure arrivals are visible at 115 and 135 ms. For the calculated waveform, the arrival times are about 91, 112, and 135 ms, respectively.

Figure 12 shows some selected wave-field snapshots for the SP 3 source position. By examining these snapshots and the full movie, we can determine the source of the individual positive arrivals. From Fig. 12, the large first arrival on the sensor 3 m from the back of the front wall is clearly seen to be a wave diffracting around the left corner of the front wall. There is also probably some transmission through the wall. In the measurement, there must also be energy propagating over the top of the wall that is ignored in the calculations.

By following the wave-fronts through a series of snapshots, the arrival at about 115 ms is identified as the reflection of the diffracted wave from the back (left) side of the side wall. Finally, the arrival at 135 ms can be identified as a diffracted wave from the bottom corner of the side wall, a wave that is rarely considered in diffraction problems. Although the pressure amplitudes of these later arrivals are small, their identification through the use of the wave-field snapshots adds to the understanding of the acoustic wave interaction with the wall. The ability to identify and elucidate wave arrivals may be especially helpful in situations with more complicated geometries, for example when full four-sided or multiple buildings are present.

V. CONCLUSIONS

Measurements of acoustic pulse propagation near a concrete right-angle wall were presented in this paper. The measurements confirm that diffraction acts like a strong low-pass filter, greatly reducing frequencies above 100 Hz and reducing the peak amplitude by a factor of about $\frac{1}{4}$.

The finite-difference method was used to calculate theoretical waveforms for comparison with the experimental measurements. Using a simplified two-dimensional representation of the wall gave fast computation times and good agreement with measured arrival times, peak amplitudes, and waveform characteristics for which no analytical techniques exist. This simulation method can be used to investigate propagation in more complicated multiple building urban environments. The comparison shows that the limitation to a two-dimensional geometry speeds up the calculations significantly while introducing only small errors in accuracy.

ACKNOWLEDGMENTS

We thank, Rich Andrejkovics, SYDET Project Officer, PM-CCS, for funding the experimental work, and the US Army Corps of Engineers Engineering Research and Development Center for funding the analysis. Michael Scanlon, Army Research Laboratory, assisted in the design of the single wall experiment, and Bill Davis, Aberdeen Test Center, safely and efficiently handled the work with the explosives. We also thank the many people who assisted with the field measurements, including Steve Decato, Dave Carbee, and Dr. Joyce Nagle of ERDC-CRREL, and Chris Reiff and Dave Gonski of ARL. Finally, we greatly appreciate the helpful comments of the Associate Editor, Lou Sutherland, and two anonymous reviewers.

APPENDIX: YEE (1966) ALGORITHM USING STAGGERED GRID AND LEAPFROG TIME-STEPPING

The method proposed by Yee[20] to solve the Maxwell equations for electromagnetic wave propagation uses a leap-

frog scheme on staggered Cartesian grids. This method can be applied to acoustic problems[23,24] using the first-order, velocity-pressure coupled differential equations:

$$\frac{\partial p}{\partial x} = -\rho \frac{\partial u}{\partial t},$$

$$\frac{\partial p}{\partial y} = -\rho \frac{\partial v}{\partial t}, \tag{A1}$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = -\frac{1}{\rho c^2}\frac{\partial p}{\partial t},$$

where u and v are the x and y components of the particle velocity, p is the pressure, ρ is the density, and c is the sound speed of the medium. [These are Eqs. (1)-(3) of Ref. 24 and Eq. (1) of this text.] For a spatial grid size of h, the first-order spatial partial derivatives appearing in the above equations can be approximated by

$$\left.\frac{\partial u}{\partial x}\right|_{t=n\Delta t} \approx \frac{1}{2h}[u(i+1,j;n) - u(i-1,j;n)] \tag{A2}$$

with similar approximations for the y and t derivatives. In the above equation, $u(i,j;n)$ represents the x component of the particle velocity at (x,y) grid point (i,j) and at time step n. Following Yee's approach and using the approximation in (A2) the equations of motion can be solved in finite-difference form [see Eqs. (7)–(9) in Ref. 24]:

$$u_{i+1/2,j}^{n+1/2} = u_{i+1/2,j}^{n-1/2} - \frac{\Delta t}{\rho_{i,j}\Delta x}(p_{i+1,j}^{n} - p_{i,j}^{n}),$$

$$v_{i+1/2,j}^{n+1/2} = v_{i+1/2,j}^{n-1/2} - \frac{\Delta t}{\rho_{i,j}\Delta y}(p_{i,j+1}^{n} - p_{i,j}^{n}), \tag{A3}$$

$$p_{i,j}^{n+1} = p_{i,j}^{n} - \Delta t \rho_{i,j} c_{i,j}^2 \left(\frac{u_{i+1/2,j}^{n+1/2} - u_{i-1/2,j}^{n+1/2}}{\Delta x} + \frac{v_{i,j+1/2}^{n+1/2} - v_{i,j-1/2}^{n+1/2}}{\Delta y} \right).$$

The staggered Cartesian grid used in Eq. (A3) is sketched in Fig. 13. Equations (A3) show how updated values of pressure and particle velocity are obtained from previous values on the staggered grid.

The Yee algorithm is robust, fast, and simple to understand and has several properties and advantages that differ from some other numerical finite-difference methods. First, the Yee method simultaneously solves two coupled physical fields (pressure and particle velocities in acoustics) through a set of first-order partial differential equations, instead of solving for a single, second-order partial differential equation for one of the two physical fields. This solution algorithm is more robust than solving either field alone.

Second, the algorithm is very flexible in taking advantage of specific features unique to each field, such as acoustically hard (particle velocity assigned to be zero) or acoustically soft (pressure to be zero) regions. Third, the finite-difference expressions for spatial derivatives are central difference in nature and second-order accurate.

Fourth, the leapfrog time-stepping used is also central difference in nature and second-order accurate. Since this approach is fully explicit, it avoids problems involved with simultaneous equations and matrix inversion. Finally, the method can be used to determine the response in a chosen frequency band in one calculation by using a pulse excitation. This cannot be achieved with frequency domain methods.

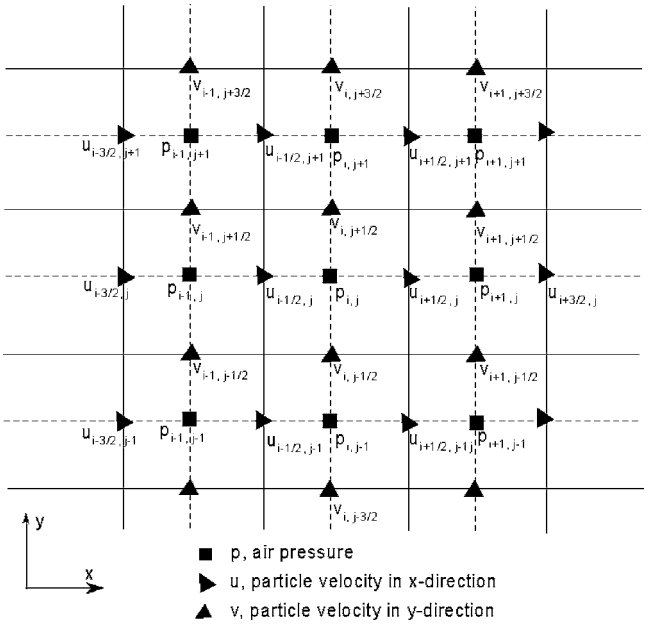


FIG. 13. A sketch of the two-dimensional staggered grid system in the x-y plane used by the finite-difference method for simulation of the acoustic field. The index i is used to indicate x coordinates, and the index j to indicate y coordinates.

The disadvantages of the Yee algorithm are those shared by all finite-difference approaches. If the Cartesian grid does not conform well to the real geometry of a particular problem, so called "stair stepping" errors will be introduced. In addition, the absence of a general subgriding scheme means that structures smaller than the resolution have to be treated by subcell models. These models are currently limited to relatively simple geometrical structures such as wires, gaps, layers, etc., and introduce additional complexity to the calculations.

[1] D. R. Raichel, *The Science and Applications of Acoustics* (Springer-Verlag, New York, 2000).
[2] E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer Academic, Dordrecht, 2001).
[3] A. D. Pierce, "Diffraction of sound around corners and over wide barriers," J. Acoust. Soc. Am. **55**, 941–955 (1974).
[4] A. D. Pierce, *Acoustics. An Introduction to its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY, 1989).
[5] D. C. Hothersall, S. N. Chandler-Wilde, and M. N. Hajmirzae, "Efficiency of single noise barriers," J. Sound Vib. **146**, 303–322 (1991).
[6] S. N. Chandler-Wilde, "The boundary element method in outdoor noise propagation," Proc. Inst. Acoust. **19** (Part 8), 27–50 (1997).
[7] T. W. Wu, ed., *Boundary Element Acoustics: Fundamentals and Computer Codes* (WIT, Boston, 2000).
[8] J. Kang, "Sound propagation in street canyons: Comparison between diffusively and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[9] J. Picaut, L. Simon, and J. Hardy, "Sound field modeling in streets with a diffusion equation," J. Acoust. Soc. Am. **106**, 2638–2645 (1999).
[10] K. K. Iu and K. M. Li, "The propagation of sound in narrow street canyons," J. Acoust. Soc. Am. **112**, 537–550 (2002).
[11] C. G. Don, "Application of hard truncated wedge theory of diffraction to

wide barriers," J. Acoust. Soc. Am. **90**, 1005–1010 (1991).
[12]A. I. Papadopoulous and C. G. Don, "A study of barrier attenuation by using acoustic impulses," J. Acoust. Soc. Am. **90**, 1011–1018 (1991).
[13]D. J. Saunders and R. D. Ford, "A study of the reduction of explosive impulses by finite sized barriers," J. Acoust. Soc. Am. **94**, 2859–2875 (1993).
[14]R. Raspet, J. Ezell, and S. V. Coggeshall, "Diffraction of an explosive transient," J. Acoust. Soc. Am. **79**, 1326–1334 (1986).
[15]H. Medwin, E. Childs, and G. M. Jebsen, "Impulse studies of double diffraction: a discrete Huygens interpretation," J. Acoust. Soc. Am. **72**, 1005–1013 (1982).
[16]K. B. Rasmussen, "Model experiments related to outdoor propagation over an earth berm," J. Acoust. Soc. Am. **96**, 3617–3620 (1994).
[17]X. Yuan, D. Borup, J. W. Wiskin, M. Berggren, R. Eidens, and S. A. Johnson, "Formulation and validation of Berenger's PML absorbing boundary for the FDTD simulation of acoustic scattering," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 816–822 (1997).
[18]H. Dong, A. M. Kaynia, C. Madshus, and J. M. Hovem, "Sound propagation over layered poro-elastic ground using a finite-difference model," J. Acoust. Soc. Am. **108**, 494–502 (2000).
[19]A. Taflove and S. C. Hagness, *Computational Electromagnetics* (Artech House, Boston, 2000).
[20]K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media," IEEE Trans. Antennas Propag. **14**, 302–307 (1966).
[21]R. A. Stephen, "A review of finite difference methods for seismo-acoustics problems at the seafloor," Rev. Geophys. **26**, 445–458 (1988).
[22]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[23]J. Virieux, "SH-wave propagation in heterogeneous media: Velocity-stress finite difference method," Geophysics **49**, 1933–1942 (1984).
[24]S. Wang, "Finite-difference time-domain approach to underwater acoustic scattering problems," J. Acoust. Soc. Am. **99**, 1924–1931 (1996).
[25]J. P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[26]G. Festa and S. Nielsen, "PML absorbing boundaries," Bull. Seismol. Soc. Am. **93**, 891–903 (2003).
[27]L. Liu and S. A. Arcone, "Numerical simulation of the wave-guide effect of the near-surface thin layer on radar wave propagation," J. Environ. Eng. Geophys. **8**, 133–141 (2003).
[28]J. W. Reed, "Atmospheric attenuation of explosion waves," J. Acoust. Soc. Am. **61**, 39–47 (1977).
[29]G. F. Kinney and K. J. Graham, *Explosive Shocks in Air* (Springer-Verlag, Berlin, 1985).
[30]C. T. Schroder and W. R. Scott, "On the stability of the FDTD algorithm for elastic media at a material interface," IEEE Trans. Geosci. Remote Sens. **40**, 474–481 (2002).

Sound generated by vortices in the presence of a porous half-cylinder mounted on a rigid plane

C. K. Lau and S. K. Tang[a)]
Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong, People's Republic of China

(Received 7 October 2004; revised 5 January 2006; accepted 12 January 2006)

The sound generated by a single vortex and by two identical vortices in the presence of a half-cylinder made of porous material mounted on a rigid horizontal plane is studied theoretically using the acoustic analogy and the matched asymptotic expansion method. Both longitudinal and transverse dipoles are observed upon the introduction of the porous cylinder, but the former is considerably stronger than the latter in all the cases studied. Results suggest that the amplitudes of the dipoles and the overall acoustical energy radiated can be higher than that in the rigid cylinder case under some suitable combinations of flow parameters, especially when the flow resistance inside the porous material seen by the vortices is very small. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2171838]

PACS number(s): 43.28.Ra, 43.50.Nm, 43.50.Gf [DRD] Pages: 2084–2095

I. INTRODUCTION

Noise is an unwanted byproduct from building services machinery, such as an air-conditioning unit, a fan coil unit, dampers, registers, louvers, guide vanes, etc. This noise propagates inside ventilation ducts conveying low Mach number turbulent flows and is usually attenuated by reactive or passive methods.[1] The most common device for this purpose is the dissipative silencer which dissipates sound energy by the viscous actions inside the porous material. Lighthill[2] established the importance of aerodynamic sound generation as a result of turbulence interaction, while Curle[3] extended the theory to include the influence of solid boundaries. Flow turbulence is expected to generate sound even in the presence of acoustically absorptive materials. The self-noise from a dissipative silencer is a typical example of such aerodynamic noise production.[4]

Theoretical studies usually deal with rigid wall boundaries,[5] and studies related to the characteristics of the porous materials on sound generation are rarely found in the existing literature. However, there are studies on the sound attenuation performance of a dissipative silencer in the absence of flow turbulence through different approaches. Examples include that of Cummings and Chang[6] who investigated the attenuation of a dissipative silencer of finite length with mean flow by mode matching technique, that of Quinn and Howe[7] who investigated theoretically the production and absorption of acoustic energy when a sound wave impinges on the edges of an acoustic lossless liner, and the finite element study of Peat and Rathi[8] on the sound fields inside dissipative silencers.

The results of Ffowcs Williams[9] show that sound is produced when flow turbulence is near sound-absorbent liners and thus provide further theoretical support to the self-noise production of dissipative silencers and acoustic wall linings. A recent study of the authors[10] deals with the vortex sound generation in the presence of a wedge made up partially of porous material. It is found that the additional acceleration of the vortex in the presence of the porous material has a significant effect on the radiated sound energy.[10]

Although vortices are a drastic simplification of real turbulence, they have provided insights into the generation processes of flow noises (for instance, Crighton).[11] In the present investigation, the sound generated by the unsteady motions of vortices in the proximity of a porous half-cylinder on an otherwise rigid horizontal plane is studied. The complex potential and the velocities of the vortices are evaluated through the use of conformal mapping as in Tang,[12] while the outer far-field potential is derived by using the matched asymptotic method as in Obermeier.[13] The present configuration is intended to represent one flow boundary inside a dissipative silencer. It is hoped that the present results can enhance the understanding of aerodynamic sound generation under the influence of porous surfaces and be useful in the future modeling of the self-noise generated by dissipative silencers.

II. THEORETICAL DEVELOPMENT

Two rectilinear vortices with circulations Γ_1 and Γ_2 initially located at the complex locations z_1 and z_2, respectively, interact with a half-cylinder composed of porous material as shown in Fig. 1(a). The height of each vortex above the rigid plane is denoted by h, while the horizontal vortex separation is represented by d. The properties of the porous material are characterized by the effective fluid density, ρ_e, and the flow resistance R_f inside its lattice.[14] The former represents the fluid inertia effect while the latter includes the effect of viscosity. Unless the fluid is perfectly inviscid, one should note that owing to the very tiny fluid passages inside the porous material, the effect of viscosity on the fluid motion inside

a)Author to whom correspondence should be addressed. Electronic mail: besktang@polyu.edu.hk

 0001-4966/2006/119(4)/2084/12/$22.50

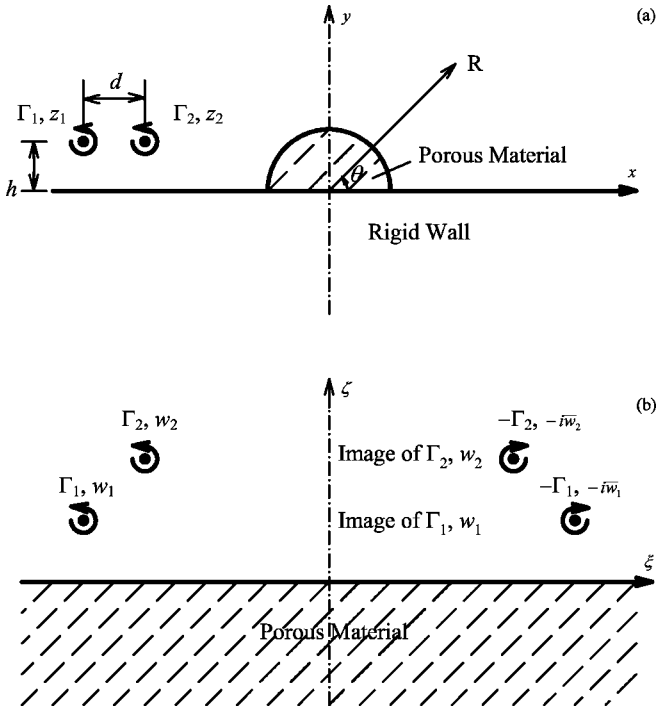


FIG. 1. Schematics of vortex model and nomenclatures: (a) Original z plane and (b) w plane.

this material cannot be neglected though the external flow outside it can be satisfactorily represented by the inviscid model.[12] Also, the introduction of the porous material results in finite surface flow impedance, which may lower the ability of this boundary to support fluid pressure and produce a pressure-releasing effect. More information on pressure-releasing boundaries can be found in, for instance, Schneider *et al.*[15]

The corresponding complex potentials and the velocities of the vortices can be worked out as in Tang and Lau.[10] With the help of the conformal mapping,[16] the original z-plane ($z=x+iy$) is transformed into the w-plane ($w=\xi+i\zeta$) as shown in Fig. 1(b), and the mapping function is

$$z = f(w) = -\frac{i+w}{i-w} \Rightarrow w = f^{-1}(z) = i\frac{z+1}{z-1}. \tag{1}$$

The stream function, Ψ_{wj}, and the transverse velocity, V_{wj}, of the jth vortex in the w plane can then be obtained by matching the fluid pressure and normal fluid velocity along the impedance boundary.[12] In the present study, all length scales are normalized by the cylinder radius a, and the strengths of the vortices are normalized by the total vortex strength Γ ($=\Gamma_1+\Gamma_2$). Here, time is normalized by a^2/Γ. V_{wj} and R_f are normalized by Γ/a and $\rho\Gamma/a^2$ respectively (ρ being the density of the incompressible fluid medium). The results of Tang[12] suggest that, for an isolated vortex j,

$$\Psi_{wj} = \frac{\Gamma_j}{4\pi}\int_{-\infty}^{\infty}\frac{1}{|k|}(e^{-|k|\zeta_j} + g_j e^{|k|\zeta_j})\frac{e^{-|k|\zeta}}{g_j}e^{ik(\xi_j-\xi)}dk, \tag{2}$$

where $g_j(k) = R_f + ikV_{wj}(1+m) \big/ -R_f + ikV_{wj}(1-m)$ and $m = \rho_e/\rho$. The corresponding vortex velocity is

$$V_{wj} = \frac{\Gamma_j}{4\pi}\int_{-\infty}^{\infty}\frac{-e^{-2|k|\zeta_j}}{g_j}dk. \tag{3}$$

The overall stream function, Ψ_w, in the presence of other vortices is therefore

$$\Psi_w = \sum_{j=1}^{4}\frac{\Gamma_j}{4\pi}\int_{-\infty}^{\infty}\frac{1}{|k|}(e^{-|k|\zeta_j} + g_j e^{|k|\zeta_j})\frac{e^{-|k|\zeta}}{g_j}e^{ik(\xi_j-\xi)}dk, \tag{4}$$

and the velocity of the jth vortex, u_{wj} and v_{wj}, in the w plane are

$$u_{wj} = V_{wj} + \sum_{l=1\neq j}^{4}\left.\frac{\partial\Psi_{wl}}{\partial\zeta}\right|_{\xi=\xi_j,\zeta=\zeta_j},$$

$$v_{wj} = -\sum_{l=1\neq j}^{4}\left.\frac{\partial\Psi_{wl}}{\partial\xi}\right|_{\xi=\xi_j,\zeta=\zeta_j}. \tag{5}$$

Also, $\Gamma_3=-\Gamma_1$, $u_{w3}=-u_{w1}$, $v_{w3}=v_{w1}$, and $\Gamma_4=-\Gamma_2$, $u_{w4}=-u_{w2}$, $v_{w4}=v_{w2}$. The paths of the vortices in the z plane are calculated by integrating Eq. (5) numerically using the standard fourth-order Runge–Kutta method together with the Routh's correction.[17]

It can be shown, as in Tang and Lau[10] with the Cauchy–Riemann principle, that

$$\phi_w = \sum_{j=1}^{4}\frac{1}{2\pi}\int_{0}^{\infty}\frac{\Gamma_j}{k}(e^{-k\zeta_j} + g_j e^{k\zeta_j})\frac{e^{-k\zeta}}{g_j}\sin(k(\xi_j-\xi))dk + C, \tag{6}$$

where C is the integration constant that can be evaluated by observing that the flow potential vanishes as $|z|\to\infty$. The incompressible flow potential in the z plane, ϕ_z, can then be found by substituting w by $f^{-1}(z)$ in Eq. (6). The far-field potential, ϕ_{zo}, can then be obtained using the matched asymptotic expansion.[13] The far-field sound pressure is

$$p = -\frac{\partial\phi_{zo}}{\partial t}, \tag{7}$$

where t is the far-field observer time and p is normalized by $\rho\Gamma^2/a^2$. The foregoing equations are specifically derived for the present study.

A. Perfectly inviscid fluid

When the flow resistance R_f inside the lattice of the cylinder vanishes, it can be shown using Ref. 18 that

$$g_j = \frac{1+m}{1-m}$$

and

$$C = -\sum_{j=1}^{4}\frac{\Gamma_j}{2\pi}\left(\frac{1}{g_j}\tan^{-1}\frac{\xi_j}{1+\zeta_j} + \tan^{-1}\frac{\xi_j}{1-\zeta_j}\right).$$

Morse and Ingard[14] remarked that $1<m<5$ for practical porous materials, but m can be very large if the porous material

is replaced by a hard solid. In the study of the transmission loss across dissipative silencers, the data in Peat and Rathi,[8] Cummings and Sormaz,[19] and Kirby[20] give $m \approx 2.5$. By substituting $w=f^{-1}(z)$ into Eq. (6), the flow potential in z plane is thus

$$\phi_z = \sum_{j=1}^{4} \frac{\Gamma_j}{2\pi} \mathrm{Im}\left\{ \frac{1}{g_j}\left[\ln\left(1-\frac{1}{z\bar{z}_j}\right) - \ln\left(1-\frac{1}{z}\right)\right] + \left[\ln\left(1-\frac{z_j}{z}\right) - \ln\left(1-\frac{1}{z}\right)\right]\right\}. \tag{8}$$

The far-field inner potential produced by the vortices in a perfectly inviscid fluid is, for large $|z|$,

$$\phi_{zi} = -\sum_{j=1}^{2} \frac{\Gamma_j}{\pi r}\left(r_j + \frac{1}{g_j r_j}\right)\sin(\theta_j)\cos\theta + O(r^{-2}), \tag{9}$$

where (r_j, θ_j) are the polar coordinates of the jth vortex and (r, θ) those of a point in the flow field.

The far field so produced in the frequency domain is the solution of the Helmholtz equation $\nabla^2\phi + k^2\phi = 0$, which is $\phi = \sum_{j=1}^{4} A_j H^{(1)}_{\alpha_j}(kr)e^{i\alpha_j\theta}$, where $H^{(1)}_{\alpha j}$ is the Hankel function of the first kind of order α_j, and k is the wave number. The matched asymptotic expansion[12] suggests that for low-frequency sound radiation, $\alpha_j = 1$ and $A_j = -\Gamma_j\omega/2c\left[\left(r_j + 1/g_j r_j\right)\sin(\theta_j)\right]^t$, where $k=\omega/c$, c is the ambient speed of sound and $[\ ]^t$ represent the Fourier transform with respect to time. At a large distance R, one obtains with the property of the Hankel function[21] that for positive ω,

$$\phi = -\sum_{j=1}^{2} \frac{\Gamma_j \omega}{2c}\left[\left(r_j + \frac{1}{g_j r_j}\right)\sin(\theta_j)\right]^t \sqrt{\frac{2c}{\pi\omega R}} e^{i(\omega R/c - 3\pi/4)}. \tag{10}$$

The far-field potential ϕ_{zo} can be obtained by using the inverse Fourier transform:

$$\phi_{zo} = -\frac{1}{2\pi}\sum_{j=1}^{2}\int_{-\infty}^{\infty} \frac{\omega\Gamma_j}{2c} \times \left[\left(r_j + \frac{1}{g_j r_j}\right)\sin(\theta_j)\right]^t \sqrt{\frac{2c}{\pi\omega R}} e^{i(\omega R/c - 3\pi/4)} e^{-i\omega t} d\omega. \tag{11}$$

Thus, the far-field sound pressure is

$$p = -\frac{\partial\phi_{zo}}{\partial t} = \frac{1}{\pi}\sqrt{\frac{1}{2c\mathrm{R}}}\sum_{j=1}^{2}\frac{\partial}{\partial t}\int_{-\infty}^{t-\mathrm{R}/c}\frac{\partial}{\partial\tau}\left[\left(r_j + \frac{1}{g_j r_j}\right)\sin(\theta_j)\right]\frac{\Gamma_j d\tau}{\sqrt{t-\tau-\mathrm{R}/c}}\cos\theta. \tag{12}$$

Equation (12) shows that the pressure generated in a perfectly inviscid fluid is a longitudinal dipole. For a rigid cylinder (very large m and $g_j \to -1$), it reproduces the result of Abou-Hussein *et al.*[22]

B. Combined effects of effective fluid density and flow resistance

When the flow resistance R_f is finite, the effects from the porous material become complicated. The flow potential in the w plane is

$$\phi_w = \sum_{j=1}^{4}\frac{\Gamma_j}{2\pi}\left\{ -\tan^{-1}\frac{\xi_j-\xi}{\zeta+\zeta_j} + \tan^{-1}\frac{\xi_j-\xi}{\zeta-\zeta_j} + \frac{2}{1+m}\int_0^\infty \frac{k e^{-k(\zeta+\zeta_j)}\sin(k(\xi_j-\xi))}{(\beta_j^2+k^2)}dk + \frac{2R_f}{|V_{wj}|(1+m)^2}\int_0^\infty \frac{e^{-k(\zeta+\zeta_j)}\cos(k(\xi_j-\xi))}{\beta_j^2+k^2}dk + C\right\}, \tag{13}$$

where

$$C = \sum_{j=1}^{4}\frac{i\Gamma_j}{2\pi}\left\{ -\mathrm{Re}[\ln(1+i\bar{w}_j)] + \mathrm{Re}[\ln(1+iw_j)] - \frac{2}{1+m}\mathrm{Im}[-ci(\beta_j\mu_j)\cos(\beta_j\mu_j) - si(\beta_j\mu_j)\sin(\beta_j\mu_j)] - \frac{2}{1+m}\mathrm{Re}[ci(\beta_j\mu_j)\sin(\beta_j\mu_j) - si(\beta_j\mu_j)\cos(\beta_j\mu_j)]\right\},$$

where $\bar{w}_j$=conjugate of w_j, $\mu_j = (1+\zeta_j) - i\xi_j$ and $\beta_j = R_f/[|V_{wj}|(1+m)]$, and *si* and *ci* are the sine and cosine integrals, respectively. The velocity of each vortex has to be estimated by iteration as in Tang and Lau[10] and Tang.[12]

The corresponding flow potential ϕ_z is

$$\phi_z = \sum_{j=1}^{4}\frac{\Gamma_j}{2\pi}\left[\begin{array}{l} -\mathrm{Im}\left\{\ln\left(1-\frac{1}{z\bar{z}_j}\right) - \ln\left(1-\frac{z_j}{z}\right)\right\} \\ +\frac{2}{1+m}\mathrm{Im}\left\{\begin{array}{l}[-ci(\beta_j\eta_j)\cos(\beta_j\eta_j) - si(\beta_j\eta_j)\sin(\beta_j\eta_j)] \\ -[-ci(\beta_j\mu_j)\cos(\beta_j\mu_j) - si(\beta_j\mu_j)\sin(\beta_j\mu_j)]\end{array}\right\} \\ +\frac{2}{1+m}\mathrm{Re}\left\{\begin{array}{l}[ci(\beta_j\eta_j)\sin(\beta_j\eta_j) - si(\beta_j\eta_j)\cos(\beta_j\eta_j)] \\ -[ci(\beta_j\mu_j)\sin(\beta_j\mu_j) - si(\beta_j\mu_j)\cos(\beta_j\mu_j)]\end{array}\right\}\end{array}\right], \tag{14}$$

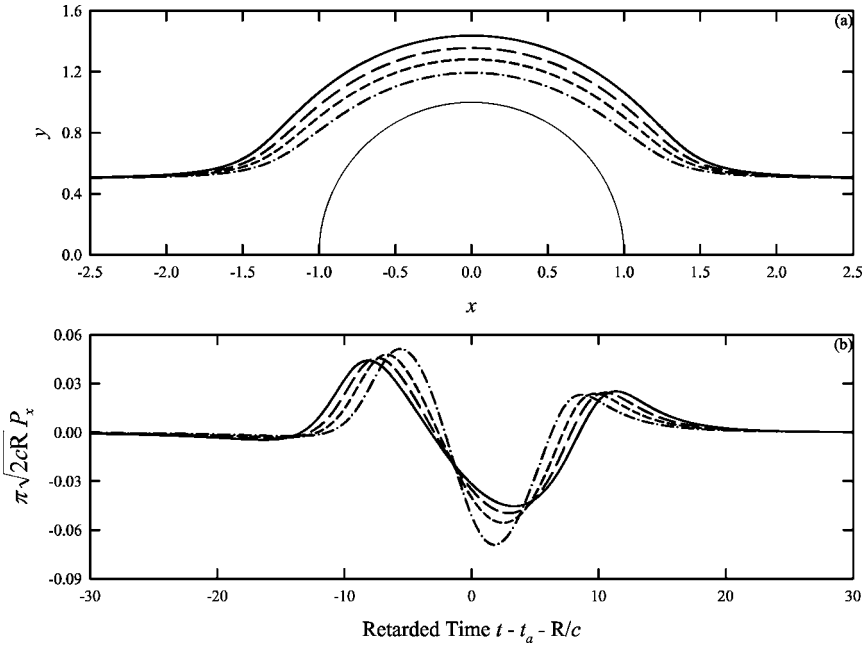


FIG. 2. Effect of pressure-releasing surface on vortex motion and sound generation: (a) Vortex path and (b) sound pressure. Initial $z_1=(-10,0.5)$. $-\cdot-$: $m=3$; - - -: $m=5$; — — —: $m=10$; and ——: Rigid cylinder.

where $\eta_j=(\zeta+\zeta_j)-i(\xi_j-\xi)$. The flow potential ϕ_{zi} at large $|z|$ becomes

$$\phi_{zi}=\sum_{j=1}^{4}\frac{\Gamma_j}{2\pi r}\left[\left(r_j-\frac{1}{r_j}\right)\sin(\theta-\theta_j)-f_{1j}\cos\theta - f_{2j}\sin\theta\right], \tag{15}$$

where

$$f_{1j}=\frac{2\beta_j}{1+m}\,\mathrm{Re}[\exp(i\gamma_j\beta_j)G(0,i\gamma_j\beta_j)+\exp(-i\gamma_j\beta_j)G(0,-i\gamma_j\beta_j)],$$

$$f_{2j}=\frac{4\beta_j}{1+m}\,\mathrm{Im}[ci(\gamma_j\beta_j)\cos(\gamma_j\beta_j)+si(\gamma_j\beta_j)\sin(\gamma_j\beta_j)],$$

$\gamma_j=1+r_je^{i\theta_j}+1/r_je^{i\theta_j}-1$, and $G(0,\chi)$ is the incomplete gamma function.[21] The far-field outer potential, ϕ_{zo}, can be obtained in the same way as in the previous two cases. Equation (15) indicates that a transverse dipole of magnitude f_{2j} exists when the flow resistance is finite. A longitudinal dipole of magnitude f_{1j} adds to the cylinder dipole. These findings have not been reported elsewhere.

One should also note that R_f in the present study is normalized by $\rho\Gamma/a^2$. Therefore, this parameter can vary over a very wide range. For weak vortex strength, R_f can be very large and it decreases as the vortex strength increases. It vanishes in the case of a perfectly inviscid fluid. The parameter ranges for real/experimental flows[19,20,23–27] are summarized in the Appendix for the sake of comparison with those in the present study. One can notice from the later discussions that the acoustic radiation with $R_f=100$ is already close to those of the rigid cylinder. In the foregoing discussions, the far-field sound pressure is evaluated at a radial distance R of 100.

III. RESULTS AND DISCUSSIONS

A. Single vortex

The paths of a single vortex translating over a rigid wall mounted half-cylinder and the corresponding sound radiations under different speeds of the mean flow have been studied by Abou-Hussein *et al.*[22] The magnitude of the sound pressure increases as h decreases. Active sound generation is observed during the period when the vortex undergoes a substantial rate of change of velocity when it is close to the cylinder. A single vortex moving over a rigid plane generates no sound.

For a perfectly inviscid fluid, the flow resistance vanishes ($R_f=0$). Figure 2(a) shows the effect of m on the vortex path with initial $h=0.5$. The path of a single vortex engaging a rigid wall mounted half-cylinder is also shown for comparison. The present theory indicates that the vortex path converges to that for the rigid wall condition at very large m. The vortex path bends toward the cylinder surface because of the pressure-releasing effect. The smaller the value of m, the greater the degree of bending toward the cylinder surface. This is similar to the case where a vortex moves in the vicinity of a wedge with inhomogeneous surface flow impedance.[12] The vortex resumes its original height as it gradually moves away from the cylinder (at $x>2$). When the initial h increases, less severe vortex path bending can be observed at a fixed m.

Equation (12) suggests that the far-field sound pressure is a longitudinal dipole (P_x) for a hard cylinder ($g_j\to-1$) or a perfectly inviscid fluid ($g_j=(1+m)/(1-m)$). The sound pressure increases [Fig. 2(b)] as the vortex comes closer to the cylinder surface and undergoes substantial longitudinal and transverse accelerations. The time t_a hereinafter, unless otherwise specified, denotes the time at which the vortex passes across the y axis ($x=0$). The pressure fluctuation patterns for various m are pulselike and are similar to that in the

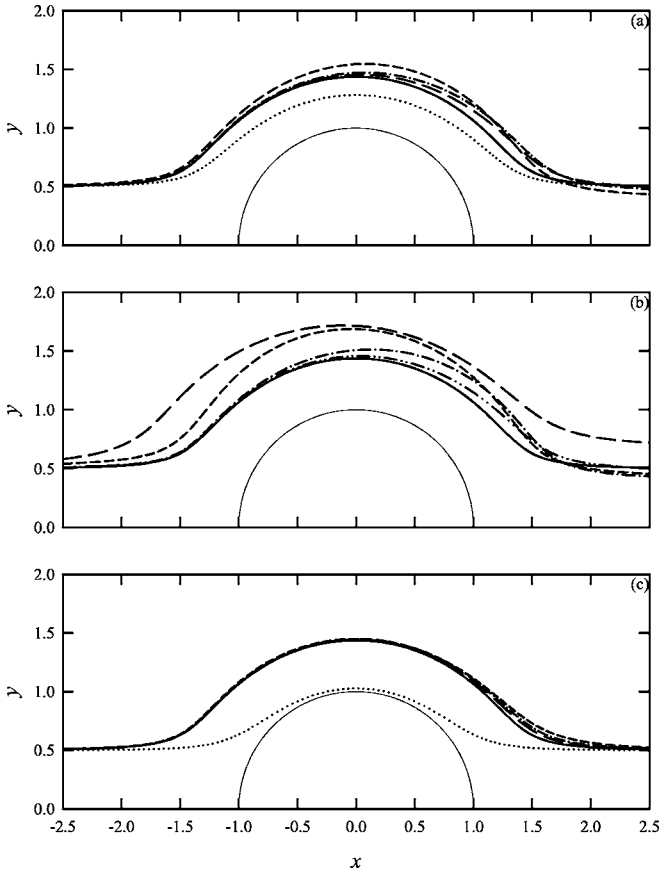


FIG. 3. Combined effects of effective fluid density and flow resistance on the vortex path. (a) $m=5$; ·····: $R_f=0$; - - -: $R_f=1$; −·−: $R_f=5$; — — —: $R_f=10$; and ——: Rigid cylinder. (b) $R_f=5$; −·−·−: $m=3$; −·−: $m=10$; - - -: $m=100$; — — —: $m=1000$; and ——: Rigid cylinder. (c) $m=1.5$. ·····: $R_f=0$; - - -: $R_f=1$; −·−: $R_f=5$; — — —: $R_f=10$; and ——: Rigid cylinder. Initial location of the vortex at $x=-10$ with initial $h=0.5$.

case of a hard cylinder, except that the duration of active sound production is reduced as m decreases. Amplifications of the first peak and trough of the pressure fluctuations are observed after the introduction of the porous material. The extent of such amplification increases with decreasing m. It is found that a decrease in either the initial h or m will lead to an increase in the strength of the far-field sound pressure fluctuation.

For a finite flow resistance ($R_f \neq 0$), previous results suggest that m and R_f produce reactive and resistive effects respectively.[12,14] Figure 3(a) shows such effects on the vortex path at a fixed m with the vortex initially located at $x=-10$ with $h=0.5$. The vortex paths for $R_f > 10$ are close to that of the rigid cylinder case and thus are not presented. The vortex moves away from the cylinder surface at $x<2$ for $1 \leqslant R_f < 10$. Further away from the cylinder at $x>2$, it is observed that the vortex height decreases when R_f is increased from 0 to 1, but it gradually rises back to $y=0.5$ at larger R_f. It is expected that when R_f is reduced toward zero, the pressure-releasing effect becomes more important such that the vortex path bends toward to the cylinder again and converges to that of the perfectly inviscid fluid case. One can also notice from Fig. 3(a) that the vortex paths under a finite R_f are not symmetrical about the y axis.

Figure 3(b) illustrates the effects of m on the vortex path with R_f fixed at 5. The vortex path bends away from the cylinder surface as in Fig. 3(a). The degree of the initial path bending increases with m for $m \leqslant 1000$. The vortex height h after the vortex passes over the cylinder first drops below 0.5 as m increases from 3, but rises up above 0.5 as m further increases from 10. One is expecting that h will resume the value of 0.5 as $m \to \infty$. When m is close to unity, the pressure-releasing effect is very strong. The vortex comes very close to the cylinder surface if the fluid is perfectly inviscid, but its path approaches that under the hard wall condition as R_f is increased as shown in Fig. 3(c).

Unlike the situation in an inviscid flow, the present far-field sound pressure consists of a longitudinal dipole, P_x, and a transverse dipole, P_y [Eq. (15)]. Figures 4(a) and 4(b) show some examples of the time variations of P_x and P_y for $m=5$ at various R_f, respectively. It is observed that the intro-

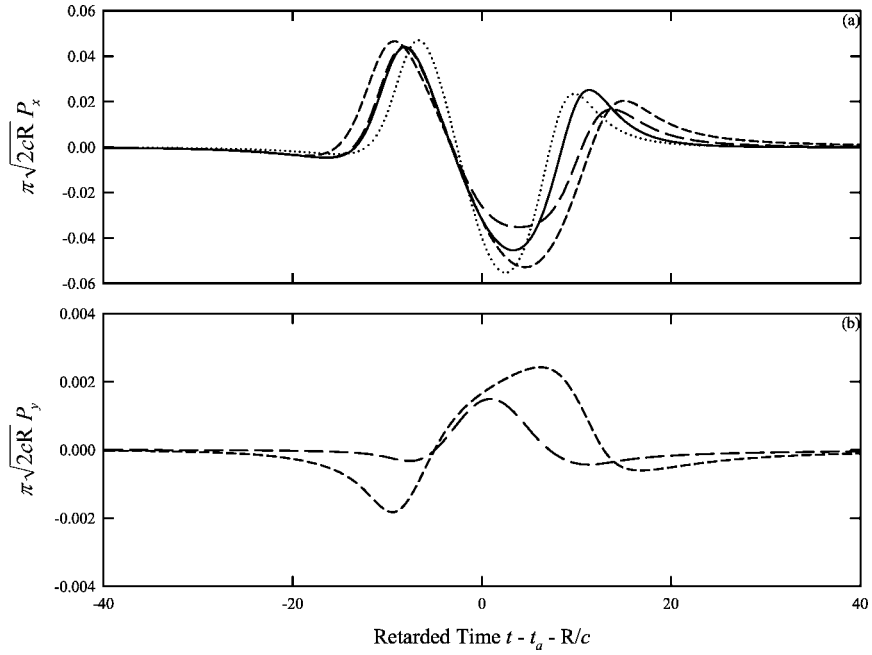


FIG. 4. Sound pressure time variation for $m=5$ at different R_f: (a) Longitudinal dipole and (b) transverse dipole. ·····: $R_f=0$; - - - -: $R_f=1$; — — —: $R_f=10$; and ——: Rigid cylinder.

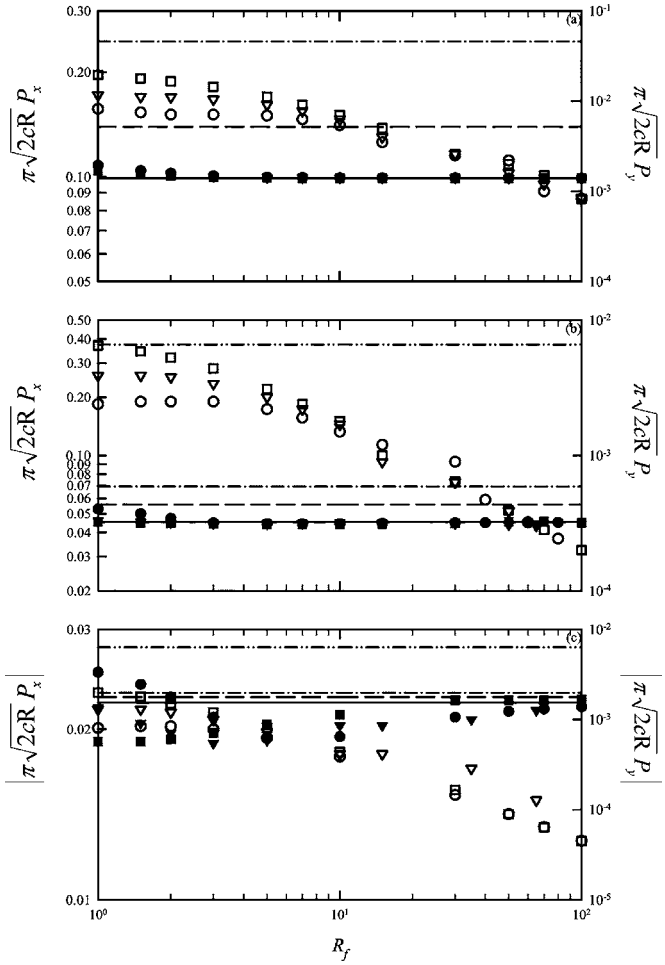


FIG. 5. Combined effects of effective fluid density, flow resistance, and initial vortex height on radiated sound amplitude: (a) $h=0.3$; (b) $h=0.5$; and (c) $h=0.8$. $-\cdot\cdot-$: P_x for $m=1.5$, $R_f=0$; $-\cdot-$: P_x for $m=3$, $R_f=0$; — — —: P_x for $m=5$, $R_f=0$; and ——: Rigid cylinder. □: $m=1.5$; ▽: $m=3$; and ○: $m=5$. Closed symbols for P_x; open symbols for P_y.

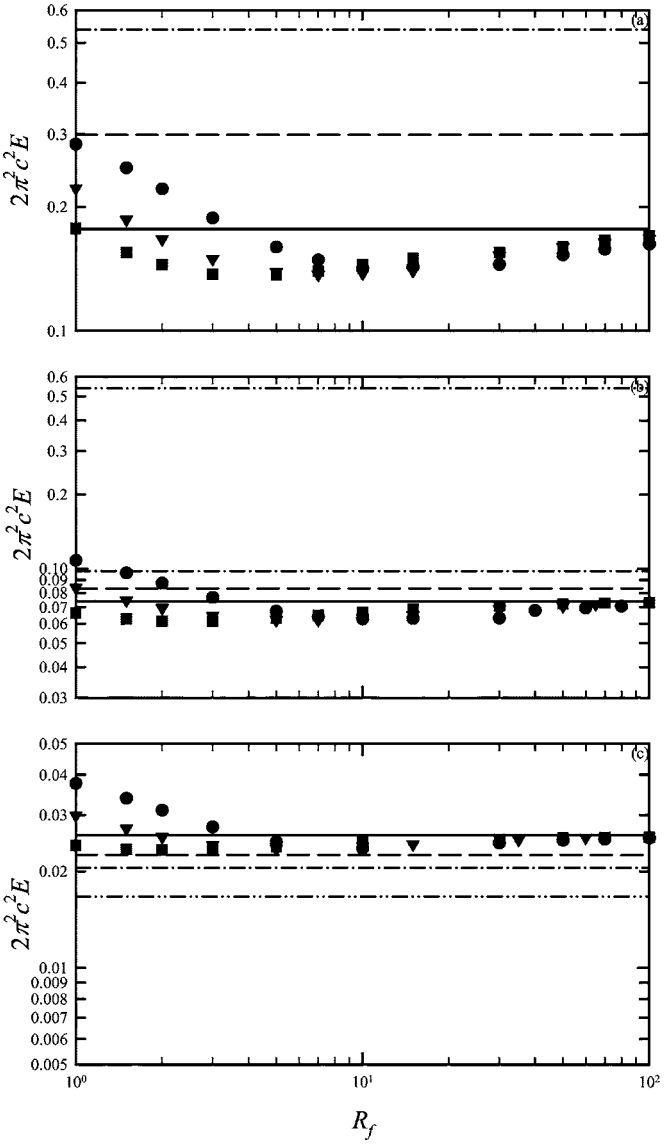


FIG. 6. Combined effects of effective fluid density, flow resistance, and initial vortex height on acoustical energy radiation: (a) $h=0.3$; (b) $h=0.5$; and (c) $h=0.8$. $-\cdot\cdot-$: P_x for $m=1.5$, $R_f=0$; $-\cdot-$: P_x for $m=3$, $R_f=0$; — — —: P_x for $m=5$, $R_f=0$; and ——: Rigid cylinder. ■: $m=1.5$; ▼: $m=3$; and •: $m=5$.

duction of a finite R_f has prolonged the duration of active sound radiation and has also resulted in the earlier radiation of sound. However, the magnitude of P_x is higher than the rigid cylinder condition for small R_f. The time variation of P_x converges to those for the rigid cylinder and perfectly inviscid fluid cases as $R_f\to\infty$ and 0, respectively. On the other hand, the introduction of the flow resistance enhances the radiation of P_y, though their magnitudes are small compared to those of P_x. The duration of the transverse dipole radiation appears longer than those of the longitudinal one. In addition, the magnitude of P_y is higher at some finite R_f. At very large R_f, the results converge to those in the rigid half-cylinder case.

Figures 5(a) to 5(c) summarize the effects of m and R_f on the amplitudes of P_x and P_y at $h=0.3$, 0.5, and 0.8 respectively for $1\leqslant R_f\leqslant 100$. The P_x for the cases of a rigid cylinder and an inviscid fluid are included for the sake of easy referencing. At an initial $h=0.3$, the amplitude of P_x is approximately equal to that of the rigid cylinder case for $R_f>1$ for all m studied [Fig. 5(a)]. The amplitude of P_y decreases as R_f is increased from 1 toward ∞ (for $R_f\to 0$ and ∞, $P_y=0$). The decrease of m increases the amplitudes of P_y for the whole range of R_f presented.

The increase of m leads to a reduction of the transverse dipole amplitude for the other two values of initial h for $R_f\geqslant 30$ [Figs. 5(b) and 5(c)]. The magnitude of the transverse dipole becomes weaker when the cylinder is less pressure releasing as anticipated by the theory (larger m and/or higher R_f). As can be expected, the increase in the initial vortex height h reduces the effects of the cylinder on the sound radiation. The results shown in Fig. 5 suggest that certain combinations of m and R_f will lead to louder sound radiation than the rigid cylinder case. Also, it is noted that the amplitudes of P_y are always below those of P_x, but the difference decreases with increasing initial h.

Figure 6 illustrates the overall acoustical energy (E) radiated by the unsteady vortex motions under the influence of m and R_f. At small initial h [Fig. 6(a)], the introduction of the porous material enhances the radiation of acoustical energy at $m=5$ and small R_f. This radiation becomes less important

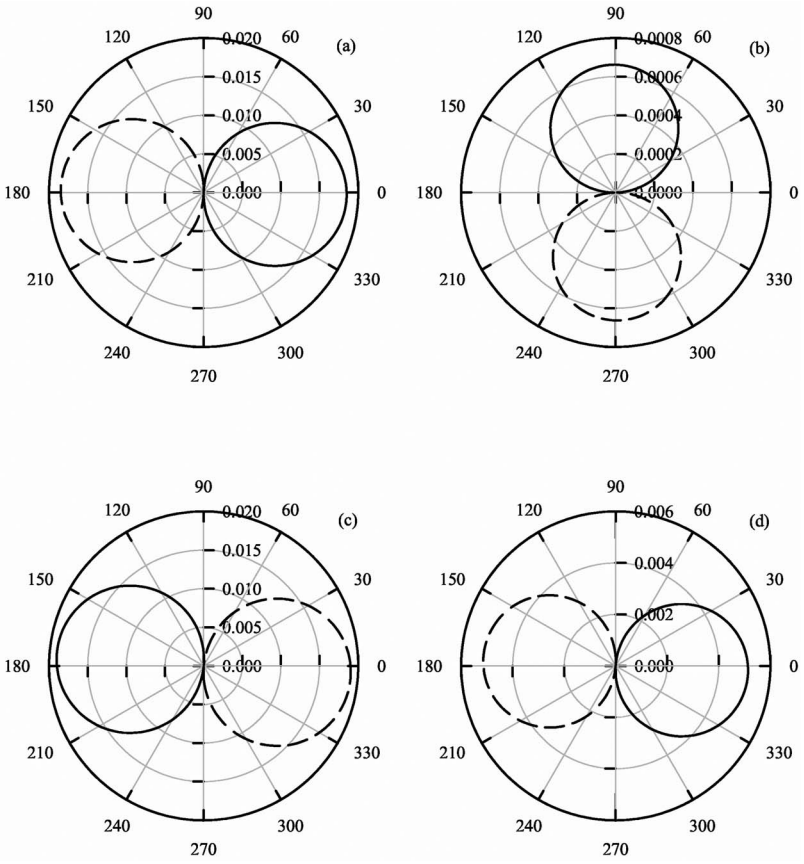


FIG. 7. Time variation of far-field directivity: (a) $t-t_a-\mathrm{R}/c=-13.55$; (b) $t-t_a-\mathrm{R}/c=-5.55$; (c) $t-t_a-\mathrm{R}/c=4.46$; and (d) $t-t_a-\mathrm{R}/c=24.89$. Initial $z_1=(-10,0.8)$, $m=1.5$ and $R_f=1.5$. — — —: negative sound pressure; ——: positive sound pressure.

as m decreases from 5 to 1.5. As R_f is increased, the strength of the energy radiation eventually falls below that in the rigid cylinder case for a fixed m. However, all the curves in Fig. 6(a) converge to $E=0.1769$, which is the energy radiated in the rigid cylinder case, for large R_f.

The situations at initial $h=0.5$, presented in Fig. 6(b), follow closely those shown in Fig. 6(a). The increase in the initial h reduces the induction effect of the cylinder, resulting in less significant sound radiation even at small m and R_f. Further increase in the initial h to 0.8 does not affect much the trend of E variation with R_f and m for $R_f>1$ [Fig. 6(c)]. In this case, less acoustical energy than in the rigid cylinder case is radiated at $R_f\rightarrow 0$ for all m.

Figures 7(a) to 7(d) show the change in the directivity patterns of the sound radiation. One can notice that the dipole axis does change with time as in Minota and Kambe,[28]

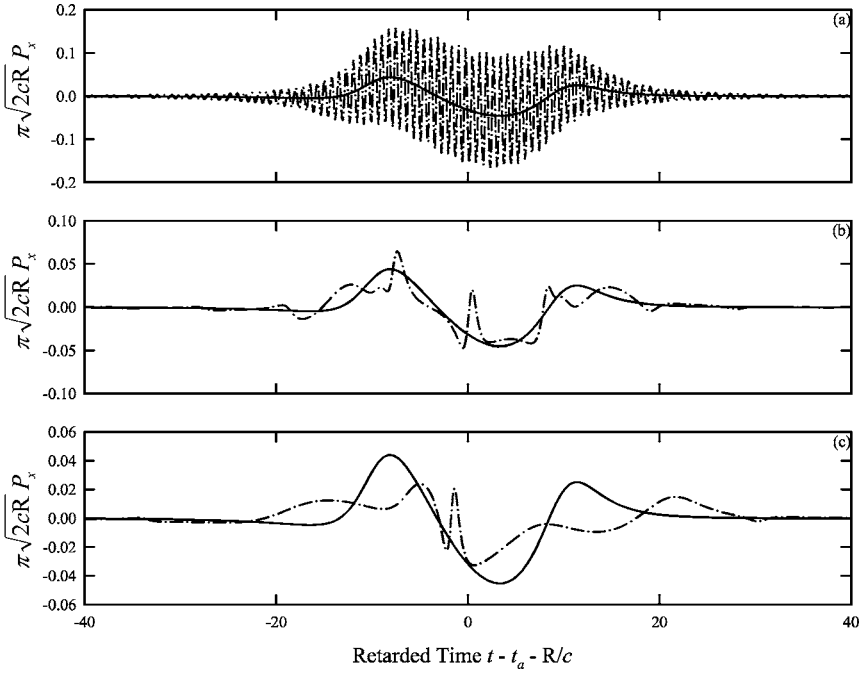


FIG. 9. Time variation of longitudinal dipole magnitude at different separation distance in the presence of a rigid cylinder: (a) $d=0.2$; (b) $d=0.8$; and (c) $d=1.6$. ——: Equivalent single vortex results; and $-\cdot-$: Two interacting identical vortices results. z_c initially at $(-10,0.5)$, $\Gamma_1=\Gamma_2=0.5$.

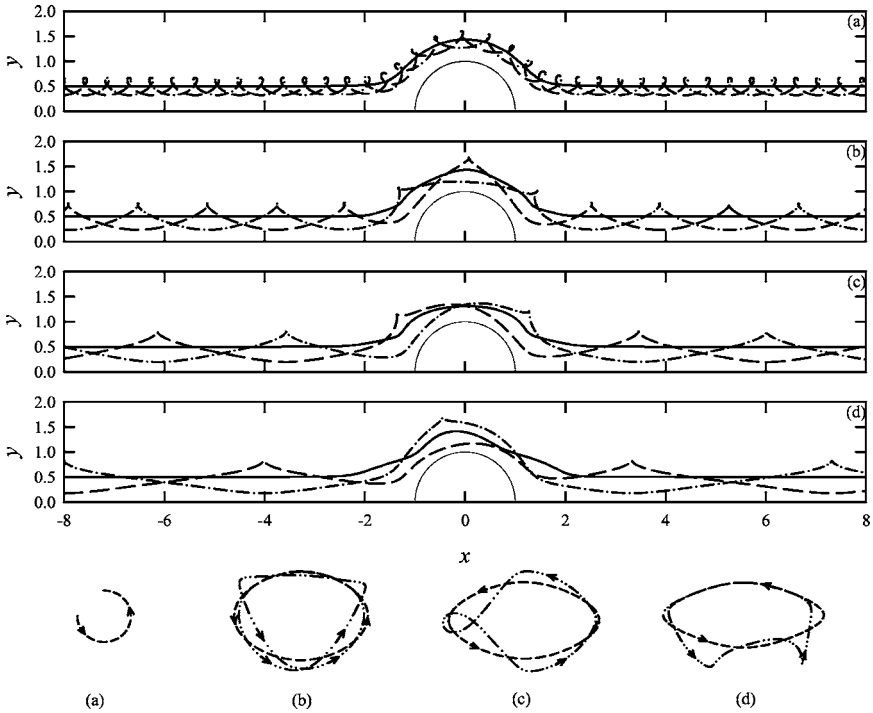


FIG. 8. Unsteady leapfrogging motions of two identical vortices near a rigid cylinder: (a) $d=0.4$; (b) $d=0.8$; (c) $d=1.2$; and (d) $d=1.6$. Initial $z_c=(-10,0.5)$, $\Gamma_1=\Gamma_2=0.5$. — — —: z_1; −·−: z_2; ——: z_c; - - - -: z_1 relative to z_c at $x<-2$ or $x>2$; and −··−: z_2 relative to z_c at $-2<x<2$.

but it should be noted that the longitudinal dipole dominates the sound field as the magnitude of P_x is nearly always higher than that of P_y (Fig. 5). The rotation of the dipole axis can only be observed when P_x is sufficiently small, which is also the instant of less significant sound radiation.

B. Two interacting identical vortices

The sound generation by two identical vortices will be examined in this section. The initial vertical height of the vortices h is set at 0.5 and the strengths of the two vortices are set equal at $\Gamma_1=\Gamma_2=0.5$. It is well known that two vortices of thin cores will undergo leapfrogging and such motion is periodic in the absence of the cylinder.[29] A quadrupole field is also created but its magnitude is too small when compared to the dipole fields in the present low Mach number vortex motion and thus is ignored. In the foregoing discussions, the vorticity centroid of the two vortices is defined as $z_c=(\Gamma_1 z_1+\Gamma_2 z_2)/\Gamma$, where $\Gamma=\Gamma_1+\Gamma_2$. Similar to the previous section, the results associated with the combinations of m and R_f under which the vortices come very close to the cylinder surface are excluded from the presentation.

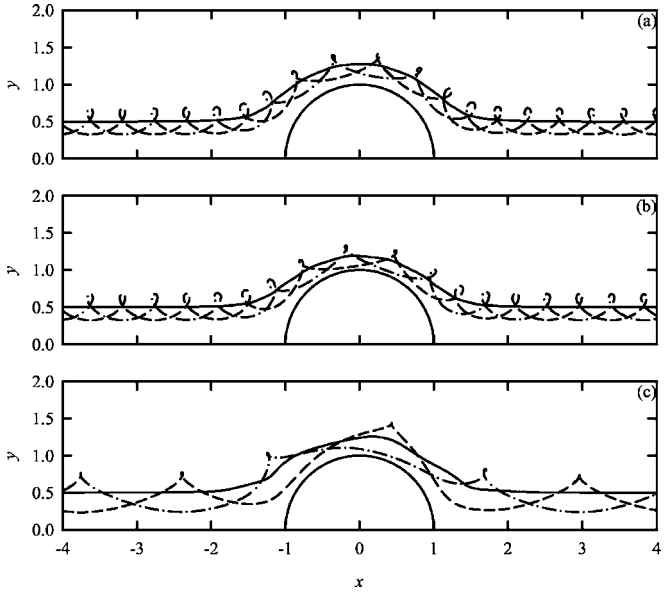


FIG. 10. Paths of two interacting vortices for perfectly inviscid fluid cases: (a) $d=0.4$, $m=5$; (b) $d=0.4$, $m=2$; and (c) $d=0.8$, $m=5$. z_c initially at $(-10,0.5)$, $\Gamma_1=\Gamma_2=0.5$. — — —: z_1; −·−: z_2; and ——: z_c.

Figure 8 illustrates some examples of the vortex paths at different d in the presence of a rigid cylinder. The paths of the individual vortices relative to z_c are also given at the bottom of the figure. For $d\leqslant 0.4$, the path of the vorticity centroid collapses with that of a single vortex of strength $\Gamma=1$ initially located at $x=-10$ with $h=0.5$. The paths of the vortices are circular relative to the vorticity centroid [Fig. 8(a)]. The presence of the cylinder does not much affect the mutual induction between the two vortices within this range of d.

At increased d, the path of z_c deviates from that shown in Fig. 8(a), and the paths of the two vortices relative to the vorticity centroid become chaotic and noncircular. The leapfrogging vortex motions become more disturbed as d increases from 0.8 to 1.6 [Figs. 8(b) to 8(d)]. The larger vortex separation weakens the mutual induction strengths between the vortices.

Figures 9(a) to 9(c) show the far-field sound pressure time fluctuations at different d. It is expected that the sound radiation is more significant when the vortices are in the proximity of the cylinder. The periodic leapfrogging vortex motions at small d result in a faster time fluctuating sound pressure [Fig. 9(a)], which carries most of the sound energy. There is a slowly varying component embedded inside the result shown in Fig. 9(a), which is similar to that produced by a single vortex of strength $\Gamma=1$ initially located at z_c

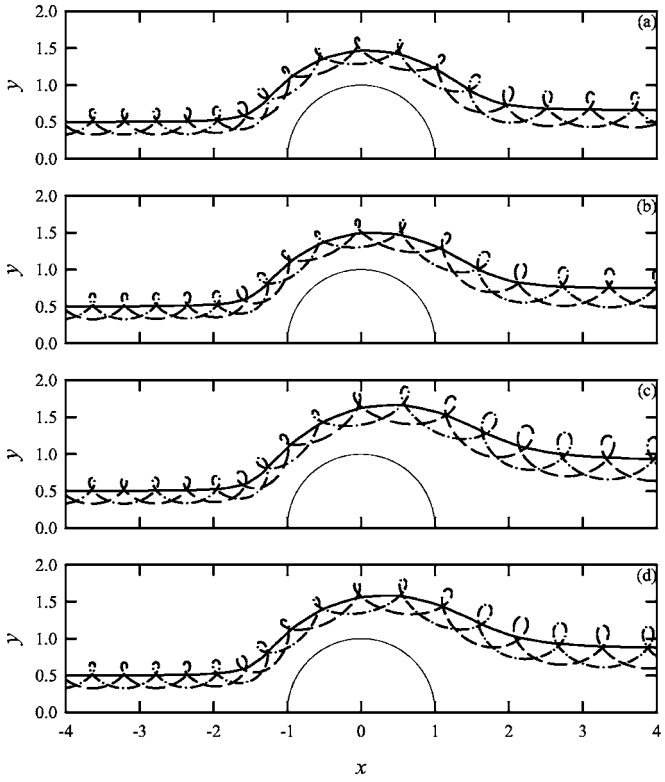


FIG. 11. Combined effects of effective fluid density and flow resistance on the vortex paths: (a) $m=5$, $R_f=10$; (b) $m=5$, $R_f=5$; (c) $m=5$, $R_f=1$; and (d) $m=3$, $R_f=1$. Initial $z_1=(-10.2,0.5)$, initial $z_2=(-9.8,0.5)$, $\Gamma_1=\Gamma_2=0.5$, $d=0.4$. — — —: z_1; –·–: z_2; and ——: z_c.

$=(-10,0.5)$. The increase in d leads to less ordered leapfrogging vortex motions. The pulses in Figs. 9(b) and 9(c) are created at the instant the vortex slip-through occurs as in the case without the cylinder.[30]

As discussed in Fig. 2(a), a finite effective fluid density inside the porous material lattice will create a pressure-releasing effect, reducing the effects of the cylinder relative to the mutual induction between the vortices. At $d=0.4$, ordered periodic vortex leapfrogging can be observed when $m=5$ and $R_f=0$ [Fig. 10(a)]. The reduction of m to 3 does not disturb much the leapfrogging vortex motions though the vortex paths are much closer to the cylinder surface [Fig. 10(b)]. The same is also true for $m=2$ (not shown here). The stronger effect from the cylinder due to shorter separation between it and the vortices does result in a slight deviation of the vortex paths relative to z_c from circular motion. The path of z_c resembles those shown in Fig. 2(a). Similar observations can be made at increased d [for instance, Fig. 10(c)] provided that the vortices do not hit the cylinder.

When the flow resistance inside the cylinder is finite, the vortex paths tend to bend away from the cylinder surface as the vortices propagate across the cylinder (Fig. 11) as in the single vortex case (Fig. 3). However, unlike the cases with a rigid cylinder or a perfectly inviscid fluid [Figs. 8 and 10, respectively], an increase in the pairing period is observed in the present two interacting vortices case upon the introduction of R_f. The separation of the vortices eventually increases due to the combined effect of m and R_f (Fig. 11). Decreasing m at a fixed R_f (for $R_f \geqslant 1$) brings the vortices closer to the x-axis after they pass over the cylinder into the region $x>2$ [Figs. 11(c) and 11(d)] and increases the frequency of the sound radiated. Such increase in sound frequency is more pronounced at small d. However, one should note that the acoustical energy radiated when the vortices are at $|x|>2$ is much less important. The dynamic at increased d is similar to those presented in Fig. 11, though one expects the two vortices will move closer to the cylinder provided that no impingement occurs. Thus, they are not presented.

Figures 12(a) and 12(b) illustrate the time variations of P_x and P_y at different R_f, respectively, at $d=0.4$ and $m=5$. Here, t_a represents the instant when the vorticity centroid passes over $x=0$. One can observe that there are fast and slow time varying components in P_x [Fig. 12(a)]. The former is due to the nominally circular motion of the vortices relative to z_c, whose frequency decreases after the vortices pass

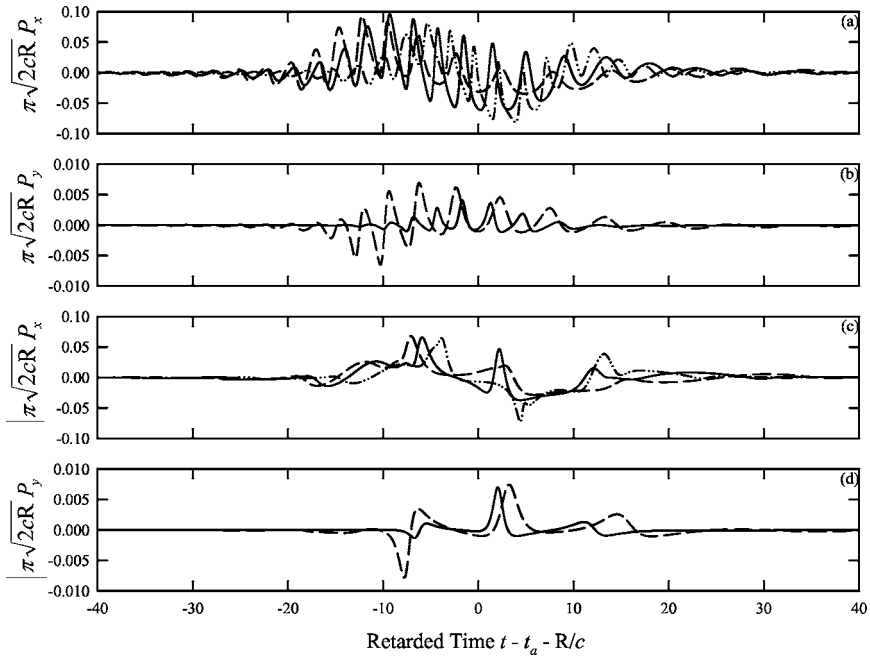


FIG. 12. Examples of time variation of dipole magnitudes at finite effective fluid density and flow resistance: (a) P_x, $d=0.4$; (b) P_y, $d=0.4$; (c) P_x, $d=0.8$; and (d) P_y, $d=0.8$. –···–: $R_f=0$; — — —: $R_f=1$; and ——: $R_f=10$. Initial $z_c=(-10,0.5)$, $\Gamma_1=\Gamma_2=0.5$.

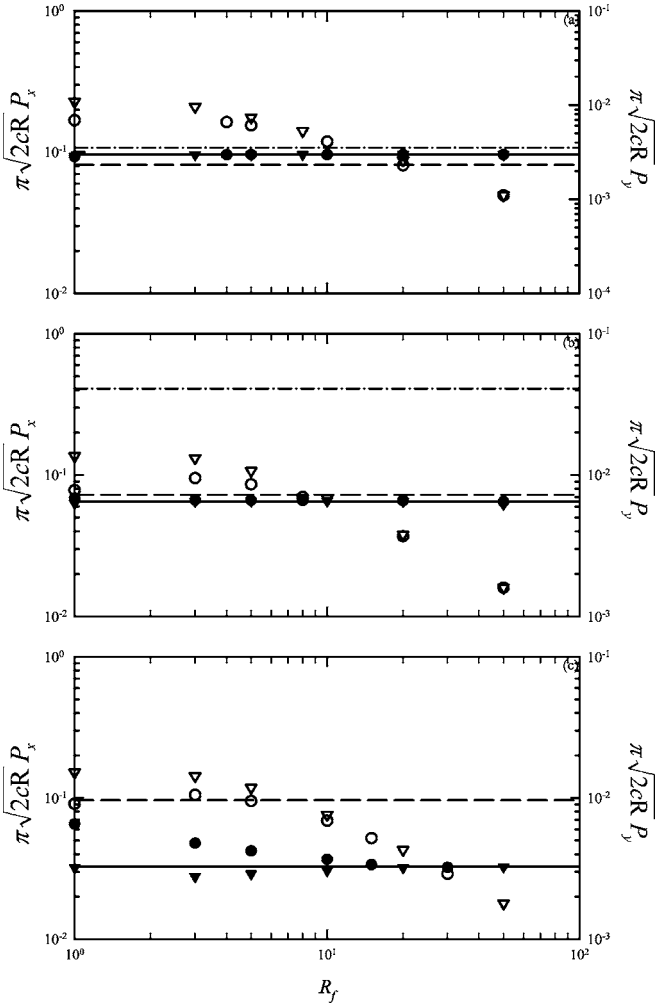


FIG. 13. Amplitudes of the dipoles produced by two interacting identical vortices: (a) $d=0.4$; (b) $d=0.8$; and (c) $d=1.6$. Initial $z_c=(-10,0.5)$, $\Gamma_1=\Gamma_2=0.5$. Legends: Same as those in Fig. 5.

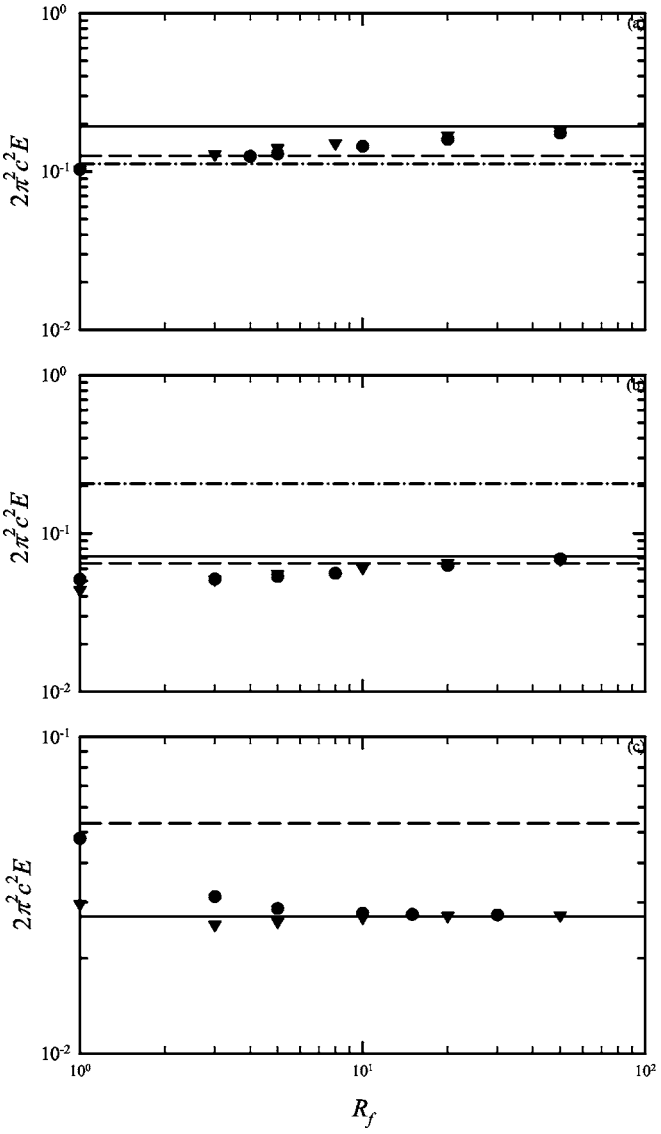


FIG. 14. Acoustical energy radiated by two interacting identical vortices: (a) $d=0.4$; (b) $d=0.8$; and (c) $d=1.6$. Initial $z_c=(-10,0.5)$, $\Gamma_1=\Gamma_2=0.5$. Legends: Same as those in Fig. 6.

over the cylinder. The strength of this component relative to the slow time varying one increases with increasing R_f for $R_f>1$. Similar time varying components are also found in P_y [Fig. 12(b)], but the amplitudes are very small when compared to those found in P_x. The amplitudes of these time varying components first increases with R_f but they decrease as R_f increases beyond unity. P_y vanishes when $R_f=0$ or $R_f\rightarrow\infty$.

The increase in the separation d to 0.8 reduces the mutual induction strength between the vortices, resulting in much less regular leapfrogging motions. The corresponding time variations of P_x and P_y with finite R_f are given in Figs. 12(c) and 12(d), respectively. The results for the hard cylinder at $d=0.8$ have been shown in Fig. 9(b). One can notice that the amplitudes of the two dipoles for $d=0.4$ and 0.8 do not differ much, but the higher-frequency fluctuations at $d=0.4$ imply more significant radiation of acoustical energy.

Figure 13 illustrates the dependence of the amplitudes of P_x and P_y on R_f, m and d. Again the amplitude of P_y is about a half- or a full-order below that of P_x. It is found that the introduction of the porous cylinder reduces in general the amplitude of the longitudinal dipole P_x for small d for $m\geqslant 3$ [Fig. 13(a)]. The vortices can be very close to the cylinder or even hit the cylinder when m drops below 3, making the whole vortex approach invalid, and the corresponding results are not presented.

The increase in d appears to have amplified P_x above its corresponding magnitude in rigid cylinder situation for $m=5$ [Figs. 13(b) and 13(c)]. The trend of P_x variation with d shown in Fig. 13 suggests louder noise will occur upon an increase of d. This implies that the presence of a porous material near a jet shear layer can be noisier than the case where the porous material is replaced by a rigid one, if the material is not located at a position where the dominant flow structures have a short wavelength (the initial shear layer mode).[31] Figure 14 further suggests that the porous material can reduce the overall acoustical energy radiation when d is small. It can also be noted that E decreases with decreasing m. The sound produced by the mutual interaction of the vortices depends very much on the unsteady leapfrogging motions. The smaller the value of d, the higher the frequency of the radiated sound, and thus weaker sound radiation can be

TABLE I. Ranges of parameters in practical and experimental flows.

Parameter	Typical range	Normalized range
Cylinder radius (cm)[a]	10–15	1
Vortex spacing (cm)[b]	0.6–10.2	0.04–1.02
Vortex height (cm)[c]	0–6.7	0–0.67
Flow speed (m/s)	0–40	0–43
Vortex circulation (m^2/s)[d]	0.14–3.4	1
Effective fluid density[e]	1–3	1–3
Flow resistance (kg/m^3s)[e,f]	2000–75 000	4.9–10 000

[a]See Ref. 24.
[b]See Ref. 25.
[c]See Refs. 34–36.
[d]See Refs. 26 and 33.
[e]See Refs. 19, 20, and 23.
[f]See Ref. 27.

expected because of the sound absorption property of the porous material. The effects of initial h in this two interacting vortices case are similar to those observed in the single vortex case and thus are not discussed further.

An external mean flow is absent in the present study. However, its presence is not expected to alter the sound generation process. On the contrary, the mean flow will result in a vortex path closer to the cylinder and the associated higher rate of change of vortex velocity will probably create a stronger sound field. The convective amplification of sound as in the case of Ffowcs Williams and Lovely[32] is also anticipated. The present results therefore represent the weakest sound radiation of the topic.

IV. REMARKS

The low Mach number condition in the present study results in the radiation of low frequency sound whose peak value depends substantially on the vortex circulation. At $R=100$ with an ambient speed of sound $c=343$ (normalized by Γ/a), $h=0.3$ and $m=1.5$, the maximum peak normalized sound pressure radiated at $R_f=1$ by a single vortex is 1.3×10^{-4} [Fig. 5(a)]. With a $\Gamma=0.14$ m^2/s, the maximum sound pressure level is around 23.5 dB, but this pressure level goes up to ~78.9 dB when $\Gamma=3.4$ m^2/s. For a rigid half-cylinder, the corresponding maximum sound pressure levels are 21 dB and 72 dB, respectively.

In the case of two vortices with $h=0.5$, $d=0.8$, $m=5$, and $R_f=1$, the sound pressure levels with $\Gamma=0.14$ m^2/s and 3.4 m^2/s are approximately 19.8 dB and 75.2 dB, respectively [Fig. 13(b)]. The corresponding values for rigid half cylinder are 19.3 dB and 74.7 dB, respectively.

The above dimensional examples illustrate that the aeroacoustics studied in the present work can be significant, and the introduction of porous material can enhance the sound radiation at certain combinations of parameters.

V. CONCLUSIONS

In the present investigation, the sound generation by the unsteady motions of vortices in the presence of a porous half cylinder on an otherwise rigid horizontal plane is studied theoretically. The far-field sound pressure so produced is evaluated through the use of conformal mapping and matched asymptotic method. The effects of the effective fluid density and flow resistance inside the porous materials on the vortex motions and the far-field sound radiation are discussed.

In the presence of a porous material with a finite flow resistance, longitudinal and transverse dipoles coexist in the far field but the latter is significantly weaker than the former in general. When a single vortex engages the porous half-cylinder, the time variation of the strength of each dipole is pulselike. The larger vortex height above the rigid plane reduces the amplitudes of the dipoles. However, the overall acoustical energy radiated remains higher than that in the rigid cylinder case at some combinations of the effective fluid density and flow resistance.

When two identical vortices exist in the proximity of the porous cylinder, both the longitudinal and transverse dipoles consist of a rapidly and a slowly varying frequency components. The former is due to the leapfrogging motions of the vortices, and the latter to the macroscopic vortex centroid motions. The overall acoustical energy so radiated is less than that in the rigid cylinder case when the two vortices are close to each other and the dipoles are dominated by the fast variation fluctuations. The opposite is observed at larger vortex separation.

The present results show that suitable combinations of the effective fluid density and the flow resistance within a porous material will enhance the radiation of sound in the presence of a turbulent shear flow, especially when the flow structures involved are of slow variation frequency.

ACKNOWLEDGMENT

This study was supported by a grant from the Research Grant Council, The Hong Kong Special Administration Region, People's Republic of China (Project No. PolyU5030/00E).

APPENDIX: PARAMETER RANGES FOR EXPERIMENTAL FLOWS

This appendix discusses the ranges of the main parameters of the present study in real flows. The circulation of the vortex in real flow cannot be easily found, but it is believed that such circulation can be approximated by the formula, $\Gamma=(4/\pi)Uh$, shown in Saffman,[33] where U and h are the mean flow speed and the vortex height, respectively. The latter is expected to be close to the turbulent boundary layer thickness δ in a real flow as concluded by Praturi and Brodkey[34] and Robinson.[35] According to Schlichting and Gersten,[36] δ is related to a Reynolds number Re_l as

$$\frac{U\delta}{\nu}=0.14\frac{\mathrm{Re}_l}{\log_e(\mathrm{Re}_l)}G(\mathrm{Re}_l), \tag{A1}$$

where $\mathrm{Re}_l=Ul/\nu$ (l is the distance from the leading edge of the boundary layer and ν the kinematic viscosity of air) and G is a function whose value depends on Re_x. $G\approx1.5$ for $10^5<\mathrm{Re}_l<10^6$ and is equal to 1 as $\mathrm{Re}_l\to\infty$.[36]

Inside a building, the location of silencer has to be far away from flow discontinuity and bends so that the flow is fully developed before it reaches the silencer. It is not diffi-

cult to find $l=10$ m with $U=40$ m/s, giving a $\mathrm{Re}_l \sim 2\times 10^7$. Assume $G\sim 1$, $\delta=6.7$ cm giving $\Gamma=(4/\pi)U\delta=3.4$ m^2/s. Table I summarizes the parameter ranges in practical and experimental flows observed, where use has been made of Refs. 19, 20, and 23 to 27.

One should also note that the fire regulation requires fire dampers to be installed in front of a silencer at the air outlet of the air handling unit.[37] The vortex strength in such a case can be large as the turbulent flow may not be parallel to the damper blades, causing R_f to drop below unity. In the present study, the analysis is focused on the range $R_f \geqslant 1$ and $R_f=0$ (the inviscid fluid case).

[1] M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley, New York, 1987).
[2] M. J. Lighthill, "On sound generated aerodynamically I. General theory," Proc. R. Soc. London, Ser. A **211**, 564–587 (1952).
[3] N. Curle, "The influence of solid boundaries upon aerodynamic sound," Proc. R. Soc. London, Ser. A **231**, 505–514 (1955).
[4] L. L. Beranek and I. L. Vér, *Noise and Vibration Control Engineering, Principles and Applications* (Wiley, New York, 1992).
[5] M. S. Howe, *Theory of Vortex Sound* (Cambridge University Press, Cambridge, UK, 2003).
[6] A. Cummings and I. J. Chang, "Sound attenuation of a finite length dissipative flow duct silencer with internal mean flow in the absorbent," J. Sound Vib. **127**, 1–17 (1988).
[7] M. C. Quinn and M. S. Howe, "On the production and absorption of sound by lossless liners in the presence of mean flow," J. Sound Vib. **97**, 1–9 (1984).
[8] K. S. Peat and K. L. Rathi, "A finite element analysis of the convected acoustic wave motion in dissipative silencers," J. Sound Vib. **184**, 529–545 (1995).
[9] J. E. Ffowcs Williams, "The acoustics of turbulence near sound-absorbent liners," J. Fluid Mech. **51**, 737–749 (1972).
[10] S. K. Tang and C. K. Lau, "Vortex sound in the presence of a wedge with inhomogeneous surface flow impedance," J. Sound Vib. **281**, 1077–1091 (2005).
[11] D. G. Crighton, "Radiation from vortex filament motion near a half plane," J. Fluid Mech. **51**, 357–362 (1972).
[12] S. K. Tang, "Effects of porous boundaries on the dynamics of an inviscid vortex filament," Q. J. Mech. Appl. Math. **51**, 65–84 (2001).
[13] F. Obermeier, "New representation of aeroacoustic source distribution. 2. 2-dimensional model flows," Acustica **42**, 62–71 (1979).
[14] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968).
[15] J. B. Schneider, C. L. Wagner, and R. J. Kruhlak, "Simple conformal methods for finite-difference time-domain modeling of pressure-release surfaces," J. Acoust. Soc. Am. **104**, 3219–3226 (1998).
[16] R. V. Churchill and J. W. Brown, *Complex Variables and Applications* (McGraw–Hill, New York, 1990).
[17] E. J. Routh, "Some applications of conjugate functions," Proc. London Math. Soc. 12, 73–89 (1881).
[18] I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, *Tables of Integrals, Series and Products* (Academic, Boston, 1994).
[19] A. Cummings and N. Sormaz, "Acoustic attenuation in dissipative splitter silencers containing mean fluid flow," J. Sound Vib. **168**, 209–227 (1993).
[20] R. Kirby, "Simplified techniques for predicting the transmission loss of a circular dissipative silencer," J. Sound Vib. **243**, 403–426 (2001).
[21] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Wiley, New York, 1972).
[22] H. Abou-Hussein, A. DeBenedictis, N. Harrison, M. Kim, M. A. Rodrigues, F. Zagadou, and M. S. Howe, "Vortex-surface interaction noise: A compendium of worked examples," J. Sound Vib. **252**, 883–918 (2002).
[23] M. E. Delany and E. N. Bazley, "Acoustical properties of fibrous absorbent materials," Appl. Acoust. **3**, 105–116 (1970).
[24] *Rectangular and Cross-Talk Silencers* (Acoustic Engineering Services, Ltd., Surrey, UK).
[25] J. Jimenez, "A spanwise structure in the plane shear layer," J. Fluid Mech. **132**, 319–336 (1983).
[26] K. B. M. Q. Zaman and A. K. M. F. Hussain, "Vortex pairing in a circular jet under controlled excitation. 1. General jet response," J. Fluid Mech. **101**, 449–492 (1980).
[27] Y. Aurégan, A. Debray, and R. Starobinski, "Low-frequency sound propagation in a cylindrical duct: Application to sudden area expansions and to dissipative silencers," J. Sound Vib. **243**, 461–473 (2001).
[28] T. Minota and T. Kambe, "Acoustic-waves emitted by a vortex ring passing near a circular cylinder," J. Sound Vib. **119**, 509–528 (1987).
[29] S. K. Tang and N. W. M. Ko, "Sound sources in the interactions of two inviscid two-dimensional vortex pairs," J. Fluid Mech. **419**, 177–201 (2000).
[30] S. K. Tang and N. W. M. Ko, "Basic sound generation mechanisms in inviscid vortex interactions at low Mach number," J. Sound Vib. **262**, 87–115 (2003).
[31] A. K. M. F. Hussain and K. B. M. Q. Zaman, "An experimental study of organized motions in the turbulent plane mixing layer," J. Fluid Mech. **159**, 85–104 (1985).
[32] J. E. Ffowcs Williams and D. J. Lovely, "Sound radiation into uniformly flowing fluid by compact surface vibration," J. Fluid Mech. **71**, 689–700 (1975).
[33] P. G. Saffman, *Vortex Dynamics* (Cambridge University Press, Cambridge, UK, 1992).
[34] A. K. Praturi and R. S. Brodkey, "A stereoscopic visual study of coherent structures in a turbulent shear flow," J. Fluid Mech. **89**, 251–272 (1978).
[35] S. K. Robinson, "Coherent motions in the turbulent boundary layer," Annu. Rev. Fluid Mech. **23**, 601–639 (1991).
[36] H. Schlichting and K. Gersten, *Boundary Layer Theory* (Springer, London, 2000).
[37] A. E. Cote, *Fire Protection Handbook*, 18th Ed. (National Fire Protection Association, Qunicy, MA1997).

Calculation of amplitudes of acoustic normal modes from the reciprocity principle

Oleg A. Godin[a)]
CIRES, University of Colorado and NOAA/Earth System Research Laboratory, DSRC, Mail Code R/PSD, 325 Broadway, Boulder, Colorado 80305-3328

(Received 18 March 2005; revised 3 November 2005; accepted 27 December 2005)

Recently, J. D. Achenbach [J. Acoust. Soc. Am. **116**, 1481–1487 (2004)] put forward, on heuristic grounds, an elegant technique to calculate amplitudes of guided waves in an elastic solid. The technique is based on application of the reciprocity principle. In this article, the technique is applied to acoustic waveguides. A mathematical justification of the technique is obtained. The technique is extended to enable calculation of excitation coefficients of modes of both discrete and continuous spectra by a given sound source. The results are shown to be identical to those derived with the traditional methods that rely on integral transforms. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2167148]

PACS number(s): 43.30.Bp, 43.20.Bi, 43.20.Mv [WPS] Pages: 2096–2100

I. INTRODUCTION

Computation of a continuous wave (CW) field in a waveguide in a layered medium is a classical problem. It involves calculation of normal modes of the waveguide, which are determined by the medium's stratification and wave frequency as well as calculation of amplitudes of the proper normal modes and the continuous spectrum excited by a given source. The problem is traditionally solved[1] by use of integral transforms with subsequent integration in the complex plane needed to separate discrete and continuous spectra of the field. For waveguides with a purely discrete spectrum, a simpler approach[2] is often used that exploits orthogonality of normal modes.

Recently, J. D. Achenbach proposed[3] and applied[4–6] to a number of elastodynamic problems a technique which allows one to calculate amplitudes of guided waves from the reciprocity principle and avoid using integral transforms. Achenbach's approach relies on an implicit assumption that one can disregard a continuous spectrum when calculating amplitudes of modes in the discrete spectrum. The question was raised[5] whether results obtained with this technique are exact or approximate. In this article, this technique is applied to acoustic waveguides in a layered fluid. By taking into account the continuous spectrum, Achenbach's approach is put on mathematically solid footing and demonstrated to be exact.

CW acoustic pressure p satisfies the reduced wave equation[1]

$$\boldsymbol{\nabla}\left(\frac{\boldsymbol{\nabla} p}{\rho}\right)+\frac{k^2}{\rho}p=i\omega A+\boldsymbol{\nabla}\left(\frac{\mathbf{F}}{\rho}\right), \tag{1}$$

where $k=\omega/c$; ω, c, ρ, A, and $\mathbf{F}$ are, respectively, wave frequency, sound speed, mass density, and volume densities of sources of volume velocity and force that generate sound. Time dependence $\exp(-i\omega t)$ of the sound sources and the resulting wave fields are assumed and suppressed throughout the article. Consider two sets of sources, $\{A^{(j)},\mathbf{F}^{(j)}\}$, $j=1, 2$, and the acoustic fields $p^{(j)}$ that the sources generate in the same medium. By multiplying Eq. (1) for $p^{(1)}$ by $p^{(2)}$ and Eq. (1) for $p^{(2)}$ by $p^{(1)}$ and subtracting the results, after integration over a spatial domain Ω with a boundary $\partial\Omega$, one obtains the identity (Ref. 1, p. 130)

$$\int_\Omega d^3\mathbf{x}[A^{(2)}p^{(1)}-\mathbf{F}^{(2)}\cdot\mathbf{v}^{(1)}-A^{(1)}p^{(2)}+\mathbf{F}^{(1)}\cdot\mathbf{v}^{(2)}]$$

$$=\int_{\partial\Omega} ds\mathbf{N}\cdot\mathbf{j}, \tag{2}$$

$$\mathbf{j}=(i\omega\rho)^{-1}[p^{(1)}\boldsymbol{\nabla}p^{(2)}-p^{(2)}\boldsymbol{\nabla}p^{(1)}]=p^{(1)}\mathbf{v}^{(2)}-p^{(2)}\mathbf{v}^{(1)}, \tag{3}$$

which is a mathematical expression of the acoustic reciprocity principle. Here, $\mathbf{N}$ is an external unit normal to $\partial\Omega$ and $\mathbf{v}=(i\omega\rho)^{-1}\boldsymbol{\nabla}p$ is the particle velocity in the sound wave.

In Sec. II we apply the identity Eq. (2) to calculate amplitudes of modes excited by a linear source in a waveguide in a fluid half-space $z>0$ with pressure-release boundary $z=0$. Parameters of the fluid are taken to be piecewise differentiable functions of vertical coordinate z. For simplicity, it is assumed that at $z>H$, sound speed and density take constant values $c=c_b$ and $\rho=\rho_b$. In Sec. III, mode excitation by point sources is considered. In Sec. IV, the results obtained from the reciprocity principle are compared to those derived from integral transforms. Our findings are summarized in Sec. V.

II. LINEAR SOURCE IN A LAYERED HALF-SPACE

Consider an acoustic field generated by a source of volume velocity uniformly distributed along a horizontal line $\{x=x_0, z=z_0\}$. For such a source, $A=A_0^{(1)}\delta(x-x_0)\cdot\delta(z-z_0)$ and $\mathbf{F}=0$ in Eq. (1). Like the source term in the reduced wave equation, the acoustic pressure is an even function of $x-x_0$ and can be written as (Ref. 1, pp. 151–153; Ref. 7)

a)Electronic mail: oleg.godin@noaa.gov

 0001-4966/2006/119(4)/2096/5/$22.50

$$p^{(1)} = \sum_n E_n f_n(z) \exp(i\xi_n|x-x_0|) + \int_0^{+\infty} E(\xi) f(\xi,z) \exp(i\xi|x-x_0|) \frac{d\xi}{d\mu} d\mu. \quad (4)$$

Here, $\mu(\xi)=\sqrt{\omega^2/c_b^2-\xi^2}$; ξ_n are propagation constants of proper normal modes, Re $\xi_n \geqslant 0$, Im $\xi_n \geqslant 0$; $f(\xi,z)$ is a solution of the one-dimensional reduced wave equation

$$\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial f(\xi,z)}{\partial z}\right) + \frac{k^2(z)-\xi^2}{\rho} f(\xi,z) = 0, \quad (5)$$

and $f_n(z)$ is a solution to Eq. (5) at $\xi=\xi_n$. The mode shape functions $f_n(z)$ and $f(\xi,z)$ satisfy the boundary condition $f(\xi,0)=0$ on the pressure release surface. For proper normal modes, Im $\mu>0$, and $f_n(z)\to 0$ when $z\to+\infty$. At $\mu>0$, solutions to Eq. (5) are sinusoidal at $z>H$ and remain finite when $z\to+\infty$. The sum of proper modes in the right side of Eq. (4) represents the discrete spectrum of the field. The integral over μ represents the continuous spectrum. Components of the acoustic field with the vertical dependence given by $f(\xi,z)$ are usually referred to as improper (or radiation) modes. The physical meaning of the continuous spectrum is especially clear when source and receiver are located at z_0, $z>H$. Then, the integral in Eq. (4) represents superposition of incident and reflected homogeneous plane waves. When μ increases from 0 to $k_b \equiv \omega/c_b$ and then to infinity, ξ varies from k_b to 0 and then becomes purely imaginary, with Im ξ increasing from 0 to infinity. Waves with the imaginary ξ in the continuous spectrum in Eq. (4) rapidly attenuate when the horizontal distance from the source increases.

A mathematical proof that the set of all proper and improper modes is complete and that the acoustic field can be represented by Eq. (4) in an arbitrarily stratified, unbounded medium can be found in Wilcox.[8] A simple reasoning shows that this expansion theorem applies also to the semi-infinite waveguide we consider. Indeed, the field due to a linear source in a layered half-space with a pressure-release (or rigid) boundary at $z=0$ equals a sum of fields of two linear sources (the original source and its mirror image) in an infinite, symmetric layered medium where sound speed and density are even functions of z. The desired modal expansion of the field in a half-space is obtained as a sum of the two modal expansions in an infinite medium. (Relations between modes in semi-infinite and corresponding infinite media are discussed in Ref. 8, pp. 121–122.)

Shape functions of normal modes are determined by medium's stratification and are supposed to be known. Our goal is to calculate excitation coefficients E_n and $E(\xi)$ of proper and improper modes. Below, we will use the orthogonality property of modes. Consider mode shape functions corresponding to two different values of ξ. Similar to the derivation of Eq. (2) from Eq. (1), we obtain from Eq. (5)

$$(\xi_a^2-\xi_b^2)\int_0^z \frac{dz'}{\rho(z')} f(\xi_a,z') f(\xi_b,z') = \frac{1}{\rho(z)}\left[f(\xi_a,z)\frac{\partial f(\xi_b,z)}{\partial z} - f(\xi_b,z)\frac{\partial f(\xi_a,z)}{\partial z}\right]. \quad (6)$$

When at least either ξ_a or ξ_b is a propagation constant of a proper mode, the right side of Eq. (6) tends to zero at $z\to+\infty$. Hence, shape functions of proper and improper modes are orthogonal with the weight ρ^{-1}:

$$\int_0^\infty \frac{dz}{\rho(z)} f_n(z) f(\xi,z) = 0, \quad (7)$$

and the proper modes can be normalized by the condition

$$\int_0^\infty \frac{dz}{\rho(z)} f_n(z) f_m(z) = \delta_{nm}. \quad (8)$$

The improper modes can be normalized by requiring that $f(\xi,z)=\cos[\mu z+\theta(\mu)]$ at $z>H$ for some $\theta(\mu)$. Then from Eq. (6) and the relation[7] $\lim_{z\to\infty} \mu^{-1}\sin(\mu z+\theta) = \pi\cos\theta\,\delta(\mu)$, it follows that

$$\int_0^\infty \frac{dz}{\rho(z)} f(\xi',z) f(\xi'',z) = \frac{-\pi}{2\rho_b}\delta[\mu(\xi')-\mu(\xi'')]. \quad (9)$$

Following Achenbach,[4] consider the reciprocity relation Eq. (2) for the field $p^{(1)}$ Eq. (4) and an auxiliary field

$$p^{(2)}(x,z) = f_m(z)\exp(i\xi_m x), \quad (10)$$

which satisfies the reduced wave Eq. (1) with $A=0$, $\mathbf{F}=0$. It is convenient to choose the spatial domain Ω as follows: $\Omega=\{-l<x-x_0<l, 0<y<L, 0<z<+\infty\}$. Because of the boundary conditions at $z=0$ and the fields' independence of the coordinate y, Eq. (2) becomes

$$\begin{aligned} -A_0^{(1)} f_m(z_0) e^{i\xi_m x_0} &= \int_0^\infty \frac{dz}{\omega\rho} f_m(z) \times\left(\xi_m e^{i\xi_m x} p^{(1)} + i e^{i\xi_m x}\frac{\partial p^{(1)}}{\partial x}\right)\Bigg|_{x=x_0-l}^{x=x_0+l} \\ &= \sum_n E_n\, \varphi(\xi_n,x)\Big|_{x_0-l}^{x_0+l} \int_0^\infty \frac{dz}{\omega\rho} f_m(z) f_n(z) \\ &\quad + \int_0^{+\infty} d\mu \frac{d\xi}{d\mu} E(\xi)\, \varphi(\xi,x)\Big|_{x_0-l}^{x_0+l} \times \int_0^\infty \frac{dz}{\omega\rho} f_m(z) f(\xi,z), \end{aligned} \quad (11)$$

where

$$\varphi(\xi,x) = \left(\xi_m - \xi\frac{x-x_0}{|x-x_0|}\right)\exp(i\xi_m x + i\xi|x-x_0|). \quad (12)$$

Because of the orthogonality relations Eqs. (7) and (8), the right-most side in Eq. (11) vanishes except for the term with $n=m$ in the sum, and we obtain from Eq. (11) the excitation coefficients of the m-th proper normal mode

$$E_m = \frac{\omega}{2\xi_m} A_0^{(1)} f_m(z_0). \tag{13}$$

As expected, the result is independent of l and L. Note that the only nonzero contribution to the integral over the boundary $\partial\Omega$ comes from $x=x_0-l$, where $p^{(2)}$ and the waves generated by the source propagate in opposite directions.

Amplitudes of waves of the continuous spectrum are found in a similar way by using an auxiliary field

$$p^{(2)} = \int_0^{+\infty} \tilde{E}(\xi) f(\xi,z) \exp(i\xi x) \frac{d\xi}{d\mu} d\mu, \tag{14}$$

where $\tilde{E}(\xi) \neq 0$ is a known function, instead of the field Eq. (10). [The function $\tilde{E}(\xi)$ should be chosen in such a way that the integral in Eq. (14) converges throughout the spatial domain Ω. Aside from that, the choice of $\tilde{E}(\xi)$ is arbitrary.] Substituting $p^{(1)}$ (4) and $p^{(2)}$ (14) into the reciprocity relation Eq. (2) and using the orthogonality relations Eqs. (7) and (9), we find that

$$E(\xi) = \frac{-\omega\rho_b}{\pi\mu(\xi)} A_0^{(1)} f(\xi,z_0). \tag{15}$$

For a linear source of vertical external force, where $A^{(1)}=0$, $\mathbf{F}^{(1)}=\mathbf{F}_0^{(1)}\delta(x-x_0)\ \delta(z-z_0)$, and $\mathbf{F}_0^{(1)}=(0,0,F_{03}^{(1)})$, determination of the mode excitation coefficients differs only by calculation of the left side in Eq. (2). For a linear source of horizontal external force, where $\mathbf{F}_0^{(1)}=(F_{01}^{(1)},0,0)$, one also must take into account that the source term in Eq. (1) and, hence, acoustic pressure are odd functions of $x-x_0$. A corresponding representation of the field in terms of modes differs from Eq. (4) by an additional factor $(x-x_0)/|x-x_0|$. Repeating the transformations presented above for a source of volume velocity, for the acoustic pressure radiated by a linear source of external force, one obtains the following mode amplitudes in Eq. (4):

$$E_m = \frac{1}{2\rho(z_0)} \left(F_{01} \frac{x-x_0}{|x-x_0|} f_m(z_0) + \frac{iF_{03}}{\xi_m} \frac{\partial f_m(z_0)}{\partial z_0} \right), \tag{16}$$

$$E(\xi) = \frac{-\rho_b}{\pi\mu(\xi)\rho(z_0)} \left(\xi F_{01} \frac{x-x_0}{|x-x_0|} f(\xi,z_0) + iF_{03} \frac{\partial f(\xi,z_0)}{\partial z_0} \right).$$

III. POINT SOURCE IN A LAYERED HALF-SPACE

Let a point source of volume velocity be located at a point $(0,0,z_0)$. For such a source, $A=A_0^{(1)}\delta(x)\delta(y)\delta(z-z_0)$ and $\mathbf{F}=0$ in Eq. (1). Acoustic pressure radiated by the source is axially symmetric and may be represented as (Ref. 1, p. 151; Ref. 2, pp. 271–273) [cf. Eq. (4)]

$$p^{(1)} = \sum_n E_n f_n(z) H_0^{(1)}(\xi_n r) + \int_0^{+\infty} E(\xi) f(\xi,z) H_0^{(1)}(\xi r) \frac{d\xi}{d\mu} d\mu, \tag{17}$$

where $r=(x^2+y^2)^{1/2}$, $H_\nu^{(1)}(u)=J_\nu(u)+iY_\nu(u)$, and J_ν, Y_ν, $H_\nu^{(1)}$ are the Bessel functions of order ν and of the first, second, and third kind, respectively. To determine excitation coefficients of the proper normal modes in Eq. (17), it is convenient to introduce an auxiliary, axially symmetric wave

$$p^{(2)} = f_m(z) J_0(\xi_m r) \tag{18}$$

and to consider the reciprocity relation Eq. (2) in a cylindrical domain $\Omega=\{0\leqslant r<R, 0<z<+\infty\}$. Using mode orthogonality relations Eqs. (7) and (8), from Eq. (2) we obtain

$$-A_0^{(1)} f_m(z_0) = \frac{2\pi}{i\omega} E_m \xi_m R \left[H_0^{(1)}(u) \frac{d}{du} J_0(u) - J_0(u) \frac{d}{du} H_0^{(1)}(u) \right] \Bigg|_{u=\xi_m R}. \tag{19}$$

When integrating over $\partial\Omega$ in Eq. (2), the only nonzero contribution to the right side of Eq. (19) comes from the cylindrical surface $r=R$. Using the identity[9] $J_\nu(u)Y_\nu'(u) - J_\nu'(u)Y_\nu(u) = 2/\pi u$, we find from Eq. (19) that

$$E_m = \frac{\omega}{4} A_0^{(1)} f_m(z_0). \tag{20}$$

It is clear from the derivation that the result, Eq. (20), is independent of the choice of the auxiliary field $p^{(2)}$ as long as $p^{(2)}$ includes a single, proper normal mode propagating toward the source and is not singular in the vicinity of the source.

To calculate amplitudes of improper modes in Eq. (17), we now choose the auxiliary field as [cf. Eq. (14)]

$$p^{(2)} = \int_0^{+\infty} \tilde{E}(\xi) f(\xi,z) J_0(\xi r) \frac{d\xi}{d\mu} d\mu. \tag{21}$$

Subsequent transformations are similar to those involved in the derivation of Eq. (20). Substituting Eqs. (17) and (21) into the reciprocity relation Eq. (2) and using mode orthogonality relations Eqs. (8) and (9), we find that

$$E(\xi) = \frac{-\omega\rho_b \xi}{2\pi\mu(\xi)} A_0^{(1)} f(\xi,z_0). \tag{22}$$

Amplitudes of modes excited by the point source of an external force can also be obtained from the reciprocity principle, as in the case of a linear source of an external force. For such a source, $A=0$ and $\mathbf{F}=\mathbf{F}_0^{(1)}\delta(x)\delta(y)\delta(z-z_0)$ in Eq. (1). One has to take into account that the field due to the source of a vertical force is axially symmetric, i.e., omnidirectional in the horizontal plane, while a source of horizontal force produces acoustic pressure with a dipole directivity in the horizontal plane. Alternatively, the field due to the source of an external force can be calculated from the known field Eqs. (17), (21), and (22) due to the source of volume velocity by differentiation with respect to source coordinates (Ref. 1, p. 142). Either approach gives

$$p = \frac{i}{4\rho(z_0)} \sum_n \left[\xi_n \frac{xF_{01} + yF_{02}}{r} f_n(z_0) H_1^{(1)}(\xi_n r) + F_{03} \frac{\partial f_n(z_0)}{\partial z_0} H_0^{(1)}(\xi_n r) \right] f_n(z) + \frac{i\rho_b}{2\pi\rho(z_0)} \int_0^{+\infty} \left[\xi \frac{xF_{01} + yF_{02}}{r} f(\xi, z_0) H_1^{(1)}(\xi r) + F_{03} \frac{\partial f(\xi, z_0)}{\partial z_0} H_0^{(1)}(\xi r) \right] f(\xi, z) d\mu. \quad (23)$$

IV. COMPARISON TO SOLUTIONS OBTAINED VIA INTEGRAL TRANSFORMS

No approximations have been made in the above derivation of mode amplitudes from the reciprocity principle. The results, therefore, are exact and should be equivalent to those obtained in a traditional way[1,2] through application of the Fourier transform (for linear and point sources) or the Hankel transform (for a point source). Because of a simple relation between the fields due to sources of volume velocity and force, here it is sufficient to consider the former source only.

Assuming $A_0^{(1)} = -i/\omega\rho(z_0)$ and $\mathbf{F} = 0$, the following integral representations:

$$p^{(l)}(x,z) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{d\xi}{w(\xi, z_0)} f(\xi, z_<) \psi(\xi, z_>) e^{i\xi|x-x_0|}, \quad (24)$$

$$p^{(p)}(r,z) = \frac{1}{4\pi} \int_{-\infty}^{+\infty} \frac{\xi d\xi}{w(\xi, z_0)} f(\xi, z_<) \psi(\xi, z_>) H_0^{(1)}(\xi r), \quad (25)$$

were obtained[1] for linear and point sources, respectively [see Eqs. (4.3.31) and (4.4.2) in Ref. 1]. Here $w(\xi,z) = f(\xi,z)\partial\psi(\xi,z)/\partial z - \psi(\xi,z)\partial f(\xi,z)/\partial z$, $z_< = \min(z, z_0)$, and $z_> = \max(z, z_0)$. In the waveguide we are considering, the function ψ can be defined as a solution to Eq. (5) such that $\psi(\xi,z) = \exp(i\mu(\xi)z)$, $\mathrm{Im}\,\mu \geqslant 0$ at $z > H$.

By deforming integration contours in the complex ξ plane, Eqs. (24) and (25) were reduced in (Ref. 1, pp. 151–153) to a sum $p_d + p_c$ of discrete and continuous spectra, where the continuous spectra p_c were given by the same integrals as in Eqs. (24) and (25) but taken over a contour γ around a branch cut in the complex ξ plane (Fig. 1). The cut, often referred to as the EJP cut, originates at the point $\xi = k_b$ and is determined by the equation $\mathrm{Im}\,\mu = 0$. The discrete spectra are given by Eqs. (4.4.52) and (4.4.54) in Ref. 1, which in our notation become

$$p_d^{(l)} = -\frac{i}{2\rho(z_0)} \sum_n \xi_n^{-1} f_n(z) f_n(z_0) e^{i\xi_n|x-x_0|}, \qquad p_d^{(p)} = -\frac{i}{4\rho(z_0)} \sum_n f_n(z) f_n(z_0) H_0^{(1)}(\xi_n r). \quad (26)$$

Comparison of Eqs. (13) and (20) to Eq. (26) shows that amplitudes of the proper normal modes calculated using integral transforms are exactly the same as those obtained from the reciprocity principle.

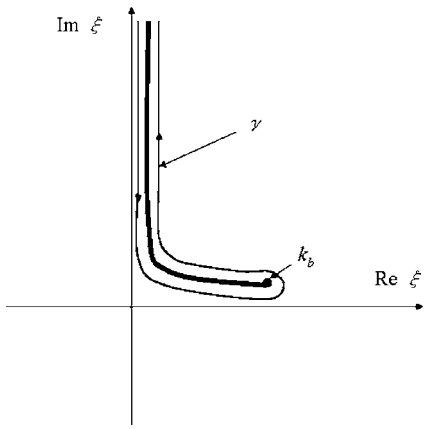


FIG. 1. Integration contour γ in the complex wave number plane. Branch cut (a boldface line) and the integration contour around it, which arises in the representation of the continuous spectrum of the acoustic field obtained with the integral transform techniques, are shown assuming a small positive imaginary part of the wave number k_b: $0 < \mathrm{Im}\,k_b \ll \mathrm{Re}\,k_b$.

The quantity $\mu(\xi)$ as well as functions $\psi(\xi,z)$ and, possibly, $f(\xi,z)$ takes different values on opposite sides of the cut in the complex ξ plane. On the part of the integration contour γ to the left and below the cut (Fig. 1), $\mu(\xi)$ varies from $+\infty$ to 0, and on the part of γ above and to the left of the cut, $\mu(\xi)$ varies from 0 to $-\infty$. Using Eqs. (24) and (25), the continuous spectrum due to the linear or point source can be written as

$$p_c^{(l)} = \int_0^{+\infty} d\mu \frac{d\xi}{d\mu} \Phi(\xi, z, z_0) e^{i\xi|x-x_0|}, \qquad p_c^{(p)} = \int_0^{+\infty} d\mu \frac{d\xi}{d\mu} \Phi(\xi, z, z_0) \frac{\xi}{2} H_0^{(1)}(\xi r), \quad (27)$$

where

$$\Phi(\xi, z, z_0) = -\frac{1}{2\pi} \frac{f(\xi, z_<)\psi(\xi, z_>)}{w(\xi, z_0)} \Bigg|_{\mu=-|\mu|}^{\mu=|\mu|}. \quad (28)$$

To simplify Eq. (28), we first note that $w(\xi,z)/\rho(z) = \mathrm{const}$. (Ref. 1, p. 143) and $w(\xi, z > H) = i\mu \exp(-i\theta)$. Hence, $w(\xi, z_0) = i\mu \exp(-i\theta)\rho(z_0)/\rho_b$. We define a function $g(\xi,z)$ as a solution to Eq. (5) with initial conditions $g = 0$, $\partial g/\partial z = 1$ at $z = 0$. The functions g and f are linearly dependent: $f(\xi,z) = a(\mu)g(\xi,z)$. As the solution of the Cauchy problem for Eq. (5), $g(\xi,z)$ is a regular function[10] of μ^2 and, therefore, is single-valued across the cut. From the single-valuedness of g at $z > H$ it follows $a(\mu) = (-1)^s a(-\mu)$, $\theta(\mu) + \theta(-\mu) = \pi s$, where s is an integer. Expressing f in terms of g, we obtain

$$\frac{f(\xi, z_<)\psi(\xi, z_>)}{w(\xi, z_0)} \Bigg|_{\mu=-|\mu|}^{\mu=|\mu|} = \frac{\rho_b}{\rho(z_0)} \frac{a(|\mu|)}{i|\mu|} g(\xi, z_<) [e^{i\theta(|\mu|)} \psi|_{\mu=|\mu|} + e^{-i\theta(|\mu|)} \psi|_{\mu=-|\mu|}].$$

The quantity in the square brackets is a solution to Eq. (5) and it equals $2\cos[|\mu|z + \theta(|\mu|)]$ at $z > H$. Hence, the quan-

tity is linearly dependent with $f(\xi,z)$. Thus, we obtain from Eq. (28)

$$\Phi(\xi,z,z_0) = \left.\frac{i}{\pi\mu}\frac{\rho_b}{\rho(z_0)}f(\xi,z)f(\xi,z_0)\right|_{\mu=|\mu|}. \tag{29}$$

Comparison of Eqs. (15) and (22) to Eq. (29) shows that the amplitudes of the improper modes obtained with the integral transforms and from the reciprocity principle are exactly the same.

V. DISCUSSION

Both the integral transform- and reciprocity-based approaches lead to exact expressions for mode amplitudes. With no approximations involved in either technique, the reciprocity-based approach does not and cannot provide any additional information about the wave field. Moreover, the application of the reciprocity principle relies on a known variation of the mode field in the horizontal plane, while the approaches based on the Hankel or Fourier transforms lead to complete solutions, including dependence of the mode field on horizontal coordinates. On the other hand, when the dependence is obvious from symmetry considerations, as it is for monopole and dipole sound sources, use of the reciprocity principle allows one to avoid relatively complicated transformations. These include integration of a multivalued function in a complex plane, as well as subsequent transformations (Ref. 1, p. 167; Ref. 2, pp. 317–319) needed to express the derivative $\partial w(\xi,z_0)/\partial\xi$ in terms of mode shape functions and present the result in a form amenable to an efficient computation. The reciprocity-based approach proposed by Achenbach is simple, intuitive, leaves little room for algebraic errors, and thus has obvious pedagogical advantages.

It is remarkable that the application of the reciprocity principle allows one to calculate amplitudes of various modes independently, and one can ignore the continuous spectrum entirely if only amplitudes of proper modes are of interest. Ultimately, this is because a proper acoustic mode is orthogonal to the other proper modes and to the continuous spectrum. The latter property is not obvious from the standard form in which continuous spectrum is obtained by the integral transforms approach. In fact, the product $f(\xi,z_<)\psi(\xi,z_>)$ in the integrand in Eqs. (24) and (25) does not satisfy the homogeneous one-dimensional reduced wave equation (5) and is *not* orthogonal to the proper mode shape function $f_n(z)$. However, the *difference* of the values of the integrand on the opposite sides of the appropriate branch cut in the complex ξ plane happens to be orthogonal to $f_n(z)$, as has been demonstrated in Sec. IV above, leading to the representations, Eqs. (4) and (17), where the discrete and continuous spectra are orthogonal.

Maupin[7] demonstrated that modes of discrete and continuous spectra are also orthogonal in a layered elastic solid that has a stress-free boundary at $z=0$, is arbitrarily stratified at $0<z<H$, and is homogeneous at $z>H$. Therefore, it is expected that in the layered solid, much like in the acoustic waveguide considered above, disregarding the continuous spectrum does not affect calculation of the amplitudes of proper modes, including surface waves, from the reciprocity principle.

Calculation of mode amplitudes in an acoustic waveguide in a motionless fluid is a well-studied problem, which we have chosen in order to discuss properties of the reciprocity-based approach proposed by Achenbach in the simplest possible setting. It is of interest to apply the technique to more complicated problems such as wave excitation by a sound source in an inhomogeneous moving medium and generation of gravity and acoustic-gravity waves in a stratified fluid. These problems will be considered elsewhere.

ACKNOWLEDGMENTS

This work was supported in part by the Office of Naval Research, Grant No. N00014-05-IP-20024.

[1] L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Medium. 2: Point Sources and Bounded Beams*, 2nd ed. (Springer-Verlag, Berlin, 1999), Chap. 4.
[2] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, 2nd printing (Springer-Verlag, New York, 2000), Chap. 5.
[3] J. D. Achenbach and Y. Xu, "Wave motion in an isotropic elastic layer generated by a time-harmonic point load of arbitrary direction," J. Acoust. Soc. Am. **106**, 83–90 (1999).
[4] J. D. Achenbach, "Calculation of surface wave motions due to a subsurface point source: An application of elastodynamic reciprocity," J. Acoust. Soc. Am. **107**, 1892–1897 (2000).
[5] J. D. Achenbach, "Simplifications for the calculation of surface wave pulses generated by laser-radiation," J. Acoust. Soc. Am. **116**, 1481–1487 (2004).
[6] J. D. Achenbach, *Reciprocity in Elastodynamics*, Cambridge Monographs on Mechanics (Cambridge University Press, Cambridge, 2004), X+255 pp.
[7] V. Maupin, "The radiation modes of a vertically varying half-space: A new representation of the complete Green's function in terms of modes," Geophys. J. Int. **126**, 762–780 (1996).
[8] C. H. Wilcox, *Sound Propagation in Stratified Fluids* (Springer-Verlag, New York, 1984). pp. 104–124.
[9] *Handbook of Mathematical Functions with Formulas, Graphs, and Tables*, Appl. Math. Ser., Vol. **55**, edited by M. Abramovitz and I. A. Stegun (Dover, New York, 1965). Sec. 9.1.
[10] E. Kamke, *Handbook of Ordinary Differential Equations* (Chelsea, New York, 1971). Part I, Sec. 5.

Mean field of a low-frequency sound wave propagating in a fluctuating ocean

Alexander G. Voronovich[a] and Vladimir E. Ostashev
NOAA/Earth System Research Laboratory, 325 Broadway, Boulder, Colorado 80305

(Received 14 November 2005; revised 25 January 2006; accepted 25 January 2006)

Statistical characteristics of low-frequency sound waves propagating over long distances in a fluctuating ocean are important for many practical problems. In this paper, using the theory of multiple scattering, the mean field of a low-frequency sound wave was analytically calculated. In these calculations, the ratio of the sound wavelength and the scale of random inhomogeneities can be arbitrary. Furthermore, the correlation function of inhomogeneities is expressed in terms of a modal spectrum (e.g., internal waves modes). The obtained mean sound field is expressed as a sum of normal modes that attenuate exponentially. It is shown that the extinction coefficients of the modes are linearly related to the spectrum of random inhomogeneities in the ocean. Measurements of the extinction coefficients can therefore be used for retrieving this spectrum. The mean sound field is calculated for both 3D and 2D geometries of sound propagation. The results obtained can be used to study the range of applicability of the 2D propagation model. [DOI: 10.1121/1.2177591]

PACS number(s): 43.30.Re, 43.30.Bp, 43.30.Qd [RMW] Pages: 2101–2105

I. INTRODUCTION

Studies of the statistical characteristics of low-frequency sound waves propagating over long ranges in a fluctuating ocean are important for several practical concerns.[1,2] The coherence function of such sound waves has been calculated in many papers (e.g., see Refs. 3–7). The theoretical results obtained are very involved and are based on several simplifying assumptions.

Calculation of the first statistical moment of a sound wave propagating in a fluctuating ocean, i.e., the mean sound field, is simpler than that for the second moment and can be made without some of the assumptions mentioned above. The mean sound field of a low-frequency wave can persist over distances of hundreds or even thousands of kilometers. This field can be measured experimentally by monitoring both the amplitude and phase of a sound wave. The mean sound field in a fluctuating ocean was calculated in a number of papers, e.g., Refs. 3, 5, 8, and 9. In the present paper, it is calculated by an approach which, in contrast to previous works, is valid for arbitrary ratio between the sound wavelength and the horizontal scale of random inhomogeneities. Furthermore, in our approach, the correlation function of inhomogeneities is expressed in terms of a modal spectrum.

Reference 10 presents a general theory based on the diagram technique for calculation of the mean field of a wave propagating in a random medium without regular refraction. Our approach coincides with that in Ref. 10 with the Green function of a free space being replaced by that of an oceanic waveguide. The mean Green function obtained is expressed as a sum of normal modes, each of which attenuates exponentially. It is shown that the extinction coefficients of the modes linearly depend on the spectrum of oceanic inhomogeneities. Therefore, measurements of these extinction coefficients for different frequencies and mode numbers can be used for remote sensing of this spectrum.[7]

The mean Green function is calculated below for both 3D and 2D sound propagation in a fluctuating ocean. Comparison between the results obtained allows one to evaluate ranges of applicability of the 2D propagation model.

II. 3D PROPAGATION

A. Starting equations

Let $\mathbf{R}=(x,y,z)$ be the Cartesian coordinates with the z axis directed upwards. Our goal is to calculate the mean value of the Green function $G(\mathbf{R},\mathbf{R}_0)$ due to a point source located at $\mathbf{R}_0=(x_0,y_0,z_0)$ in a fluctuating ocean.

The Green function G satisfies the Helmholtz equation:

$$\left(\frac{\partial^2}{\partial z^2}+\Delta_\perp+k^2(z)-\epsilon(\mathbf{R})\right)G(\mathbf{R},\mathbf{R}_0)=\delta(\mathbf{r}-\mathbf{r}_0)\delta(z-z_0). \tag{1}$$

Here, $\mathbf{r}=(x,y)$, $\mathbf{r}_0=(x_0,y_0)$, $\Delta_\perp=\partial^2/\partial x^2+\partial^2/\partial y^2$, $k(z)$ is the sound wave number in the ocean without fluctuations, and $\epsilon(\mathbf{R})=k_0^2\tilde{n}(\mathbf{R})$, where k_0 is the reference sound wave number and $\tilde{n}(\mathbf{R})$ are fluctuations of the refractive index in the ocean. The Helmholtz equation can be written in the equivalent integral form:

$$G(\mathbf{R},\mathbf{R}_0)=G_0(\mathbf{R},\mathbf{R}_0)+\int d\mathbf{R}'G_0(\mathbf{R},\mathbf{R}')\epsilon(\mathbf{R}')G(\mathbf{R}',\mathbf{R}_0). \tag{2}$$

In this equation, $G_0(\mathbf{R},\mathbf{R}_0)$ is the Green function of the unperturbed problem, which satisfies the following equation:

[a]Electronic mail: alexander.voronovich@noaa.gov

$$\left(\frac{\partial^2}{\partial z^2}+\Delta_\perp+k^2(z)\right)G_0=\delta(\vec{r}-\vec{r}_0)\,\delta(z-z_0). \tag{3}$$

The value of G_0 is well known. It can be represented as follows:

$$G_0(\mathbf{R},\mathbf{R}_0)=G_0(\mathbf{r}-\mathbf{r}_0;z,z_0)$$
$$=-\frac{i}{4}\sum_n u_n(z)u_n(z_0)H_0^{(1)}(\xi_n|\mathbf{r}-\mathbf{r}_0|). \tag{4}$$

Here, the summation index n corresponds to all modes, $H_0^{(1)}$ is the Hankel function, and u_n and ξ_n are the eigenfunctions and propagation constants of the standard boundary problem:

$$\left(\frac{d^2}{dz^2}+k^2(z)-\xi_n^2\right)u_n(z)=0,\quad u_n(z=0)=u_n(z=-\infty)$$
$$=0, \tag{5}$$

where the level $z=0$ corresponds to the ocean surface. Note that the boundary condition $u_n\ (z=-\infty)=0$ means that the modes attenuate in the bottom. The eigenfunctions u_n are normalized as follows:

$$\int_{-\infty}^{0} u_n u_m\,dz=\delta_{nm}. \tag{6}$$

Here, δ_{nm} is the Kronecker symbol.

B. Dyson equation

It is worthwhile to write Eq. (2) in the operator form:

$$G=G_0+\hat{G}_0\epsilon G. \tag{7}$$

Here, $\hat{G}_0$ is the integral operator defined as $\hat{G}_0 f=\int d\mathbf{R}'G_0(\mathbf{R},\mathbf{R}')f(\mathbf{R}')$, where f is an arbitrary function. A formal solution of Eq. (7) is given by the Born series:

$$G=G_0+\hat{G}_0\epsilon G_0+\hat{G}_0\epsilon\hat{G}_0\epsilon G_0+\cdots. \tag{8}$$

Averaging both sides of this equation, assuming that the random field ϵ has the Gaussian statistics, and separating irreducible terms, we obtain

$$\bar{G}=G_0+\hat{G}_0\hat{Q}G_0+\hat{G}_0\hat{Q}\hat{G}_0\hat{Q}G_0+\cdots. \tag{9}$$

Here, $\bar{G}$ is the mean Green function, and the operator $\hat{Q}$ is the sum of all strongly connected diagrams.[10] The kernel Q of this operator is the sum of terms of order ϵ^{2m} (where $m=1,2,3,\ldots$):

$$Q(\mathbf{R},\mathbf{R}')=G_0(\mathbf{R},\mathbf{R}')B(\mathbf{R},\mathbf{R}')+\int\int G_0(\mathbf{R},\mathbf{R}_1)G_0(\mathbf{R}_1,\mathbf{R}_2)G_0(\mathbf{R}_2,\mathbf{R}')B(\mathbf{R},\mathbf{R}_2)B(\mathbf{R}_1,\mathbf{R}')\,d^3\mathbf{R}_1\,d^3\mathbf{R}_2$$
$$+\int\int G_0(\mathbf{R},\mathbf{R}_1)G_0(\mathbf{R}_1,\mathbf{R}_2)G_0(\mathbf{R}_2,\mathbf{R}')B(\mathbf{R},\mathbf{R}')B(\mathbf{R}_1,\mathbf{R}_2)\,d^3\mathbf{R}_1\,d^3\mathbf{R}_2+\cdots. \tag{10}$$

Here, $B(\mathbf{R},\mathbf{R}')=\langle\epsilon(\mathbf{R})\epsilon(\mathbf{R}')\rangle$ is the correlation function of the random field $\epsilon(\mathbf{R})$.

Equation (9) can be written as the following integral equation for $\bar{G}$:

$$\bar{G}=G_0+\hat{G}_0\hat{Q}\bar{G}. \tag{11}$$

Indeed, solving this equation by iterations, one obtains Eq. (9). Equation (11) is called the Dyson equation. Applying the operator $\partial^2/\partial z^2+\Delta_\perp+k^2(z)$ to both sides of Eq. (11), we obtain the Dyson equation in the integro-differential form:

$$\left(\frac{\partial^2}{\partial z^2}+\Delta_\perp+k_0^2(z)-\hat{Q}\right)\bar{G}=\delta(\mathbf{r}-\mathbf{r}_0)\,\delta(z-z_0). \tag{12}$$

C. Bourret approximation

In the Bourret approximation,[10] the kernel Q is approximated by the first term on the right-hand side of Eq. (10). A necessary condition of this approximation is that the mean sound field attenuates weakly on the horizontal scale of inhomogeneities.[10,11] In what follows, we assume that random inhomogeneities in the ocean have horizontally homogeneous statistics, i.e., $B(\mathbf{R},\mathbf{R}')=B(\mathbf{r}-\mathbf{r}';z,z')$. In this case, using Eqs. (10) and (4), we obtain the following expression for the kernel Q:

$$Q(\mathbf{R},\mathbf{R}')=-\frac{i}{4}\sum_n u_n(z)u_n(z')H_0^{(1)}(\xi_n|\mathbf{r}-\mathbf{r}'|)B(\mathbf{r}-\mathbf{r}';z,z')=\int\beta(\boldsymbol{\kappa};z,z')e^{i\boldsymbol{\kappa}\cdot(\mathbf{r}-\mathbf{r}')}\frac{d\boldsymbol{\kappa}}{(2\pi)^2}, \tag{13}$$

where

$$\beta(\boldsymbol{\kappa};z,z')=-\frac{i}{4}\sum_n u_n(z)u_n(z')\int H_0^{(1)}(\xi_n r)\times B(\mathbf{r};z,z')e^{-i\boldsymbol{\kappa}\cdot\mathbf{r}}\,d\mathbf{r} \tag{14}$$

is a symmetric kernel: $\beta(\boldsymbol{\kappa};z,z')=\beta(\boldsymbol{\kappa};z',z)$.

Without loss of generality, we can assume that $\mathbf{r}_0=0$ in Eq. (12). Performing a Fourier transform of Eq. (12), we find that

$$\left(\frac{d^2}{dz^2}+k^2(z)-\kappa^2\right)\bar{g}(\boldsymbol{\kappa};z,z_0)$$
$$-\int_{-\infty}^{0}\beta(\boldsymbol{\kappa};z,z')\bar{g}(\boldsymbol{\kappa};z',z_0)\,dz'=\delta(z-z_0), \tag{15}$$

where

$$\bar{g}(\boldsymbol{\kappa};z,z_0)=\int \bar{G}(\mathbf{r};z,z')e^{-i\boldsymbol{\kappa}\cdot\mathbf{r}}\,d\mathbf{r}. \tag{16}$$

Equation (15) is similar to the Fourier transform of Eq. (3) with respect to the coordinates $\mathbf{r}$ and determines mean field modes which generally differ from the unperturbed ones. The boundary conditions for Eq. (15) coincide with those for the unperturbed problem; see Eq. (5). Equation (15) can be solved by a perturbation technique provided that corrections to eigenvalues of Eq. (15) are significantly smaller than the difference between eigenvalues. The physical meaning of this condition is that the inverse of the decrement of attenuation of the mean field is less than the ray skip distance. This condition holds for sufficiently low-frequency sound waves scattered by internal waves.

D. Solution of the Dyson equation

The perturbative solution of Eq. (15) results in the well-known corrections $\delta\xi_n$ to the eigenvalues ξ_n [see Eq. (38.6) from Ref. 12 or any other textbook on quantum mechanics]:

$$\delta\xi_n=-\frac{\langle u_n|\beta(\xi_n;z,z')|u_n\rangle}{2\xi_n}$$
$$=-\frac{1}{2\xi_n}\int_{-\infty}^{0}dzdz'u_n(z)u_n(z')\beta(\xi_n;z,z')$$
$$=\frac{i\pi}{4\xi_n}\sum_m\int_{-\infty}^{0}dzdz'u_n(z)u_n(z')u_m(z)u_m(z')$$
$$\times\int_0^\infty H_0^{(1)}(\xi_m r)B(r;z,z')J_0(\xi_n r)r\,dr. \tag{17}$$

Here, J_0 is the Bessel function, and it is assumed that the random inhomogeneities are statistically isotropic in the horizontal plane.

The mean Green function can be represented as follows:

$$\bar{G}(\mathbf{R},\mathbf{R}_0)=-\frac{i}{4}\sum_n u_n(z)u_n(z_0)H_0^{(1)}((\xi_n+\delta\xi_n)|\mathbf{r}-\mathbf{r}_0|). \tag{18}$$

Here, we neglected small (first-order) variations in the eigenfunctions.

Let us express the correlation function B in terms of its spectrum:

$$B(r;z,z')=\sum_j\int_0^\infty P_j(\kappa)J_0(\kappa r)\Phi_j(\kappa,z)\Phi_j(\kappa,z')\kappa\,d\kappa. \tag{19}$$

Here, $\Phi_j(\kappa,z)$ is an eigenfunction of the symmetric integral operator with kernel $B(\kappa;z,z')$, which is the Fourier-Bessel transform of the kernel $B(r;z,z')$, and $P_j(\kappa)$ is an appropriate modal spectrum. For the case of linear internal waves (IW), functions $\Phi_j(\kappa,z)$ coincide with the IW modes. For instance, in the case of the Garret-Munk (GM) spectrum of IW, one has

$$P_j(\kappa)=E_0\frac{2}{\pi^3}\frac{j_*j}{j^2+j_*^2}\frac{\kappa_*\kappa}{(j^2\kappa_*^2+\kappa^2)^2}, \tag{20}$$

where $\kappa_*=\pi\omega_i/(B_{\mathrm{M}}n_0)$, ω_i is the inertial frequency, B_{M} is the width of Munk's canonical profile, n_0 is a Brunt-Väisälä frequency $n(z)$ extrapolated to the surface [an exponential profile of $n(z)=n_0\exp(-z/B_M)$ is assumed], and j_* and E_0 are parameters of the GM spectrum (the values $j_*=3$ and $E_0=4\ \mathrm{m^3s^{-2}}$ are frequently used.) IW modes are normalized as follows:

$$\int_{-H}^{0}(n^2(z)-\omega_i^2)\Phi_j(\kappa,z)\Phi_{j'}(\kappa,z)\,dz=\delta_{jj'}, \tag{21}$$

where $z=-H$ is the level of the bottom.

Substituting Eq. (19) into Eq. (17), one finds that

$$\delta\xi_n=\frac{i\pi}{4\xi_n}\sum_{m,j}\int_0^\infty d\kappa\kappa P_j(\kappa)\left[\int_{-H}^{0}u_n(z)\Phi_j(\kappa,z)u_m(z)\,dz\right]^2$$
$$\times\int_0^\infty H_0^{(1)}(\xi_m r)J_0(\kappa r)J_0(\xi_n r)r\,dr. \tag{22}$$

In this equation, the Hankel $H_0^{(1)}$ function can be represented as

$$-\frac{i}{4}H_0^{(1)}(\kappa r)=\frac{1}{(2\pi)^2}\int\frac{e^{i\boldsymbol{\xi}\cdot\mathbf{r}}d\boldsymbol{\xi}}{\kappa^2-\xi^2+i0}$$
$$=\frac{1}{2\pi}\int_0^\infty\frac{J_0(\xi r)\xi d\xi}{\kappa^2-\xi^2+i0}. \tag{23}$$

Using this formula, Eq. (22) can be written as

$$\delta\xi_n=-\frac{1}{4\pi\xi_n}\sum_{m,j}\int_0^\infty d\kappa\kappa P_j(\kappa)$$
$$\times\left[\int u_n(z)\Phi_j(\kappa,z)u_m(z)dz\right]^2$$
$$\times\int_0^\infty\frac{\xi d\xi}{\xi_m^2-\xi^2+i0}\frac{1}{\Delta(\xi_n,\xi,\kappa)}. \tag{24}$$

Here

$$\frac{1}{\Delta(\xi_n,\xi,\kappa)}=2\pi\int_0^\infty J_0(\xi_n r)J_0(\xi r)J_0(\kappa r)r\,dr \tag{25}$$

and $\Delta(\xi_n,\xi,\kappa)$ is an area of the triangle formed by segments with lengths ξ_n, ξ, and κ [the integral in Eq. (25) equals zero if such a triangle does not exist]. An explicit expression for $\Delta(\xi_n,\xi,\kappa)$ is

$$\Delta(\xi_n,\xi,\kappa)=\frac{1}{4}\sqrt{(\xi_n+\xi)^2-\kappa^2}\sqrt{\kappa^2-(\xi_n-\xi)^2}. \tag{26}$$

Since

$$\mathrm{Im}\,\frac{1}{\xi_m^2-\xi^2+i0}=-\pi\delta(\xi_m^2-\xi^2),$$

for the decrement of attenuation of the nth propagating mode, we find that

$$\mathrm{Im}\,\delta\xi_n=\frac{1}{8\xi_n}\sum_{m,j}\int_{|\xi_n-\xi_m|}^{\xi_n+\xi_m}P_j(\kappa)\times\left[\int_{-H}^{0}u_n(z)\Phi_j(\kappa,z)u_m(z)dz\right]^2\frac{\kappa d\kappa}{\Delta(\xi_n,\xi_m,\kappa)}. \quad (27)$$

Here, for simplicity, we assumed that u_n and Φ_j are real functions.

Equations (24) and (27) for ξ_n and its imaginary part are the main results obtained in the present paper. These equations differ from those known in the literature, e.g., Refs. 5, 8, and 9. First, Eqs. (24) and (27) are valid for an arbitrary ratio between the sound wavelength and the horizontal scale of random inhomogeneities while those obtained previously are valid when this ratio is small so that a one-way propagation or small angle approximation can be used. Second, when deriving Eqs. (24) and (27), the correlation function $B(r;z,z')$ was expressed in terms of a modal spectrum (e.g., internal waves modes) [see Eq. (19)]. This allowed us to simplify formulas for ξ_n and Im $\delta\xi_n$. Note that, in the often referenced paper of Ref. 3, the mean sound field is equal to zero due to approximations adopted in that paper.

III. 2D PROPAGATION

To simplify numerical calculations of sound propagation in a fluctuating ocean, 2D propagation models are often considered. In this section, we calculate the mean Green function for the case of 2D sound propagation in a fluctuating ocean. The obtained mean Green function will be compared with that in 3D case, Eq. (18).

In 2D, Eq. (1) takes the form

$$\left(\frac{\partial^2}{\partial z^2}+\frac{\partial^2}{\partial x^2}+k^2(z)-\epsilon(x,z)\right)G=\delta(x-x_0)\delta(z-z_0). \quad (28)$$

This case can be considered similarly to the previous section. Equation (4) is replaced by the expression

$$G_0(\mathbf{R},\mathbf{R}_0)=G_0(x-x_0;z,z_0)=\sum_n u_n(z)u_n(z_0)\frac{\exp(i\xi_n|x-x_0|)}{2i\xi_n}, \quad (29)$$

and Eq. (15) becomes

$$\left(\frac{d^2}{dz^2}+k^2(z)-\kappa^2\right)\bar{g}(\kappa;z,z_0)-\int_{-H}^{0}\beta(\kappa;z,z')\bar{g}(\kappa;z',z_0)\,dz'=\delta(z-z_0). \quad (30)$$

Here,

$$\bar{g}(\kappa;z,z_0)=\int\bar{G}(x;z,z_0)e^{-i\kappa x}\,dx, \quad (31)$$

and

$$\beta(\kappa;z,z')=\sum_m u_m(z)u_m(z')\times\int_{-\infty}^{\infty}\frac{\exp(i\xi_m|x|)}{2i\xi_m}B(x;z,z')e^{-i\kappa x}\,dx. \quad (32)$$

In this equation, $B(x;z,z')$ is replaced with the right-hand side of Eq. (19). Furthermore, Eq. (30) is solved by the perturbation method used above. As a result, the mean Green function $\bar{G}(x-x_0;z,z_0)$ is given by the right-hand side of Eq. (29) if ξ_n is replaced with $\xi_n+\delta\xi_n$. However, the corrections to the propagation constants are given by a formula different from Eq. (22):

$$\delta\xi_n=\frac{1}{4\xi_n}\sum_{m,j}\frac{1}{\xi_m}\int_0^\infty d\kappa\kappa P_j(\kappa)\times\left[\int_{-H}^{0}u_n(z)\Phi_j(\kappa,z)u_m(z)\,dz\right]^2\times\int_{-\infty}^{\infty}e^{i\xi_m|x|-i\xi_n x}J_0(\kappa x)\,dx \quad (33)$$

For the decrement of attenuation, one obtains

$$\mathrm{Im}\,\delta\xi_n=\frac{1}{4\xi_n}\sum_{m,j}\frac{1}{\xi_m}\int_0^\infty d\kappa\kappa P_j(\kappa)\times\left[\int_{-H}^{0}u_n(z)\Phi_j(\kappa,z)u_m(z)\,dz\right]^2\times\mathrm{Re}\int_{-\infty}^{\infty}e^{i\xi_m|x|-i\xi_n x}J_0(\kappa x)\,dx. \quad (34)$$

In this equation,

$$\mathrm{Re}\int_{-\infty}^{\infty}e^{i\xi_m|x|-i\xi_n x}J_0(\kappa x)\,dx=2\int_0^\infty\cos(\xi_n x)\cos(\xi_m x)J_0(\kappa x)\,dx=\int_0^\infty\cos[(\xi_n+\xi_m)x]J_0(\kappa x)\,dx+\int_0^\infty\cos[(\xi_n-\xi_m)x]J_0(\kappa x)\,dx. \quad (35)$$

Using the relationship

$$\int_0^\infty \cos(\xi x) J_0(\kappa x)\, dx = \begin{cases} (\kappa^2 - \xi^2)^{-1/2}, & \kappa^2 > \xi^2, \\ 0, & \kappa^2 < \xi^2, \end{cases} \tag{36}$$

Eq. (34) can be represented as

$$\text{Im}\ \delta\xi_n = \frac{1}{4\xi_n} \sum_{m,j} \int_{|\xi_n - \xi_m|}^{\xi_n + \xi_m} P_j(\kappa) \left[\int_{-H}^{0} u_n(z) \Phi_j(\kappa, z) u_m(z)\, dz \right]^2 \frac{\kappa\, d\kappa}{\xi_m \sqrt{\kappa^2 - (\xi_n - \xi_m)^2}} + \frac{1}{4\xi_n} \sum_{m,j} \int_{\xi_n + \xi_m}^{\infty} d\kappa \kappa P_j(\kappa)$$
$$\times \left[\int_{-H}^{0} u_n(z) \Phi_j(\kappa, z) u_m(z)\, dz \right]^2 \times \frac{1}{\xi_m} \left(\frac{1}{\sqrt{\kappa^2 - (\xi_n - \xi_m)^2}} + \frac{1}{\sqrt{\kappa^2 - (\xi_n + \xi_m)^2}} \right). \tag{37}$$

We now assume that the spectrum of inhomogeneities $P_j(\kappa)$ for small κ is much greater than that for $\kappa \geqslant \xi_n + \xi_m$. In this case, the second term on the right-hand side of Eq. (37) can be ignored. Then, using Eq. (26), one can represent Eq. (37) as follows:

$$\text{Im}\ \delta\xi_n \approx \frac{1}{8\xi_n} \sum_{m,j} \int_{|\xi_n - \xi_m|}^{\xi_n + \xi_m} P_j(\kappa) \times \left[\int_{-H}^{0} u_n(z) \Phi_j(\kappa, z) u_m(z)\, dz \right]^2 \times \frac{\sqrt{(\xi_n + \xi_m)^2 - \kappa^2}}{2\xi_m} \frac{\kappa\, d\kappa}{\Delta(\xi_n, \xi_m, \kappa)}. \tag{38}$$

Here, the main contribution to the integral over κ comes from $\kappa \ll \xi_n + \xi_m$. Therefore, one can neglect κ^2 in the square root in Eq. (38):

$$\frac{\sqrt{(\xi_n + \xi_m)^2 - \kappa^2}}{2\xi_m} \approx \frac{\xi_n + \xi_m}{2\xi_m} = 1 + \frac{\xi_n - \xi_m}{2\xi_m}. \tag{39}$$

If in Eq. (39) the term $(\xi_n - \xi_m)/(2\xi_m)$ can be neglected, Eq. (38) coincides with the 3D result, Eq. (27). If the interaction between the nth and mth modes is significant (i.e., the integral over depth of the triple product of the modes' profiles is not small), corrections to the mode attenuations increase with the increase of the maximum angle of the water-borne ray supported by the waveguide.

IV. CONCLUSIONS

The mean Green function of a sound field propagating in a fluctuating ocean was calculated. The formula for the mean Green function was expressed as a sum of normal modes, which attenuate exponentially with the distance of propagation. The extinction coefficients of the mean modes linearly depend on the spectrum of IW. Therefore, measurements of the extinctions coefficients can be used for retrieving the IW spectrum.

The mean Green function was also calculated for 2D sound propagation in a fluctuating ocean. The difference between the Green functions in 3D and 2D cases can be used to verify applicability of the 2D propagation models which are often used in numerical simulation of sound propagation through a fluctuating ocean.

The 2D and 3D mean Green functions calculated in this paper are valid for any ratio between the sound wavelength and the horizontal scale of ocean inhomogeneities.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research Grant No. N0001404IP20009.

[1] P. F. Worcester and R. C. Spindel, "North Pacific Acoustic Laboratory," J. Acoust. Soc. Am. **117**(3) (Pt. 2), 1499–1510 (2005).
[2] A. G. Voronovich, V. E. Ostashev, and the NPAL group (J. A. Colosi, B. D. Cornuelle, B. D. Dushaw, M. A. Dzieciuch, B. M. Howe, J. A. Mercer, W. H. Munk, R. C. Spindel, and P. F. Worcester), "Horizontal refraction of acoustic signals retrieved from NPAL billboard array data," J. Acoust. Soc. Am. **117**(3) (Pt. 2), 1527–1537 (2005).
[3] L. Dozier and F. Tappert, "Statistics of normal mode amplitudes in a random ocean," J. Acoust. Soc. Am. **63**, 353–365 (1978).
[4] E. Yu. Gorodetskaya, A. I. Malekhanov, A. G. Sazontov, and N. K. Vdovicheva, "Deep-water acoustic coherence at long-ranges: Theoretical prediction and effects on large-array signal processing," IEEE J. Ocean. Eng. **24**, 156–170 (1999).
[5] S. Frankenthal and M. J. Beran, "Propagation in random stratified waveguides—A modal-spectral treatment," J. Acoust. Soc. Am. **104**, 3282–3295 (1998).
[6] J. A. Colosi, "Coupled mode theory for sound propagation through random internal wave fields," J. Acoust. Soc. Am. **116** (Pt. 2), 2535 (2005).
[7] A. G. Voronovich, "Low-frequency sound propagation through random internal waves with applications to measurements of internal wave spectra by acoustical means," in *Proceedings of the Fourth European Conference on Underwater Acoustics*, edited by A. Alippi and G. B. Canelli, Rome, 1998, Vol. **2**, pp. 751–756.
[8] C. Penland, "Acoustic normal mode propagation through a three-dimensional internal wave field," J. Acoust. Soc. Am. **78**, 1356–1365 (1985).
[9] A. L. Vyrovlyanskii, A. G. Kosterin, and A. N. Sazontov, "Mode fluctuations in a canonical underwater channel," Sov. Phys. Acoust. **35**, 138–142 (1989).
[10] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskii, *Principles of Statistical Radio Physics. Part 4, Wave Propagation through Random Media* (Springer, Berlin, 1989).
[11] V. E. Ostashev, *Acoustics in Moving Inhomogeneous Media* (E&FN SPON, London, 1997).
[12] L. M. Landau and E. M. Liftshitz, *Course of Theoretical Physics. V. 3, Quantum Mechanics (Non-relativistic Theory)* (Elsevier Butterworth-Heinemann, Amsterdam, 2005).

Multiple ping sonar accuracy improvement using robust motion estimation and ping fusion

Lian Yu
Department of Mathematics, Beijing Normal University, Beijing 100875, P. R. China

Nicola Neretti
Institute of Brain and Neural Systems, Brown University, Providence, Rhode Island 02906

Nathan Intrator
The School of Computer Science, Tel-Aviv University, Ramat Aviv 69978, Israel

(Received 21 May 2005; revised 20 December 2005; accepted 6 January 2006)

Noise degrades the accuracy of sonar systems. We demonstrate a practical method for increasing the effective signal-to-noise ratio (SNR) by fusing time delay information from a burst of multiple sonar pings. This approach can be useful when there is no relative motion between the sonar and the target during the burst of sonar pinging. Otherwise, the relative motion degrades the fusion and therefore, has to be addressed before fusion can be used. In this paper, we present a robust motion estimation algorithm which uses information from multiple receivers to estimate the relative motion between pings in the burst. We then compensate for motion, and show that the fusion of information from the burst of motion compensated pings improves both the resilience to noise and sonar accuracy, consequently increasing the operating range of the sonar system. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2169913]

PACS number(s): 43.30.Vh [WMC] Pages: 2106–2113

I. INTRODUCTION

Sonar system measures the distance between sonar and targets by emitting sonar pings and measuring the delay to the received echoes from the targets. Through a beam focusing method and other processing steps, a two- or three-dimensional (2D or 3D) representation of the environment is created. For a multisensor sonar array system, using beamforming or image reconstruction technique, a 2D image can be obtained. With increasing range, the received signal is attenuated, and therefore more affected by background noise. In fact, when the SNR is below a certain threshold, according to the Woodward equation,[1] accuracy degrades rapidly.[2] The challenge of a sonar system is to cover a long range without sacrificing its accuracy. One possible solution to this problem is to average returns over multiple sonar pings. Such averaging is likely to reduce the effect of noise by a factor of $1/\sqrt{n}$, where n is the number of successive sonar pings. Neretti *et al.*[3] have demonstrated methods which are more robust than simple averaging to improve on echo delay accuracy at high levels of noise. Their approach is useful when there is no relative motion between the sonar and the targets during the burst of pings which is used for time delay estimation. When relative motion exists, their method cannot be readily used. This paper addresses this problem by introducing a robust motion estimation method which is applied before the fusion of information from multiple pings. The robust motion estimation introduced in the paper relies on the use of multiple-sensors to record the returning echo. While sending n pings requires n times the energy level of a single ping, there are several reasons why sending a high peak energy is less desirable than sending n times a lower energy signal. This has to do with the properties of the sound transducer, as well as security issues attempting to stay undetected.

Although there are six (three translational and three rotational) possible motion variables for a sonar system, only two have significant effect on the multisensor array. They are the sway (the horizontal shift from the mean position of the array) and the yaw (the horizontal angular rotation around the center of the array). There are three options to determine the motion errors: Using hardware systems to record the different degrees of freedom of the sonar system, using motion estimation algorithms, and using information from the reconstructed sonar image. Current navigation instruments, such as inertial units, are either expensive or unable to provide adequate level of accuracy.[4] The problem of motion estimation has been dealt in the past in the context of synthetic aperture sonar (SAS) where motion estimation is used to create a larger (virtual) sensor array. Thus, we refer to previous work on motion estimation which exists in the SAS literature, although in this paper, we do not deal with SAS, but only with motion estimation for a rigid multiple-sensor sonar. Algorithms, which estimate motion that are based on the raw sonar returns, often rely on correlation or phase estimation. The correlation algorithms include amplitude-only envelope correlation, complex base band signal correlation[5] and the shear average[6] method. The performance of the phase estimation technique relies heavily on the choice of the phase estimation kernels. A detailed comparison of the different phase estimation kernels is given by Callow.[7] A popular phase estimation technique is the "displaced phase center" (DPC) algorithm,[8] also known as the "redundant phase center" (RPC) algorithm.[9] This algorithm uses the phase and time shifts in the cross-correlation of the positions of the

 0001-4966/2006/119(4)/2106/8/$22.50

collocated phase centers to interpret the displacements of the receiver array. The image based motion estimation algorithms use 2D cross-correlation on the reconstructed images to estimate yaw and sway. A "beam to beam correlation" technique is described by Tonard and Brussieux[10] where a 2D cross-correlation of the adjacent ping images is used to estimate the relative yaw angle. Gough and Miller[11] introduce a "displaced ping imaging autofocus" (DPIA) algorithm performing a 2D cross-correlation on the Fourier domain of the low-accuracy images to estimate the yaw. A high-order correlation (HOC) scheme[12] has been developed for post processing the beamformed images in order to significantly minimize the incident of false detections. Generally, the raw-data based motion estimation algorithms simply average the estimation result of multiple sensors, so they can only estimate sway and be suitable for higher signal-to-noise-ratio (SNR) conditions. On the other hand, the image based approaches using information from multiple sensors are more resilient to lower SNR. However, they rely on a dominant target return, which limits the usefulness of image based approaches when detecting more complicated structure objects.

This paper introduces a robust motion estimation technique to estimate the relative motion while sending multiple pings from a relatively fixed location. Thus, the motion estimation method can be useful for a slow moving target (with respect to the burst of pings), but is not applicable for a fast moving target such as Torpedo. This technique relies on information obtained from all receiving sensors for each returning echo. Following the motion compensation for a burst of pings, the ideal situation would be that the target becomes stationary relative to the sonar during the burst. We then demonstrate that this burst of multiple pings provides information which can be used to increase sonar accuracy and resiliency to noise.

We first present a model for robust estimation of sway and yaw of the relative motion between sonar and target during the transmission of the burst. The motion estimated, can be a motion of the sonar, or the target or both, as we concentrate on the estimation of the relative motion between them. Following the compensation of the estimated relative motion, we demonstrate improved resiliency to noise after fusing the realigned burst of pings. This is done using the Woodward curves[1] which provides echo delay accuracy [measured by the variability or root mean square error (RMSE)] of time-delay estimation versus SNR. We study and compare several methods for motion estimation/compensation and several methods for fusing the motion compensated data from the burst of pings.

This paper does not deal with the multiple path issue of acoustic signals. It is assumed that the relative motion during the burst is small, and then, since each returning ping is analyzed separately and compared to its corresponding neighbors in the adjacent ping, the multiple-path problem will not interfere with the robust estimation of the motion from multiple sensors as the shift in echoes will be coherent and small to all the sensors. This assumption will be tested in a subsequent paper.

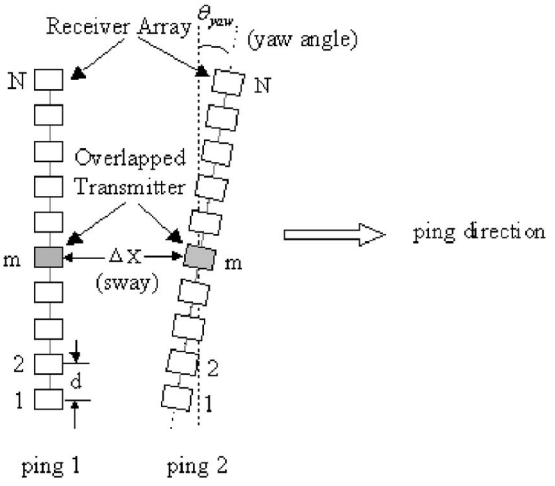


FIG. 1. Modeling motion between two pings. The sonar system consists of one transmitter and multiple receivers. The transmitter (gray block) is located at the center of the receiver array. The receiver is labeled from 1 to N, where N is the total number of receiving sensors. The center receiver's index is m. d is the space between adjacent receivers. The relative motion of the receiver array between two pings is shown. As the sensors are situated on a straight line, the motion between any two pings is a rigid motion. The displacement ΔX of sensor m is the sway. The angle θ_{yaw} between two receiver arrays is the yaw angle. If ΔX and θ_{yaw} can be estimated correctly, then through shift and tilt, ping 2 can be aligned to ping 1. This is implemented by adding a certain time delay to the received signal of the relevant receiver of ping 2 and aligning each receiver's signal of ping 2 to the corresponding receiver's signal of ping 1.

II. METHODOLOGY

A. Motion model

The proposed sonar accuracy enhancement is based on averaging multiple echo returns obtained from multiple sonar pings. Such averaging can effectively increase accuracy if the relative motion of the sonar and targets are compensated. After motion compensation, the echoes will be aligned, to give an effective stationary relation between the sonar and the targets. We estimate the relative motion between echoes of the first ping and all the other pings so as to align all pings to the first ping.

Similar assumptions hold:[11] (1) The medium is stable during pinging (i.e., there is good temporal correlation between pings) so that shifts in the cross-correlation of the positions of the sensors can be interpreted as specific displacements of the receiver array; (2) the "stop and hop" scenario where the sonar is assumed to be stationary during the period between transmitting at a particular position and receiving all of the echoes before next ping transmission. Since all the sensors are mounted on a rigid frame, the sensors move as a rigid body. That is to say, the sensors move as a line and the displacement between any two pings also is a hyperbolic curve which in first approximation can be considered a line. This follows from the assumption that the target range is much larger than the relative displacement during the burst of pings, which is referred to by sway and yaw changes.

The sway and yaw motion between two pings are depicted in Fig. 1. The sonar system model has one transmitting sensor and N receiving sensors where each sensor has an integer index $n \in [1, N]$ and the middle receiving sensor

(overlaid with transmitting sensor) is m. The sensors are equally separated by distance d. The displacement ΔX of sensor m is the sway between two pings. The angle θ_{yaw} of ping 2 with ping 1 is the yaw angle (we assume that the sensors of the first ping, i.e., ping 1 in Fig. 1 are all aligned in the direction perpendicular to the ping direction). If ΔX and θ_{yaw} can be estimated correctly, then through shift and tilt, ping 2 can be aligned to ping 1. This is implemented by adding a certain time delay to the received signal of the relevant sensor of ping 2 and aligning each sensor's signal of ping 2 to the corresponding sensor's signal of ping 1. Due to noise, the estimated displacements of the sensors are not exactly on a straight line, but are distributed around some certain line. Therefore, a robust line fitting method should be used to estimate this line. We have found that the least square line fitting (LSLF) method for estimating the line properties is not sufficiently robust in our case. This is due to the outliers in the displacement estimation which exist when noise level is high. We propose a more robust line fitting method—an improved mode line fitting (IMLF) method which bases on the mode of the distribution of the displacements. By getting the estimated fitting line of the displacements between sensors, the sway and yaw between two pings are estimated, respectively, by the center and the angle of the line.

For multiple pings, one can always set the first ping as the baseline. The relative motion between each ping to the first ping is estimated and each ping is aligned to the first one by correcting the relative motion. The aligned multiple pings are then used for an averaged time-delay estimation or image reconstruction. As the motion between the pings is corrected, the accuracy of the sonar should be improved with the increasing of the number of pings.

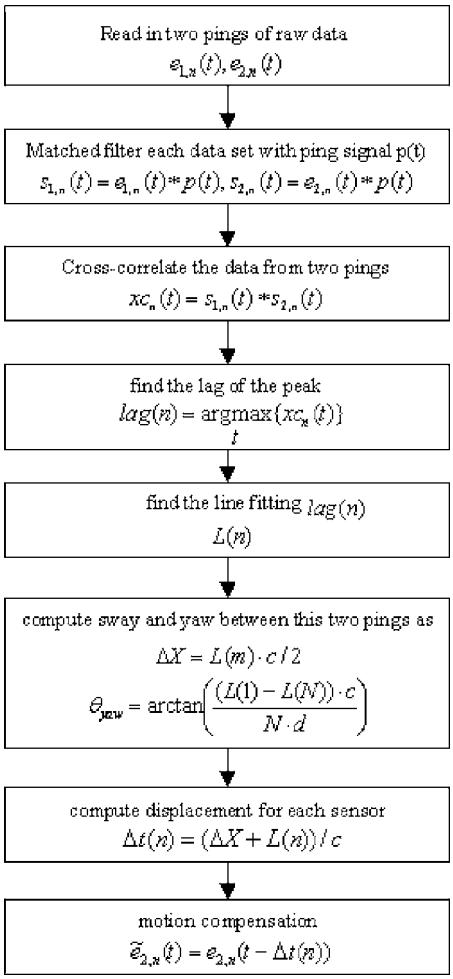


FIG. 2. Flow diagram for estimation and compensation of sway and yaw between two pings.

B. Motion estimation

The motion estimation approach is based on the raw data from one transmitting sensor and multiple receive sensors. Here one can consider any two pings. The displacement of each receiving sensor depends on the sway and yaw between the two pings. Thus, by following the process summarized in the flow diagram of Fig. 2, the actual sway and yaw required to align the two pings into coincidence can be estimated simultaneously. We denote the received raw echo data from sensor n of two pings as $e_{1,n}(t), e_{2,n}(t)$ respective to pings 1 and 2, where t is the delay time. The received echo is passed through a matched filter for cross correlation with the pinging signal

$$s_{1,n}(t) = e_{1,n}(t) * p(t),$$
$$s_{2,n}(t) = e_{2,n}(t) * p(t), \tag{1}$$

where $p(t)$ is the ping signal and $*$ defines cross-correlation in t. The cross-correlation of the outputs of the matched filter from two pings by the same sensor yields:

$$xc_n(t) = s_{1,n}(t) * s_{2,n}(t), \tag{2}$$

where n is the index of the sensor. The lag of the peak of the cross-correlation result is

$$\text{lag}(n) = \arg\max_t \{xc_n(t)\}, \tag{3}$$

which is the temporal displacement between two pings for sensor n. The collection of the displacement data for all sensors consists of a set of points $P=\{(1,\text{lag}(1)), (2,\text{lag}(2)), \ldots, (N,\text{lag}(N))\}$, where $1,2,\ldots,N$ is the index of the sensors (as the intervals of the sensors are equal, one can always use the index to substitute the real distance). Using a suitable line fitting method (which is introduced further below), we seek the best fitting line of P. We denote the value of the line at index n by $L(n)$, which is the estimated displacement between two pings of sensor n, for $n=1,2,\ldots,N$. The sway and yaw angle between two pings is computed as follows:

$$\Delta X = L(m) \cdot c/2, \tag{4}$$

$$\theta_{\text{yaw}} = \arctan\left(\frac{(L(1) - L(N)) \cdot c}{N \cdot d}\right), \tag{5}$$

where ΔX is the sway and θ_{yaw} is the yaw angle, m is the index of the middle sensor, c is the speed of sound, d is the space between adjacent sensors in the receiving array. Because the received data is a two-way echo, which is transmitted from the sonar pinging source to the targets

and returned from the targets to the receiving sensor, the time delay of the received echo between two pings for each sensor due to the displacement of the pings can be calculated by

$$\Delta t(n) = (\Delta X + L(n))/c, \tag{6}$$

where $\Delta t(n)$ is the time delay of the received echo data between two pings for sensor n, ΔX is the displacement of the transmitting sensor, $L(n)$ is the estimated displacement of the receiving sensor n. The motion compensation for each sensor received data of ping 2 is

$$\tilde{e}_{2,n}(t) = e_{2,n}(t - \Delta t(n)), \tag{7}$$

where $\tilde{e}_{2,n}(t)$ is the motion corrected data for sensor n of ping 2. After motion compensation implemented to each sensor's received echo data, ping 2 would be aligned to ping 1. The two pings are independent, so the scheme adapts to any pair of pings. Especially, for a set of multiple pings, using the scheme, all the pings will align to the first ping. The aligned multiple pings may be treated by coherent addition. In this case, the RMSE of the echo delay goes down by a factor of $\sqrt{n}$. As the motion between the pings is corrected, the accuracy of the sonar in a noisy environment will be greatly improved with the increasing of the number of pings.

C. Least square line fitting (LSLF) method

A popular line fitting method is the least square line fitting. In this section we demonstrate that the middle point of the least square fitting line segment of a set of points is equal to the mean of the points. Suppose there are a set of points $(x_i, y_i), i = 1, 2, \ldots, n$, the linear fitting curve through the points using least square method is

$$\tilde{y} = ax + b,$$

where,

$$a = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} x_i y_i\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)}{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} = \frac{\overline{xy} - \overline{xy}}{\overline{x}^2 - (\overline{x})^2}$$

$$b = \left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) - a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \overline{y} - a\overline{x}, \tag{8}$$

for each point, the fitting point should be

$$\tilde{y}_i = ax_i + b. \tag{9}$$

Let the middle point of the fitting line segment is (x_m, y_m) which can be expressed as

$$x_m = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \overline{x}, \tag{10}$$

$$y_m = \left(\frac{1}{n}\sum_{i=1}^{n} \tilde{y}_i\right) = \left(\frac{1}{n}\sum_{i=1}^{n} (ax_i + b)\right)$$

$$= a\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) + b = a\overline{x} + b = \overline{y}. \tag{11}$$

This is to say that the middle point of the least square fitting line segment of a set of points is equal to the mean of the set of points. Given a set of displacements of sensors for two pings which immersed in noise, using least square line fitting (LSLF) method to search the fitting line of the displacements and using the middle of the fitting line as the estimated sway of the pings is equivalent to using the mean of the displacements to estimate the sway. It is proved by Neretti *et al.*[3] that the mean of the distribution (before motion compensation) is not contributing at all to the resilience to noise. So the LSLF method is not suitable for a high noise environment.

D. Improved mode line fitting (IMLF) method

It is argued by Neretti *et al.*[3] that a more robust statistics method for estimating time-delay, such as the median, can improve the resiliency to noise as the number of pings increases. Best results are obtained by using the mode of distribution of the estimates from the multiple pings. In our case, using multiple pings is equivalent to using multiple sensors as the sensors lie on a straight line. The time delay estimation from multiple pings is equivalent to the sway estimation from multiple sensors' displacement. However, the mode can not be used directly. In the noise free case, although the displacements of the sensors are positioned on a straight line (a rigid moving of sensor array), the line can be tilted due to the yaw. In a noisy case, the displacements will be distributed around a tilted straight line with some possible outliers (Fig. 3). When the distribution of the echo displacements between any two pings is analyzed along a line that is parallel to the tilted sensor array, the observed distribution has minimal variance and the mode is a good estimation of the center of the displacements. In the tilt angle distribution case, we divide the range into intervals with the same width. Each interval forms a bin. In the simplest case, we define the mode by the bin which contains the largest number of points. A line through the mode with the tilt degree of the distribution can be the fitting line for the displacements. The full description of IMLF algorithm is given below.

Consider a set of points from 1 to N. Consider all the lines which go through at least two points (with some limitation on the tilt degree). For every line, we compute a bin with a certain width that is centered at the line. We further compute the number of points contained inside the bin. We choose the fitting line to the collection of echo returns as the line which contains the largest number of points in its bin.

The steps of the algorithm are:

step 0: Input a set of points $p(i) = (x(i), y(i))$, $i = 1 \ldots N$, N is the number of points, select an engaged gap e as the width of the bin which is centered around a selected line, select a degree $\theta_{\max}$ as the maximum tilt degree of a suitable line, set $k = 1$;

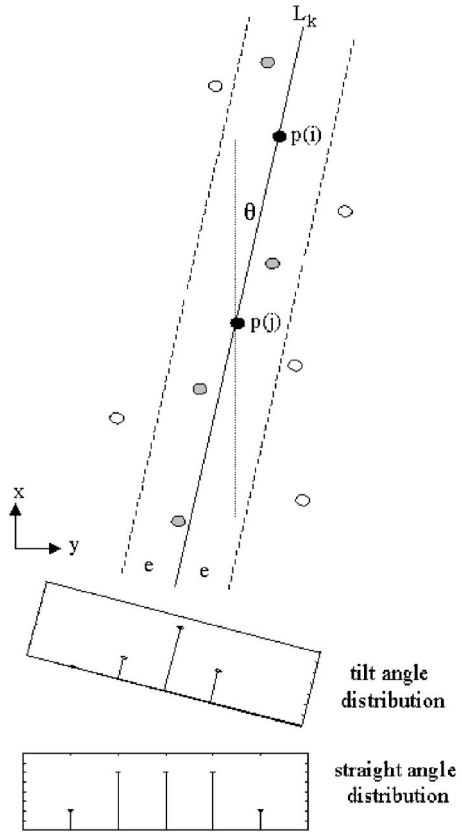


FIG. 3. The sketch of improved mode line fitting method. The sensor array is aligned along x-direction. The position along y-direction of each point expresses the displacement of each sensor's signal between two pings. Black points $p(i)$ and $p(j)$ are the selected pair of points. Solid line L_k is the fitting line through $p(i)$ and $p(j)$. The gap between two dashed line is the bin of L_k. The distance e between solid line and dashed line is the width of the bin. θ is the tilt degree between line L_k and x-direction. Gray points are the points inside the bin. Circle points are the points outside the bin. The histograms of the number of points in the bin are shown at the bottom for tilt angle distribution and straight angle distribution, respectively.

step 1: Select a pair of points $\{p(i),p(j)\}$, $i,j=1,\ldots,N$, $i\neq j$;

step 2: Compute the line l through the point pair $\{p(i),p(j)\}$;

step 3: Calculate the tilt degree θ between line l and x direction, if $abs(\theta)>\theta_{\max}$, then ignore line l, go to step 1, else keep line l and put l in the suitable line set as L_k, continue step 4;

step 4: Search the points whose distance to line L_k is less than e, save the number of these points as $num(k)$, $k=k+1$;

step 5: Iterate steps 1–4 for all the point pairs;

step 6: Search the line L_K which contains the maximum number of points as $K=\arg\max_k\{num(k)\}$;

step 7: Output line L_K as the fitting line of the point set p.

The sketch of this algorithm is shown in Fig. 2.

III. SIMULATION AND RESULTS

A. System setup

A set of simulations using Field II[13] has been performed for investigating the accuracy of the line fitting algorithm and the sonar echo delay accuracy following motion compensation based on several line fitting algorithms. We used one transmitter and a linear array of 10 receivers. The pinging signal was created by timing a hanning window on a sinusoidal signal. It was a Mexican hat at a center frequency of 300 kHz with a bandwidth of 200 kHz and a 6.7 μs duration. Both the pinging signal and it's auto-correlation had a sharp peak and two symmetric negative part with lower magnitude. The repetition rate is five pings per second which simulates a target that is within a 300 meters range. We note that it is possible to increase the effective range while maintaining the burst duration by changing the pinging signal during the burst and applying multiple matched filters on the returns. The target was a point reflector surrounded by scatterers. For each ping, we simulated a random motion (sway and yaw) of the receivers array. For simplicity, the motion was considered relative to the array location at the first ping. The sway of the motion was modeled by a uniformly distributed displacements in the range $[-2\lambda,2\lambda]$. where λ is the wave length of the pinging signal. The yaw of the motion was uniformly distributed random degree between $[-2^\circ,2^\circ]$. White Gaussian noise was added to each sensors' received signal. The SNR was measured in decibels as a function of the ratio between the total energy of the echo (measured in Ws), and the spectral density of the noise (measured in W/Hz). Using this measurement of SNR, the spectrum of the noise which actually influences the detection of the signal becomes equal to the spectrum of the signal. The received data of each sensor which was contaminated by noise was the raw data. The raw data was cross-correlated with the pinging signal to reduce noise and output as the preprocessed data. We estimated the motion of the pings using LSLF and IMLF methods. The estimated sway and yaw were compared to the simulated sway and yaw and the errors were recorded. For multiple pings, the root mean square error of motion estimation was calculated as a function of SNR. The sampling rate of the system was 100 MHz.

B. Sway and yaw estimation error

To test the sway and yaw estimation procedure, the sway and yaw between any ping and the first ping was calculated. For the test ping, the preprocessed data from each sensor was cross-correlated with the preprocessed data from the same sensor for the first ping. The lag of the peak of the cross-correlated result was computed. For 10 sensors, there were 10 lags. They were the relative displacement of each sensor between the test ping and the first ping. The lags were applied to the LSLF and IMLF algorithm, respectively, to find the fitting line. The center of the line and the tilt degree of the line were the estimated sway and yaw between the test ping and the first ping, respectively. For the sway estimation, we also calculated the median of the lags as the estimated sway (MedLF). The error between the estimated sway and yaw to the actual sway and yaw was computed. 100 pings were repeated and the errors of sway and yaw estimation were recorded. The root mean square error (RMSE) of 100 pings was calculated for the sway and yaw, respectively. The procedure was implemented for different SNRs. The sway errors and yaw errors are plotted in Figs. 4(a) and 4(b).

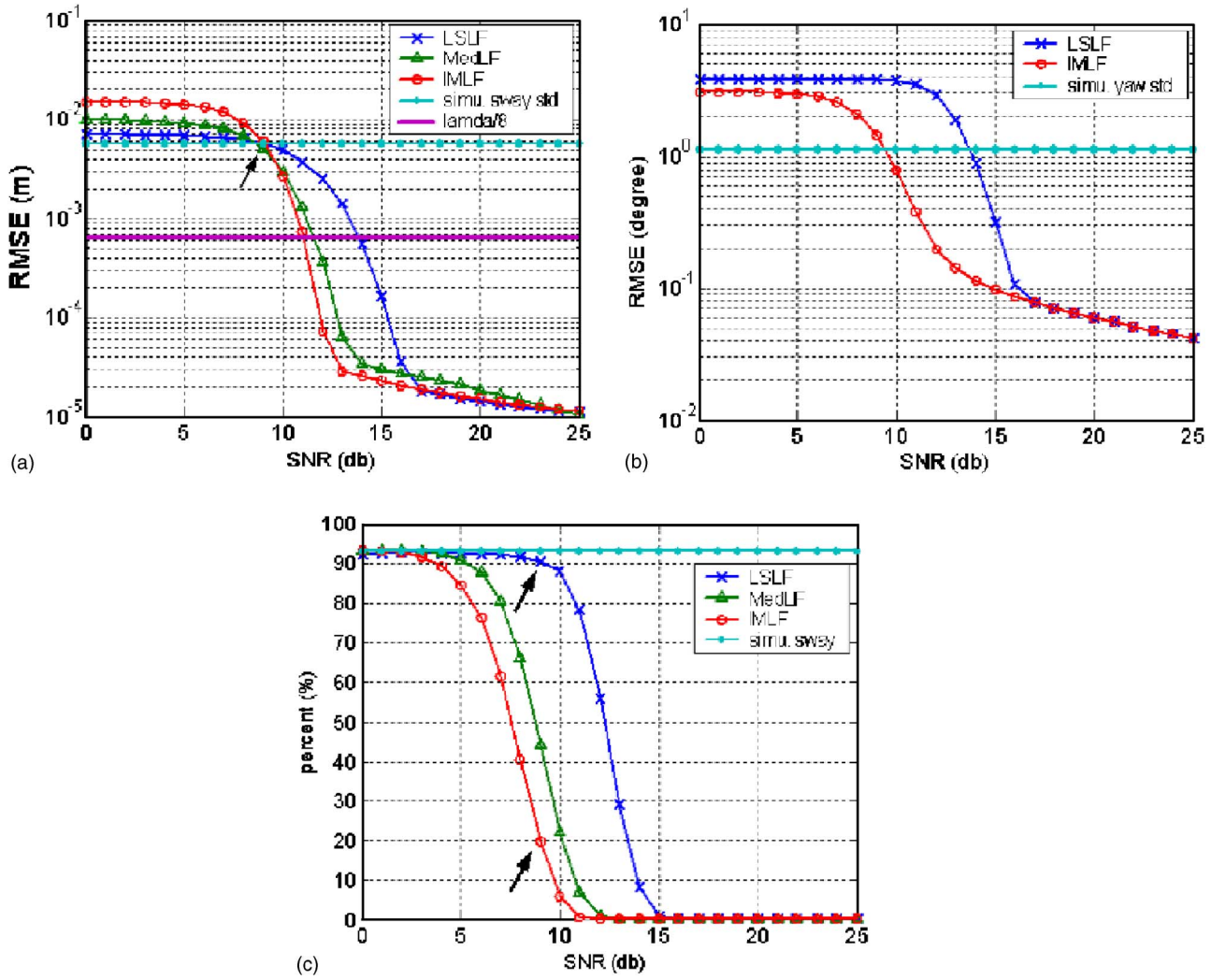


FIG. 4. (Color online) Comparison of sway and yaw estimation methods. (A) Sway estimation error (RMSE) (*m*) vs SNR. Each point is the root mean square error of sway estimation. The error is between the estimated sway to the simulated sway. (B) Yaw estimation error (RMSE) (degree) vs SNR. Each point is the root mean square error of yaw estimation. The error is between the estimated yaw to the simulated yaw. (C) Sway estimation error (percentage of pings whose error over $\lambda/8$) (%) vs SNR. Each point depicts the percentage of the pings, whose sway estimation error is over $\lambda/8$. In each plot, the blue cross curve shows the result using LSLF estimation method and the red circle curve shows the result using IMLF estimation method. The green triangle curve in plot (A) shows the result using MedLF estimation method to estimate sway. The cyan point lines are the standard deviation of the simulated sway and yaw distribution and the percentage for the simulated sway over $\lambda/8$, respectively, in plots (A)–(C), they are independent of SNR. The magenta line in plot (A) shows the position of $\lambda/8$, $\lambda/8=6.4\times10^{-4}$ m, λ is the wave length of the ping signal. The RMSE and percentage are calculated for SNR from 0 db to 25 db.

Figure 4(a) depicts the sway estimation error (RMSE) (*m*) versus SNR. LSLF (blue cross curve), MedLF (green triangle curve), and IMLF (red circle curve) estimation error are shown and it indicates that the best sway estimation is obtained by the proposed IMLF method. The IMLF has the smallest estimation error when SNR is higher than 10 db and also has the smallest breakpoint at 13 db as the breakpoints of MedLF and LSLF are at 14 and 17 db, respectively. Also in Fig. 4(a), the blue, green, and red curves are above the cyan curve which is the standard deviation of the simulated sway when SNR is lower than 8 db. This means that any motion estimation algorithm does not obtain the correct estimation if SNR is lower than 8 db. Figure 4(b) depicts the yaw estimation errors by using LSLF (blue cross curve) and IMLF (red circle curve) estimation, respectively. It shows that the IMLF is better than LSLF. The breakpoint of IMLF is 13 db and LSLF is 16 db. Meanwhile, the RMSE of IMLF is over the standard deviation of the simulated yaw [cyan line in Fig. 4(b)] at 9 db, but the LSLF is at 13 db. From Figs. 4(a) and 4(b), it shows that both for sway and yaw estimation in the low noise region (SNR is larger than 16 db) IMLF and LSLF have the same smaller estimation error, in the moderate high noise region (SNR is between 16 and 9 db) IMLF has smaller RMSE than LSLF, finally in the much higher noise region (SNR is smaller than 9 db) each method can not give estimation better than simulated motion. This follows that the IMLF algorithm is more accurate and reliant to noise than LSLF algorithm for both the sway and yaw estimation.

We further propose another method to evaluate the motion error sense. It is discussed by Cutrona[14] that if the motion perturbations is held to within 1/8 of a wavelength then the influence of the motion will be tolerant. Thus after motion estimation, among the multiple pings, only those pings whose motion estimation errors are over $\lambda/8$ will corrupt the fusion of multiple pings. The more pings of this kind exist,

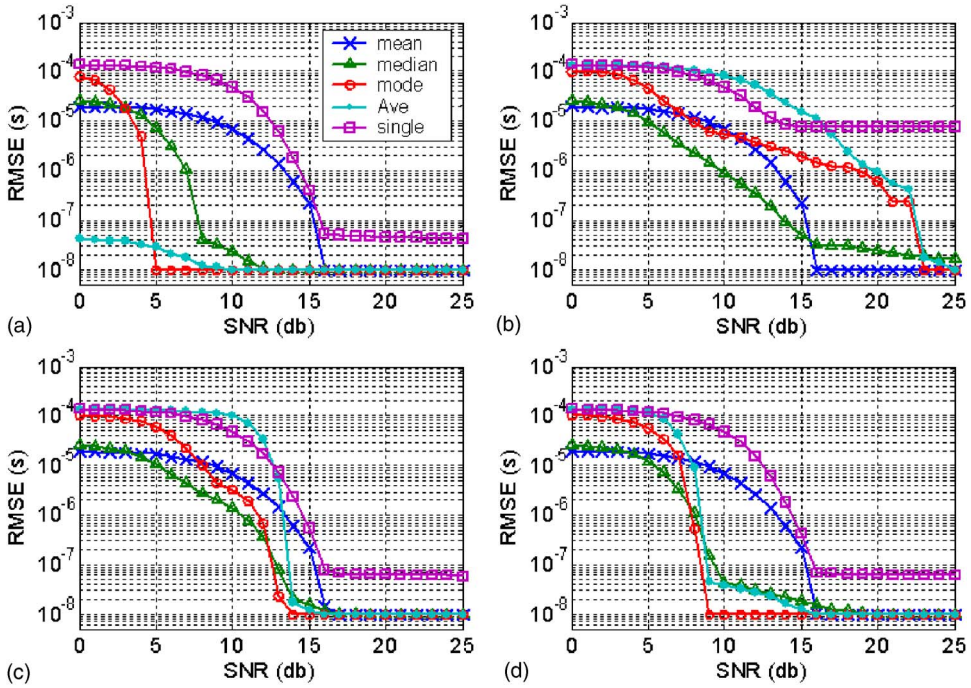


FIG. 5. (Color online) Comparison of the motion corruption and the improvement after motion estimation for time delay estimation accuracy. Plot (A) is the ideal case where there is no motion between pings. Plot (B) is the worst case where the motion between pings is not corrected. Plot (C) is a moderate case where LSLF motion estimation is used to correct the motion between pings. Plot (D) is the best case where IMLF motion estimation is used to correct the motion. In each plot, time-delay estimation error (RMSE) (s) as a function of SNR is shown for different time-delay estimation methods. Each point is the root mean square error of time-delay estimation. The RMSE is calculated for SNR form 0 to 25 db. The magenta square curve is the estimation from a single ping (single). The blue cross curve, green triangle curve, and red circle curve are the estimation based on the temporal mean (mean), temporal median (median) and temporal mode (mode) of multiple pings' estimation, respectively. The cyan point curve is the result of using the analog average (Ave) of multiple pings to estimate the time delay.

the worse the corruption will be. We calculate the percentage of the pings whose errors are over $\lambda/8$ after motion estimation. Figure 4(c) shows the percentage of the sway error as a function of SNR for LSLF and IMLF algorithm. It can be seen that more than 93% of the simulated sways (cyan point curve) are over $\lambda/8$. This is because we suppose the sway is uniformly distributed in $[-2\lambda, 2\lambda]$. For IMLF algorithm, almost all the sway errors are within $\lambda/8$ when SNR higher than 11 db. And for LSLF algorithm, it needs SNR larger than 15 db to ensure most of the sway errors are within $\lambda/8$. However, if comparing IMLF sway estimation error at 9 db SNR by RMSE and percentage [arrows in Figs. 4(a) and 4(c)], one can see that although RMSE of IMLF is over the std of simulated sway in Fig. 4(a), there are nearly 80% pings of IMLF whose errors are within $\lambda/8$ in Fig. 4(c) (the other 20% pings' sway estimation are far from the simulated sway so as to causing a large RMSE). But for the LSLF algorithm, although it has less RMSE than IMLF [Fig. 4(a) arrow], but there is only less than 10% pings whose errors are within $\lambda/8$. The correctly estimated pings by IMLF are 70% more than by LSLF. This observation will contribute improvement by IMLF estimation to the accuracy of multiple pings fusion which is discussed below.

C. Time-delay estimation

The next simulation aims at demonstrating the improvement of sonar accuracy for multiple pings fusion resulting by the improved motion estimation. The setup for the sonar array was the same as before. 50 pings were transmitted and the received signals were contaminated by noise for 10 sensors separately and corrupted by motion for each ping. The preprocessing and motion estimation were implemented using LSLF and IMLF algorithm, respectively. The relative motion between any ping and the first ping were corrected. All the pings were aligned to the first ping. The time delay of the target was estimated for each ping. We used multiple pings fusion methods introduced by Neretti *et al.*[3] to estimate the time-delay by multiple pings. The estimation by the temporal mean, the temporal median and the temporal mode of multiple pings' observation were computed. The estimation by using single ping and the analog average of multiple pings were also calculated. The procedure was repeated by 200 times and RMSE of the time delay estimation as a function of SNR was computed for each time delay estimation method. The influence of motion and motion estimation to the time delay error is depicted in Fig. 5.

Figure 5 plots the Woodward curve of the delay estimation error (RMSE) (s) as a function of SNR and delay estimation methods. Figure 5(a) shows the ideal case where there is no motion between pings. As one can predict, the analog average (Ave) of multiple pings (cyan point curve) gives the best time delay estimation due to the reduction of noise. In this case, the temporal mode (mode) of multiple pings (red circle curve) shows more robust statistics than the temporal median (median) of multiple pings (green triangle curve). Using the simple temporal mean (mean) of multiple pings (blue cross curve) does not show significant improvement in the resiliency to noise comparing with using the

single ping's (single) estimation (magenta square curve). On the other side, the worst case is shown in Fig. 5(b), where a random motion presented between pings and not any motion estimation is used to correct it. In this case, the estimation error of a single ping at higher SNR (larger than 14 db) is larger than that is in the ideal case. The jump gap is the standard deviation of the simulated motion between pings. Using the analog average of multiple pings will eliminate the gap at higher SNR, but will not improve noise resiliency. This is due to the fact that although the average of multiple pings reduces the effect of noise, at the same time the signal power also averaged because there is no signal overlay between pings due to the motion. SNR will not be improved by using average multiple pings. In some instances (such as a large motion), SNR even becomes lower after averaging. In this case, the temporal mode and the temporal median produce better results while the temporal mean is not changed, this is because we assume that the motion is uniformly distributed and the mean of motion is zero.

Figures 5(c) and 5(d) show the improvement by using LSLF and IMLF motion estimation, respectively, to correct the motion before using multiple ping fusion to estimate the time delay. Figure 5(c) is the moderate result after using LSLF to correct the sway and yaw. The breakpoints of the temporal mode function and the analog average function are reduced to 14 db. This attributes to the fact that the LSLF can correct the motion of most pings within $\lambda/8$ when SNR is above 14 db [Fig. 4(c)]. Figure 5(d) shows the accuracy after using IMLF to correct the motion. The IMLF motion estimation makes an evident improvement of the time delay estimation accuracy. Due to the fact that most pings are within $\lambda/8$ after motion compensation [Fig. 4(c)], using the temporal mode function of multiple pings' fusion, the breakpoint is decreased to 9 db. The RMSE of IMLF [in Fig. 5(d)] is smaller than that of LSLF [in Fig. 5(c)] when the SNR is above the 9 db for any time delay estimation method (mean, median, mode, and ave). This verifies that IMLF gives better motion correction than LSLF. It demonstrates that IMLF method can estimate the motion correctly in a high noise environment. Subsequently both the accuracy and the resilience to noise are improved by the fusion of multiple pings after realignment according to the motion estimation.

Also notice that the breakpoint of the mode using IMLF in Fig. 5(d) is 9 db, not 13 db in Fig. 4(a) for the sway estimation by IMLF. This attributes to the fact that, after using IMLF motion estimation, there are almost 80% pings which are aligned to the first ping with an error less than $\lambda/8$. Although RMSE of all the estimated pings are very high, there are also 80% pings which are correctly estimated and can contribute to the accuracy of time delay estimation using multiple pings mode fusion function. This is in accordance with the study of Fig. 4(c) which demonstrates that IMLF algorithm can estimate the motion correctly even when very strong noise (9 db SNR) is present.

IV. CONCLUSIONS

This paper presented a method for robust motion estimation during the transmission of a burst of pings. It then demonstrated improved sonar accuracy and resiliency to noise when the motion estimation was utilized within a multiple ping sonar system. The motion estimation was based on the rigid motion property of the sonar, which enabled analysis of the motion using the multisensor signals. It then applied an improved mode line fitting algorithm (IMLF) which provided a robust estimation of the motion in the presence of noise and estimated sway and yaw between two pings. As the estimated motion is relative between the sonar and the target, our results still hold when the target is moving (relative to the sonar) at a slow speed (in a direction perpendicular to the sonar). The motion in a direction that is not perpendicular will be discussed elsewhere. We demonstrated that a robust algorithm based on multisensors, such as IMLF, significantly decreases the signal-to-noise ratio breakpoint for motion estimation. We have further demonstrated that it is possible to take advantage of the motion estimation, to compensate the motion during transmission of a burst of multiple pings. Specifically, fusing that information (in few methods) improves the accuracy and the resiliency of time-delay estimation to background noise.

Using a burst of 50 pings, we have shown that it is possible to estimate an adequate motion and then fuse the information from the multiple pings to get good time delay estimation up to a break point at 9 db when the single ping estimation breaks at 16 db.

[1] P. Woodward, "Theory of radar information," IEEE Trans. Inf. Theory **1**(1), 108–113 (1953).

[2] R. J. McAulay and L. P. Seidman, "A useful form of the Barankin lower bound and its application to PPM threshold analysis," IEEE Trans. Inf. Theory **IT-15**, 273–279 (1969).

[3] N. Neretti, N. Intrator, and L. N. Cooper, "Pulse-train based time-delay estimation improves resiliency to noise," IEEE International Workshop on Machine Learning for Signal Processing 213–222 (Brazil, 2004).

[4] J. M. Silkaitis, B. L. Douglas, and H. Lee, "A cascade algorithm for estimating and compensating motion error for synthetic aperture sonar imaging," IEEE Processing ICIP'94 **1**, 905–909 (1994).

[5] G. A. Shippey, P. Ulriksen, and Q. Liu, "Quasi-barrowband processing of wideband sonar echoes," Proc. of the 4th European Conf. on Underwater Acous. (ECUA), Rome, Italy, 63–68 (1998).

[6] K. A. Johnson, M. P. Hayes, and P. T. Gough, "Estimating sub-wavelength sway of sonar towfish," IEEE J. Ocean. Eng. **20**(4), 258–267 (1995).

[7] H. J. Callow, "Signal processing for synthetic aperture sonar image enhancement," (Ph.D. dissertation, Department of Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand, April 2003).

[8] A. Bellettini and M. A. Pinto, "Experimental investigation of synthetic aperture sonar micronavigation," Proc. 5th European Conf. on Underwater Acous. (ECUA), Lyon, France, 445–450 (2000).

[9] J. T. Christoff, "Motion compensated high-frequency synthetic aperture sonar," J. Acoust. Soc. Am. **1**, 2950 (1998).

[10] V. Tonard and M. Brussieux, "Towards development of autofocusing schemes for phase compensation of synthetic aperture sonars," Proc. of the 1997 Oceans Conference **2**, 803–808 (1997).

[11] P. T. Gough and M. A. Miller, "Displaced ping imaging autofocus for a multi-hydrophone SAS," IEE Proc., Radar Sonar Navig. **151**(3), 163–170 (2004).

[12] C. Yuan, M. R. Azimi-Sadjadi, J. Wilbur, and G. J. Dobeck, "Underwater target detection using multichannel subband adaptive filtering and high-order correlation schemes," IEEE J. Ocean. Eng. **25**(1), 192–205 (2000).

[13] J. A. Jensen, "User's guide for the Field II program, Version 2.72," Technical Report, Department of Information Technology, Technical University of Denmark, May 26, 1999.

[14] L. J. Cutrona, "Comparison of sonar system performance achievable using synthetic aperture techniques with the performance achievable by more conventional means," J. Acoust. Soc. Am. **58**(2), 336–348 (1975).

Elastic moduli approximation of higher symmetry for the acoustical properties of an anisotropic material

Andrew N. Norris[a)]
Department of Mechanical and Aerospace Engineering, Rutgers University, 98 Brett Road, Piscataway, New Jersey 08854-8058

(Received 5 November 2005; revised 19 January 2006; accepted 20 January 2006)

The issue of how to define and determine an optimal acoustical fit to a set of anisotropic elastic constants is addressed. The optimal moduli are defined as those which minimize the mean-squared difference in the acoustical tensors between the given moduli and all possible moduli of a chosen higher material symmetry. The solution is shown to be identical to minimizing a Euclidean distance function, or equivalently, projecting the tensor of elastic stiffness onto the appropriate symmetry. This has implications for how to best select anisotropic constants to acoustically model complex materials. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2173525]

PACS number(s): 43.35.Cg, 43.20.Jr [YHB] Pages: 2114–2121

I. INTRODUCTION

The theory of wave propagation in anisotropic elastic solids was historically developed for homogeneous crystals with *a priori* known symmetry.[1] The number of elastic moduli are then well defined. However, the symmetry is not always apparent in noncrystalline and generally complex materials, for which measurements can yield as many as 21 anisotropic elastic coefficients. Whether one has some idea of the underlying material symmetry or simply prefers to deal with fewer parameters, the question arises of how to best fit the given set of elastic moduli to, for instance, a transversely isotropic material model. This issue occurs in acoustical measurements of composites composites,[2] and in geophysical applications[3] where laboratory measurements might yield 21 moduli, but seismic modeling requires a higher symmetry, such as transverse isotropy. The purpose of this paper is to provide a simple but unambiguous means to find the reduced set of anisotropic elastic constants that is in a certain sense the best acoustical fit to the given moduli. The optimal material minimizes the mean-square difference in the slowness surfaces of the given moduli and of all possible sets of elastic constants of the chosen symmetry.

The prevailing approach to finding a reduced set of moduli does not invoke acoustical properties, but views the moduli as elements in a vector space which are projected onto the higher elastic material symmetry. This is achieved by defining a Euclidean norm for the moduli $\mathbb{C}$ according to $\|\mathbb{C}\|^2 = C_{ijkl}C_{ijkl}$, where C_{ijkl} are the elements of the stiffness tensor. Lower case Latin suffices take on the values 1, 2, and 3, and the summation convention on repeated indices is assumed here and in Eq. (A1). This provides a natural definition for distance, from which one can find the elastic tensor of a given symmetry nearest to an anisotropic elastic tensor, or equivalently, define a projection appropriate to the higher symmetry. Gazis *et al.*[4] outline the procedure in terms of fourth-order tensors, while more recently Browaeys and Chevrot[3] provide projection matrices for $\mathbb{C}$ expressed as a 21-dimensional vector. Helbig[5] provides a useful overview of the problem from a geophysical perspective, Cavallini[6] examines isotropic projection specifically, and Gangi[7] gives formulas for several other symmetries. One drawback of the approach is that, while it minimizes the Euclidean distance between the original and projected stiffness tensors, it does not provide the analogous closest *compliance* (the inverse of stiffness). In this sense the stiffness projection method, while simple and attractive, is unsatisfactory because it is not invariant under inversion. Alternative procedures based on non-Euclidean norms such as the Riemannian[8] or log-Euclidean[9] metrics do not have this deficiency, and, in principle, provide a unique "projection" regardless of whether one uses the stiffness or the compliance tensor.

An apparently quite different approach is to try to find higher-symmetry moduli which in some way better approximate the acoustic properties of the given moduli. Thus, Fedorov[10] considered the question of what elastically *isotropic* material is the best acoustic fit to a given set of anisotropic moduli. He defined best fit to mean the effective bulk and shear moduli $\{\kappa, \mu\}$ which minimize the mean-square difference between the slowness surfaces of the original anisotropic material and the isotropic material characterized by $\{\kappa, \mu\}$ (density is unaffected). Fedorov obtained explicit expressions for the moduli, Eq. (26.19) of Ref. 10, or in the present notation

$$\kappa = \tfrac{1}{9}C_{iijj}, \quad \mu = \tfrac{1}{10}C_{ijij} - \tfrac{1}{30}C_{iijj}. \tag{1}$$

Fedorov's procedure for finding a suitable set of higher-symmetry moduli is physically appealing, especially as it seeks to approximate acoustical properties. Also, as Fedorov and others[10,4,3] have shown, κ and μ are precisely the isotropic moduli found by the stiffness projection method. However, Fedorov only considered effective isotropic moduli and it does not appear that anyone has attempted to generalize his method to symmetries other than isotropic. The purpose of this paper is to solve what may be termed the generalized Fedorov problem. The solution is found for the acoustically best-fitting moduli of arbitrary symmetry, which is of higher symmetry than the given moduli but lower than isotropy. The

a)Electronic mail: norris@rutgers.edu

 0001-4966/2006/119(4)/2114/8/$22.50

central result is that the solution of the generalized Fedorov problem possesses the same crucial property as the solution obtained by Fedorov,[10] that is, the best-fit moduli are identical to those obtained by the stiffness projection method. This result provides a strong physical and acoustical basis for using the Euclidean projection scheme that has been absent until now.

We begin in Sec. II with the definition of elastic tensors and associated notation. The generalized Fedorov problem is introduced and solved in Sec. III. Examples are given in Sec. IV.

II. PRELIMINARIES

A. The elasticity tensor and related notation

The solution of the generalized Fedorov problem is most easily accomplished using tensors, which are reviewed here along with some relevant notation. Boldface lower case Latin quantities indicate three-dimensional vectors, such as the orthonormal basis $\{\mathbf{e}_i, i=1,2,3\}$, open-face symbols such as $\mathbb{C}$ indicate fourth-order elasticity tensors, and boldface capitals, e.g., $\mathbf{A}$, are second-order symmetric tensors, with some exceptions. Components defined relative to the basis vectors allow us to represent arbitrary tensors in terms of the fundamental tensors formed from the basis vectors, thus, $\mathbf{A}=A_{ij}\mathbf{e}_i\otimes\mathbf{e}_j$, $\mathbb{C}=C_{ijkl}\mathbf{e}_i\otimes\mathbf{e}_j\otimes\mathbf{e}_k\otimes\mathbf{e}_l$, where $\otimes$ is the tensor product. The identity tensors of second and fourth order are $\mathbf{I}$ and $\mathbb{I}$, respectively, and we will also use the fourth-order tensor $\mathbb{J}$. These have components

$$I_{ij}=\delta_{ij},\quad I_{ijkl}=\tfrac{1}{2}(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}),\quad J_{ijkl}=\tfrac{1}{3}\delta_{ij}\delta_{kl}. \tag{2}$$

Isotropic elasticity tensors can be expressed as a linear combination of $\mathbb{I}$ and $\mathbb{J}$. Using Lamé moduli μ and $\lambda=\kappa-\frac{2}{3}\mu$, for instance,

$$\mathbb{C}=2\mu\mathbb{I}+3\lambda\mathbb{J}. \tag{3}$$

Products of tensors are defined by summation over pairs of indices: $(\mathbb{C}\mathbf{A})_{ij}=C_{ijkl}A_{kl}$ and $(\mathbb{A}\mathbb{B})_{ijkl}=A_{ijpq}B_{pqkl}$. Thus, $\mathbb{I}\mathbf{I}=\mathbb{J}\mathbf{I}=\mathbf{I}$, $\mathbb{I}\mathbb{I}=\mathbb{I}$, $\mathbb{J}\mathbb{J}=\mathbb{J}$, and $\mathbb{I}\mathbb{J}=\mathbb{J}\mathbb{I}=\mathbb{J}$.

The inner product of a pair of tensors of the same order is defined

$$\langle u,v\rangle=\mathrm{tr}(uv), \tag{4}$$

where tr $\mathbf{A}=A_{ii}$ and tr $\mathbb{A}=A_{ijij}$, and the Euclidean norm of a tensor is

$$\|u\|\equiv\langle u,u\rangle^{1/2}. \tag{5}$$

This will be used to compare differences between tensors.

An anisotropic elastic stiffness tensor $\mathbb{C}$ relates stress $\mathbf{T}$ and strain $\mathbf{E}$ according to the generalized form of Hooke's law,

$$\mathbf{T}=\mathbb{C}\mathbf{E}. \tag{6}$$

Stress and strain are symmetric second-order tensors, implying that their components are also symmetric, $T_{ij}=T_{ji}$, $E_{ij}=E_{ji}$, which in turn implies the first two of the following identities:

$$C_{ijkl}=C_{ijlk},\quad C_{ijkl}=C_{jikl},\quad C_{ijkl}=C_{klij}. \tag{7}$$

The third property is a consequence of the existence of a positive strain energy of the form $W=\frac{1}{2}\langle\mathbf{E},\mathbb{C}\mathbf{E}\rangle$, which also constrains the moduli to be positive definite. The squared norm of the elasticity tensor is

$$\|\mathbb{C}\|^2\equiv\langle\mathbb{C},\mathbb{C}\rangle=C_{ijkl}C_{ijkl}. \tag{8}$$

The 21-dimensional space of elasticity tensors can be decomposed as a 15-dimensional space of totally symmetric fourth-order tensors plus a six-dimensional space of asymmetric fourth-order tensors,[11,12] by

$$\mathbb{C}=\mathbb{C}^{(s)}+\mathbb{C}^{(a)}, \tag{9}$$

where

$$C^{(s)}_{ijkl}=\tfrac{1}{3}(C_{ijkl}+C_{ilkj}+C_{ikjl}). \tag{10}$$

The elements of the totally symmetric part satisfy the relations $C^{(s)}_{ijkl}=C^{(s)}_{ikjl}$ in addition to (7). Thus, $C^{(s)}_{ijkl}$ is unchanged under any rearrangement of its indices, and $\mathbb{C}^{(s)}$ has at most 15 independent elements. This can be seen by the explicit representation of the asymmetric part in terms of the remaining six independent elements, which are the components of the symmetric second-order tensor

$$D_{ij}=C_{ijkk}-C_{ikjk}. \tag{11}$$

Thus,

$$C^{(a)}_{ijkl}=\tfrac{1}{3}[2D_{ij}I_{kl}+2D_{kl}I_{ij}-D_{ik}I_{jl}-D_{il}I_{jk}-D_{jk}I_{il}-D_{jl}I_{ik}+D_{mm}(I_{ijkl}-3J_{ijkl})]. \tag{12}$$

An asymmetric tensor satisfies, by definition,[11]

$$C^{(a)}_{ijkl}+C^{(a)}_{ilkj}+C^{(a)}_{ikjl}=0. \tag{13}$$

Note that the symmetric part of an asymmetric tensor is zero, and vice versa, and that the decomposition (9) is orthogonal in the sense of the Euclidean norm, $\|\mathbb{C}\|^2=\|\mathbb{C}^{(s)}\|^2+\|\mathbb{C}^{(a)}\|^2$. The partition of $\mathbb{C}$ as a sum of totally symmetric and traceless asymmetric tensors is the first step in Backus' harmonic decomposition of elasticity tensors,[11] a partition that has proved useful for representations of elastic tensors[13] and has also been used to prove that there are exactly eight distinct elastic symmetries.[14] Finally, for future reference, note that the totally symmetric part of the fourth-order identity is

$$I^{(s)}_{ijkl}=\frac{1}{3}(\delta_{ij}\delta_{kl}+\delta_{ik}\delta_{jl}+\delta_{il}\delta_{kj})\Leftrightarrow\mathbb{I}^{(s)}=\frac{2}{3}\mathbb{I}+\mathbb{J}. \tag{14}$$

B. The acoustical tensor and the tensor $\mathbb{C}^*$

The acoustical tensor, also known as Christoffel's matrix,[1] arises in the study of plane crested waves with displacement of the form $\mathbf{u}(\mathbf{x},t)=\mathbf{b}h(\mathbf{n}\cdot\mathbf{x}-vt)$, where $\mathbf{b}$ is a fixed unit vector describing the polarization, $\mathbf{n}$ is the phase direction, also a unit vector, v is the phase velocity, and h is an arbitrary but sufficiently smooth function. Substituting this wave form into the equations of motion

$$C_{ijkl}\frac{\partial^2 u_k}{\partial x_j \partial x_l} = \rho \ddot{u}_i, \tag{15}$$

where ρ is the density, implies that it is a solution if and only if $\mathbf{b}$ and ρv^2 are eigenvector and eigenvalue of the second order tensor $\mathbf{Q}$,

$$Q_{ik}(\mathbf{n}) = C_{ijkl} n_j n_l. \tag{16}$$

This definition of the acoustical tensor is not the product of a fourth-order tensor with a second-order tensor. In order to express it in this form, which simplifies the analysis later, introduce the related fourth-order tensor $\mathbb{C}^*$ defined by

$$C^*_{ijkl} = \tfrac{1}{2}(C_{ikjl} + C_{iljk}). \tag{17}$$

Thus,

$$Q_{ij}(\mathbf{n}) = C^*_{ijkl} n_k n_l \Leftrightarrow \mathbf{Q}(\mathbf{n}) = \mathbb{C}^* \mathbf{n} \otimes \mathbf{n}. \tag{18}$$

The operation defined by * is of fundamental importance in solving the generalized Fedorov problem, and therefore some properties are noted. First, * is a linear operator that commutes with taking the symmetric and asymmetric parts of a tensor: $\mathbb{C}^{(s)*} = \mathbb{C}^{*(s)} = \mathbb{C}^{(s)}$ and $\mathbb{C}^{(a)*} = \mathbb{C}^{*(a)} = -\frac{1}{2}\mathbb{C}^{(a)}$. Accordingly, $\mathbb{C}^*$ is partitioned as

$$\mathbb{C}^* = \mathbb{C}^{(s)} - \tfrac{1}{2}\mathbb{C}^{(a)}, \tag{19}$$

and repeating the * operation n times yields $\mathbb{C}^{n*} = \mathbb{C}^{(s)} + (-1/2)^n \mathbb{C}^{(a)}$. Taking $n=2$ implies the identity

$$\mathbb{C} = 2\mathbb{C}^{**} - \mathbb{C}^*, \tag{20}$$

from which $\mathbb{C}$ can be found from $\mathbb{C}^*$. Hence the mapping $\mathbb{C} \leftrightarrow \mathbb{C}^*$ is one-to-one and invertible, that is bijective. This property is important in the inverse problem of determining elastic moduli from acoustic data.[15] Acoustic wave speeds and associated quantities are related primarily to $\mathbb{C}^*$ through the acoustical tensor also known as the Christoffel matrix,[1] and this can be determined uniquely from $\mathbb{C}$ using (20). Decompositions of $\mathbb{C}$ and $\mathbb{C}^*$ into totally symmetric and asymmetric parts are unique, and knowledge of one decomposition implies the other. In particular, $\mathbb{C} = \mathbb{C}^*$ if and only if the asymmetric parts of both are zero. This occurs if the moduli satisfy $C_{ijkk} = C_{ikjk}$, which together with the symmetries (7) are equivalent to the Cauchy relations; see Sec. 4.5 of Musgrave.[1] We note that the operation * is self-adjoint in the sense that the following is true for any pair of elasticity tensors:

$$\langle \mathbb{A}, \mathbb{B}^* \rangle = \langle \mathbb{A}^*, \mathbb{B} \rangle. \tag{21}$$

Elastic moduli are usually defined by the Voigt notation: $C_{ijkl} \equiv c_{IJ}$, where $I,J = 1,2,3,\ldots,6$ and $I = 1,2,3,4,5,6$ correspond to $ij = 11,22,33,23,13,12$, respectively, i.e.,

$$\mathbb{C} \leftrightarrow \mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ & & c_{33} & c_{34} & c_{35} & c_{36} \\ & & & c_{44} & c_{45} & c_{46} \\ S & Y & M & & c_{55} & c_{56} \\ & & & & & c_{66} \end{pmatrix}. \tag{22}$$

The components of $\mathbb{C}^*$ are then

$$\mathbb{C}^* \leftrightarrow \mathbf{C}^* = \begin{pmatrix} c_{11} & c_{66} & c_{55} & c_{56} & c_{15} & c_{16} \\ & c_{22} & c_{44} & c_{24} & c_{46} & c_{26} \\ & & c_{33} & c_{34} & c_{35} & c_{45} \\ & & & \frac{1}{2}(c_{44}+c_{23}) & \frac{1}{2}(c_{45}+c_{36}) & \frac{1}{2}(c_{46}+c_{25}) \\ S & Y & M & & \frac{1}{2}(c_{55}+c_{13}) & \frac{1}{2}(c_{56}+c_{14}) \\ & & & & & \frac{1}{2}(c_{66}+c_{12}) \end{pmatrix}. \tag{23}$$

III. FEDOROV'S PROBLEM FOR PARTICULAR SYMMETRIES

A. Definition of the problem

We assume as known the elasticity tensor $\mathbb{C}$ of arbitrary material symmetry with as many as 21 independent components. A particular material symmetry is chosen with prescribed symmetry axes or planes, for instance, transverse isotropy with symmetry axis in the direction $\mathbf{a}$, or cubic symmetry with orthogonal cube axes $\mathbf{a},\mathbf{b},\mathbf{c}$. Fedorov's problem for particular symmetries is to determine the elastic stiffness $\mathbb{C}_{\text{sym}}$ of the chosen material symmetry which is the best fit in the sense that it minimizes the orientation-averaged squared difference of the acoustical tensors. We introduce the acoustical distance function

$$f(\mathbb{C},\mathbb{C}_{\rm sym}) \equiv \frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \sin\theta\, \mathrm{d}\theta \,\|\mathbf{Q}(\mathbb{C},\mathbf{n}) - \mathbf{Q}(\mathbb{C}_{\rm sym},\mathbf{n})\|^2, \tag{24}$$

with $\mathbf{n}=\sin\theta(\cos\phi\mathbf{e}_1+\sin\phi\mathbf{e}_2)+\cos\theta\mathbf{e}_3$. The same function was considered by Fedorov[10] with $\mathbb{C}_{\rm sym}$ restricted to isotropic elasticity. Thus, substituting $\mathbb{C}_{\rm sym}=\alpha_1\mathbb{I}+\alpha_2\mathbb{J}$ into (24) one obtains, after simplification, a positive-definite quadratic in the two unknowns α_1 and α_2. A simple minimization yields $\alpha_1=2\mu$ and $\alpha_2=3\kappa-2\mu$, where κ and μ are defined in Eq. (1). The symmetry of $\mathbb{C}_{\rm sym}$ can be considered as arbitrary for the general problem addressed here, although Fedorov's isotropic result is recovered as a special case of the general solution discussed in Sec. IV.

The integral over directions in (24) can be removed using

$$\frac{1}{4\pi}\int_0^{2\pi} \mathrm{d}\phi \int_0^{\pi} \sin\theta\, \mathrm{d}\theta\, \mathbf{n}\otimes\mathbf{n}\otimes\mathbf{n}\otimes\mathbf{n} = \frac{1}{5}\mathbb{I}^{(s)}, \tag{25}$$

where $\mathbb{I}^{(s)}$ is defined in Eq. (14). The identity follows by noting that the integral must be a totally symmetric isotropic fourth-order tensor of the form $a\mathbb{I}^{(s)}$. Taking the trace of both sides and using tr $\mathbb{I}^{(s)}=I^{(s)}_{ijij}=5$, tr$(\mathbf{n}\otimes\mathbf{n}\otimes\mathbf{n}\otimes\mathbf{n})=1$, gives $a=1/5$. Thus, since

$$\mathbf{Q}(\mathbb{C},\mathbf{n}) = \mathbb{C}^*\mathbf{n}\otimes\mathbf{n}, \quad \mathbf{Q}(\mathbb{C}_{\rm sym},\mathbf{n}) = \mathbb{C}^*_{\rm sym}\mathbf{n}\otimes\mathbf{n}, \tag{26}$$

the distance function reduces to

$$f = \tfrac{1}{5}\langle(\mathbb{C}^*-\mathbb{C}^*_{\rm sym}),\, \mathbb{I}^{(s)}(\mathbb{C}^*-\mathbb{C}^*_{\rm sym})\rangle. \tag{27}$$

Define the modified inner product for elasticity tensors

$$\langle\mathbb{A},\mathbb{B}\rangle_a \equiv \mathrm{tr}(\mathbb{I}^{(s)}\mathbb{A}\mathbb{B}) = \langle\mathbb{A},\mathbb{I}^{(s)}\mathbb{B}\rangle, \tag{28}$$

and norm

$$\|\mathbb{A}\|_a = \langle\mathbb{A},\mathbb{A}\rangle_a^{1/2}. \tag{29}$$

Then, Fedorov's problem for particular symmetries amounts to

$$\text{Fedorov} \Leftrightarrow \text{minimize}\|\mathbb{C}^*-\mathbb{C}^*_{\rm sym}\|_a. \tag{30}$$

The reason the problem is expressed in this form is to make the connection with the notion of projection onto the chosen elastic symmetry, or equivalently of finding the elastic tensor of the chosen symmetry nearest to the given elasticity $\mathbb{C}$. As mentioned in the Introduction, this question has been addressed by several authors and has an explicit solution. The issue is to find the elastic tensor $\mathbb{C}_{\rm sym}$ which minimizes the Euclidean distance function d, where

$$d(\mathbb{C},\mathbb{C}_{\rm sym}) \equiv \|\mathbb{C}-\mathbb{C}_{\rm sym}\|. \tag{31}$$

Comparing Eqs. (30) and (31), the two problems appear tantalizingly similar. It should be realized that Fedorov's problem involves $\mathbb{C}^*$, not $\mathbb{C}$, and that the norms are distinct. However, it will be proved that the problems share the same solution,

$$\{\mathbb{C}_{\rm sym}:\min_{\mathbb{C}_{\rm sym}}\|\mathbb{C}^*-\mathbb{C}^*_{\rm sym}\|_a\} \equiv \{\mathbb{C}_{\rm sym}:\min_{\mathbb{C}_{\rm sym}}\|\mathbb{C}-\mathbb{C}_{\rm sym}\|\}, \tag{32}$$

which gives the central result of this paper, i.e.,

$$\text{Fedorov} \Leftrightarrow \text{Euclidean projection}. \tag{33}$$

This equivalence enables us to provide an explicit solution to the generalized Fedorov problem, e.g., using the methods of Gazis *et al.*,[4] Browaeys and Chevrot,[3] or others. The remainder of this section develops a proof by construction of the solution of the Fedorov problem, which is shown to be identical to the Euclidean projection. Some further concepts and notation are required and introduced next.

B. Basis tensors

The solution uses a fundamental decomposition of the chosen material symmetry using basis tensors. These form a vector space for the symmetry in the sense that any elasticity tensor of that symmetry may be expressed uniquely in terms of N linearly independent tensors $\mathbb{V}_1, \mathbb{V}_2, \ldots, \mathbb{V}_N$, where $2 \leqslant N \leqslant 13$ is the dimension of the vector space for the material symmetry. For instance, isotropic elasticity tensors are of the form $\mathbb{C}_{\rm iso}=\alpha_1\mathbb{I}+\alpha_2\mathbb{J}$, $\alpha_1,\alpha_2>0$. The procedure is analogous for other material symmetries, cubic, transversely isotropic, etc., and is described in detail by Walpole,[16] who provides expressions for the base tensors of the various symmetries. Thus, $N=2,3,5,9$ for isotropy, cubic symmetry, transverse isotropy, and orthorhombic symmetry, respectively. $N=13$ corresponds to monoclinic, which is the lowest symmetry apart from triclinic (technically $N=21$), which is no symmetry. The precise form of the basis tensors is irrelevant here (examples are given in Sec. IV, full details are in Ref. 16, and Kunin[17] develops a similar tensorial decomposition); all that is required is that they be linearly independent, and consequently any tensor with the desired symmetry can be written

$$\mathbb{C}_{\rm sym} = \sum_{i=1}^{N} \beta_i \mathbb{V}_i. \tag{34}$$

The coefficients follow by taking inner products with the basis tensors. Let $\mathbf{\Lambda}$ be the $N\times N$ symmetric matrix with elements

$$\Lambda_{ij} \equiv \langle\mathbb{V}_i,\mathbb{V}_j\rangle. \tag{35}$$

$\mathbf{\Lambda}$ is invertible by virtue of the linear independence of the basis tensors, and therefore

$$\beta_i = \sum_{j=1}^{N} \Lambda^{-1}_{ij}\langle\mathbb{C}_{\rm sym},\mathbb{V}_j\rangle. \tag{36}$$

Conversely, Eq. (34) describes all possible tensors with the given symmetry, and in particular, $\mathbb{C}_{\rm sym}=0$ if and only if $\beta_i=0$, $i=1,2,\ldots,N$.

C. The elastic projection

It helps to first derive the solution that minimizes the Euclidean distance function of (31). By expressing the unknown projected solution in the form (34), it follows that the minimum of $\|\mathbb{C}-\mathbb{C}_{\rm sym}\|^2$ is determined by setting to zero the derivatives with respect to β_i, which gives the system of simultaneous equations

$$\sum_{j=1}^{N} \langle \mathbb{V}_i, \mathbb{V}_j \rangle \beta_j = \langle \mathbb{V}_i, \mathbb{C} \rangle, \quad i = 1,2, \ldots, N. \tag{37}$$

The inner products are the elements of the invertible matrix $\mathbf{\Lambda}$, and so

$$\mathbb{C}_{\text{sym}} = \sum_{i,j=1}^{N} \langle \mathbb{C}, \mathbb{V}_i \rangle \Lambda_{ij}^{-1} \mathbb{V}_j. \tag{38}$$

Furthermore, the distance function at the optimal $\mathbb{C}_{\text{sym}}$ satisfies

$$\|\mathbb{C} - \mathbb{C}_{\text{sym}}\|^2 = \|\mathbb{C}\|^2 - \|\mathbb{C}_{\text{sym}}\|^2, \tag{39}$$

as expected for a projection using the Euclidean norm. This is essentially the method used by Arts *et al.*[18] Helbig[5] describes the procedure as the 'optimum approximation' of an arbitrary elasticity tensor by projection of a 21-dimensional vector onto a subspace of fewer dimensions, and Browaeys and Chevrot[3] list the explicit forms of the 21D projection operators for various symmetries.

Define the projection operator $\mathcal{P}_{\text{sym}}$ which maps $\mathbb{C}$ onto the chosen symmetry,

$$\mathcal{P}_{\text{sym}} \mathbb{C} \equiv \sum_{i,j=1}^{N} \langle \mathbb{C}, \mathbb{V}_i \rangle \Lambda_{ij}^{-1} \mathbb{V}_j. \tag{40}$$

Equations (38)–(40) imply that $\mathcal{P}_{\text{sym}}\mathbb{C}$ is the Euclidean projection, also equal to the closest elasticity tensor of the chosen symmetry to $\mathbb{C}$. We note the following important property:

$$\mathcal{P}_{\text{sym}} \mathbb{C}^* = (\mathcal{P}_{\text{sym}} \mathbb{C})^*. \tag{41}$$

In other words, the operation * commutes with the projection operator. This is not surprising if one considers that the * operation is a linear mapping on the symmetric and asymmetric parts of $\mathbb{C}$, and therefore * maintains the material symmetry of $\mathbb{C}$. However, a more detailed proof of the identity (41) is provided in the Appendix.

D. Solution of the generalized Fedorov problem

We now calculate the optimal $\mathbb{C}_{\text{sym}}$ for the generalized Fedorov problem, and show that it is equivalent to the moduli from the Euclidean norm. The starting point this time is to express the unknown $\mathbb{C}^*_{\text{sym}}$ (rather than $\mathbb{C}_{\text{sym}}$) in terms of the basis tensors,

$$\mathbb{C}^*_{\text{sym}} = \sum_{i=1}^{N} \alpha_i \mathbb{V}_i. \tag{42}$$

This is justified by the fact that all tensors of the given symmetry are linear combinations of the basis tensors. Furthermore, the coefficients α_i are related to those in Eq. (34) by the $N \times N$ matrix $\mathbf{P}$ introduced in the Appendix. Let $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote the $N \times 1$ arrays with elements α_i and α_i; then,

$$\boldsymbol{\beta} = \mathbf{P}^t \boldsymbol{\alpha} \Leftrightarrow \boldsymbol{\alpha} = 2\mathbf{P}^t \boldsymbol{\beta} - \boldsymbol{\beta}. \tag{43}$$

Equation (30) implies that the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_N$ satisfy the system of N equations,

$$\sum_{j=1}^{N} \langle \mathbb{V}_i, \mathbb{V}_j \rangle_a \alpha_j = \langle \mathbb{V}_i, \mathbb{C}^* \rangle_a, \quad i = 1,2, \ldots, N, \tag{44}$$

or, in matrix format,

$$\mathbf{S}\mathbf{\Lambda}\boldsymbol{\alpha} = \mathbf{S}\boldsymbol{\gamma}. \tag{45}$$

Here, $\boldsymbol{\gamma}$ with elements γ_i and the $N \times N$ matrix $\mathbf{S}$ are defined by

$$\gamma_i = \langle \mathbb{V}_i, \mathbb{C}^* \rangle, \tag{46a}$$

$$\sum_{j=1}^{N} S_{ij} \mathbb{V}_j = \tfrac{1}{2}(\mathbb{I}^{(s)} \mathbb{V}_i + \mathbb{V}_i \mathbb{I}^{(s)}) \equiv \mathbb{U}_i. \tag{46b}$$

The tensors $\mathbb{U}_i$, $i = 1,2, \ldots, N$ form a linearly independent set of basis tensors for the given symmetry. This may be seen by assuming the contrary, i.e., that there is a set of nonzero coefficients a_i such that

$$\sum_{i=1}^{N} a_i \mathbb{U}_i = 0. \tag{47}$$

Let $\mathbb{A}$ be the nonzero tensor,

$$\mathbb{A} = \sum_{i=1}^{N} a_i \mathbb{V}_i. \tag{48}$$

Then, Eq. (47) requires that

$$\tfrac{2}{3}\mathbb{A} + \tfrac{1}{2}(\mathbb{A}\mathbf{I}) \otimes \mathbf{I} + \tfrac{1}{2}\mathbf{I} \otimes (\mathbb{A}\mathbf{I}) = 0. \tag{49}$$

Multiplication by $\mathbf{I}$ implies

$$\tfrac{13}{6}\mathbb{A}\mathbf{I} + \tfrac{3}{2}\langle \mathbb{J}, \mathbb{A} \rangle \mathbf{I} = 0, \tag{50}$$

and taking the inner product with $\mathbf{I}$ gives

$$11 \langle \mathbb{J}, \mathbb{A} \rangle = 0. \tag{51}$$

Therefore,

$$\mathbb{A} = 0, \tag{52}$$

and the $\mathbb{U}_i$ tensors form a linearly independent basis for the symmetry. In particular, the tensors $\mathbb{V}_i$ can be expressed in terms of this alternate basis, and so the matrix $\mathbf{S}$ is invertible. Hence,

$$\boldsymbol{\alpha} = \mathbf{\Lambda}^{-1} \boldsymbol{\gamma}. \tag{53}$$

We are now in a position to determine the optimal $\mathbb{C}^*_{\text{sym}}$ and hence $\mathbb{C}_{\text{sym}}$. Equations (42) and (53), along with (40), imply

$$\mathbb{C}^*_{\text{sym}} = \sum_{i,j=1}^{N} \langle \mathbb{C}^*, \mathbb{V}_i \rangle \Lambda_{ij}^{-1} \mathbb{V}_j = \mathcal{P}_{\text{sym}} \mathbb{C}^*. \tag{54}$$

Using the fundamental property of the projection operator (41), the optimal elasticity as determined by (54) is seen to be exactly the same as the Euclidean projection, i.e., of Eq. (38). This completes the proof of the main result, the equivalence (33).

IV. EXAMPLES

The general procedure for projection is illustrated for several symmetries, and an example application is discussed in this section.

A. Basis functions for isotropic, cubic, and hexagonal materials

The tensor decomposition procedure is described for the three highest symmetries: isotropic, cubic, and transversely isotropic. The fundamental matrices $\mathbf{\Lambda}$ and $\mathbf{P}$ are given explicitly.

1. Isotropic approximation (*N*=2)

Let the basis tensors be $\mathbb{V}_1=\mathbb{J}$, $\mathbb{V}_2=\mathbb{K}\equiv\mathbb{I}-\mathbb{J}$. Then, $\mathbf{\Lambda}$ is a 2×2 diagonal matrix, $\mathbf{\Lambda}=\mathrm{diag}(1,5)$, and the optimal moduli are

$$\mathbb{C}_{\mathrm{iso}}=3\kappa\mathbb{J}+2\mu\mathbb{K}, \tag{55}$$

where

$$\kappa=\tfrac{1}{3}\langle\mathbb{J},\mathbb{C}\rangle,\quad \mu=\tfrac{1}{10}\langle\mathbb{K},\mathbb{C}\rangle. \tag{56}$$

This is precisely Fedorov's original result,[10] Eq. (1). The * operation is defined by the matrix $\mathbf{P}$ of Eq. (A1), which is

$$\mathbf{P}=\begin{pmatrix}\frac{1}{3} & \frac{1}{3}\\ \frac{5}{3} & \frac{1}{6}\end{pmatrix}. \tag{57}$$

2. Cubic materials (*N*=3)

Let $\mathbf{a},\mathbf{b},\mathbf{c}$ be the cube axes, and select as basis tensors[16] $\mathbb{V}_1=\mathbb{J}$, $\mathbb{V}_2=\mathbb{L}\equiv\mathbb{I}-\mathbb{H}$, and $\mathbb{V}_3=\mathbb{M}\equiv\mathbb{H}-\mathbb{J}$, where

$$\mathbb{H}=\mathbf{a}\otimes\mathbf{a}\otimes\mathbf{a}\otimes\mathbf{a}+\mathbf{b}\otimes\mathbf{b}\otimes\mathbf{b}\otimes\mathbf{b}+\mathbf{c}\otimes\mathbf{c}\otimes\mathbf{c}\otimes\mathbf{c}. \tag{58}$$

Then, $\mathbf{\Lambda}=\mathrm{diag}(1,3,2)$ and the optimal moduli of cubic symmetry are

$$\mathbb{C}_{\mathrm{cub}}=3\kappa'\mathbb{J}+2\mu'\mathbb{L}+2\eta\mathbb{M}, \tag{59}$$

where

$$\kappa'=\tfrac{1}{3}\langle\mathbb{J},\mathbb{C}\rangle,\quad \mu'=\tfrac{1}{6}\langle\mathbb{L},\mathbb{C}\rangle\quad \eta=\tfrac{1}{4}\langle\mathbb{M},\mathbb{C}\rangle. \tag{60}$$

It is assumed here that the axes of the cubic material are known; otherwise, a numerical search must be performed to find the axes which give the closest, i.e., largest, projection. This additional step is discussed in the numerical example below. Also,

$$\mathbf{P}=\begin{pmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ 1 & \frac{1}{2} & -\frac{1}{2}\\ \frac{2}{3} & -\frac{1}{3} & \frac{2}{3}\end{pmatrix}. \tag{61}$$

3. Transverse isotropy/hexagonal symmetry (*N*=5)

Let $\mathbf{a}$ be the direction of the symmetry axis, and define[16]

$$\mathbb{V}_1=\mathbf{A}\otimes\mathbf{A}, \tag{62}$$

$$\mathbb{V}_2=\tfrac{1}{2}\mathbf{B}\otimes\mathbf{B},\quad \mathbb{V}_3=\tfrac{1}{2}(\mathbf{A}\otimes\mathbf{B}+\mathbf{B}\otimes\mathbf{A}),\quad \mathbb{V}_4=2\mathbb{V}_3^*,\quad \mathbb{V}_5=2\mathbb{V}_2^*-\mathbb{V}_2,$$

where $\mathbf{A}=\mathbf{a}\otimes\mathbf{a}$ and $\mathbf{B}=\mathbf{I}-\mathbf{A}$. Then, $\mathbf{\Lambda}=\mathrm{diag}(1,1,1,2,2)$ and the optimal transverse isotropy (TI) moduli are given by

$$\mathbb{C}_{\mathrm{TI}}=\sum_{i=1}^{3}\langle\mathbb{C},\mathbb{V}_i\rangle\mathbb{V}_i+\tfrac{1}{2}\sum_{i=4}^{5}\langle\mathbb{C},\mathbb{V}_i\rangle\mathbb{V}_i. \tag{63}$$

Also,

$$\mathbf{P}=\begin{pmatrix}1 & 0 & 0 & 0 & 0\\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2}\\ 0 & 0 & 0 & \frac{1}{2} & 0\\ 0 & 0 & 1 & \frac{1}{2} & 0\\ 0 & 1 & 0 & 0 & 0\end{pmatrix}. \tag{64}$$

B. Application to acoustically measured data

Let us assume that the material has cubic symmetry but the cube axes orientations are unknown. The effective cubic moduli are, in coordinates coincident with the cube axes,

$$\mathbf{C}_{\mathrm{cub}}=\begin{pmatrix}c_{11}^c & c_{12}^c & c_{12}^c & 0 & 0 & 0\\ c_{12}^c & c_{11}^c & c_{12}^c & 0 & 0 & 0\\ c_{12}^c & c_{12}^c & c_{11}^c & 0 & 0 & 0\\ 0 & 0 & 0 & c_{66}^c & 0 & 0\\ 0 & 0 & 0 & 0 & c_{66}^c & 0\\ 0 & 0 & 0 & 0 & 0 & c_{66}^c\end{pmatrix}, \tag{65}$$

where the three constants can be expressed in terms of the bulk modulus and the two shear moduli,

$$c_{11}^c=\kappa'+\tfrac{4}{3}\eta,\quad c_{12}^c=\kappa'-\tfrac{2}{3}\eta,\quad c_{66}^c=\mu'. \tag{66}$$

The effective bulk modulus of (60) is an isotropic invariant that is independent of the cube axes orientation,

$$\kappa'=\tfrac{1}{9}C_{iijj}=\tfrac{1}{9}(c_{11}+c_{22}+c_{33}+2c_{12}+2c_{23}+2c_{31}). \tag{67}$$

Similarly, the combination $(3\mu'+2\eta)/5=\mu$, the isotropic shear modulus implying

$$6\mu'+4\eta=C_{ijij}-\tfrac{1}{3}C_{iijj}=\tfrac{2}{3}(c_{11}+c_{22}+c_{33}-c_{12}-c_{23}-c_{31})+2(c_{44}+c_{55}+c_{66}). \tag{68}$$

Only one of the three cubic parameters depends upon the orientation of the axes, and it follows by considering the squared length of the projected elastic tensor,

$$\|\mathbb{C}_{\rm cub}\|^2 = 9\kappa'^2 + 12\mu'^2 + 8\eta^2 = 9\kappa'^2 + \tfrac{1}{5}(6\mu' + 4\eta)^2 + \tfrac{6}{5}(2\mu' - 2\eta)^2. \quad (69)$$

The first two terms are independent of the cube axes orientation, and therefore the closest cubic projection maximizes the final term. Note that

$$2\mu' - 2\eta = \tfrac{1}{3}C_{ijij} + \tfrac{1}{6}C_{iijj} - \tfrac{5}{6}\langle\mathbb{C},\mathbb{H}\rangle = \tfrac{5}{2}\kappa' + 2\mu' + \tfrac{4}{3}\eta - \tfrac{5}{6}\langle\mathbb{C},\mathbb{H}\rangle. \quad (70)$$

The moduli κ', μ', and η must all be positive in order for the material to have positive-definite strain energy. Therefore, $(2\mu' - 2\eta)^2$ is maximum when $\langle\mathbb{C},\mathbb{H}\rangle$ is minimum. The latter is also a positive quantity, which may be expressed as

$$\langle\mathbb{C},\mathbb{H}\rangle = C_{ijkl}H_{ijkl} = c'_{11} + c'_{22} + c'_{33}, \quad (71)$$

where c'_{IJ} are the elements of $\mathbb{C}$ of (74) in the coordinate system coincident with the cube axes.

We can also write

$$\langle\mathbb{C},\mathbb{H}\rangle = \tilde{\mathbf{a}}^t\mathbf{C}\tilde{\mathbf{a}} + \tilde{\mathbf{b}}^t\mathbf{C}\tilde{\mathbf{b}} + \tilde{\mathbf{c}}^t\mathbf{C}\tilde{\mathbf{c}}, \quad (72)$$

where

$$\tilde{\mathbf{a}}^t = (a_1^2, a_2^2, a_3^2, 2a_2a_3, 2a_3a_1, 2a_1a_2), \text{ etc.} \quad (73)$$

The minimum can be found by numerically searching over all possible orientations using Euler angles to transform from the fixed vector basis to the cube axes. The rotated (cube) axes are simply the columns of the 3×3 transformation matrix $\mathbf{R} \in SO(3)$.

As a specific application we reconsider the elastic moduli obtained from ultrasonic measurements by Francois *et al.*[2] In the notation of Eq. (22), the raw data for the stiffness tensor are

$$\mathbf{C} = \begin{pmatrix} 243 & 136 & 135 & 22 & 52 & -17 \\ 136 & 239 & 137 & -28 & 11 & 16 \\ 135 & 137 & 233 & 29 & -49 & 3 \\ 22 & -28 & 29 & 133 & -10 & -4 \\ 52 & 11 & -49 & -10 & 119 & -2 \\ -17 & 16 & 3 & -4 & -2 & 130 \end{pmatrix} \text{(GPa)}. \quad (74)$$

The element c_{53} is given as 49 in Ref. 2, which appears to be a typographical error.

We find that

$$\mathbf{C}_{\rm cub} = \begin{pmatrix} 213.4 & 148.5 & 148.5 & 0 & 0 & 0 \\ 148.5 & 213.4 & 148.5 & 0 & 0 & 0 \\ 148.5 & 148.5 & 213.4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 139.8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 139.8 & 0 \\ 0 & 0 & 0 & 0 & 0 & 139.8 \end{pmatrix} \text{(GPa)}, \quad (75)$$

in agreement with Francois *et al.*[2] The rotation from the coordinate system $\{\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3\}$ of (74) can be expressed as a rotation about a single axis $\mathbf{n}$ through angle θ using Euler's theorem.[19] The angle and axis follow from $2\cos\theta = tr\mathbf{R} - 1$ and $2\sin\theta\mathbf{n} = \mathbf{skew}\,\mathbf{R}$, where $\mathbf{skew}\,\mathbf{Y} = -\epsilon_{ijk}Y_{ij}\mathbf{e}_k$, and ϵ_{ijk} is the third-order alternating tensor. Thus, we find $\theta = 96°$ and $\mathbf{n} = (0.18, 0.06, 0.98)$, and the full set of moduli in the rotated frame is

$$\mathbf{C}' = \begin{pmatrix} 228 & 141 & 148 & -5 & -1 & -2 \\ 141 & 209 & 156 & -6 & 23 & -1 \\ 148 & 156 & 203 & -7 & -2 & 4 \\ -5 & -6 & -7 & 144 & 12 & -3 \\ -1 & 23 & -2 & 12 & 139 & 11 \\ -2 & -1 & 4 & -3 & 11 & 136 \end{pmatrix} \text{(GPa)}. \quad (76)$$

This set of cube axes is unique within rotation under the group of transformations congruent with cubic symmetry. In this preferred coordinate system, the projection onto the cubic moduli in this frame is simply

$$c^c_{11} = \tfrac{1}{3}(c'_{11} + c'_{22} + c'_{33}) = 213.4 \text{ (GPa)},$$

$$c^c_{12} = \tfrac{1}{3}(c'_{12} + c'_{23} + c'_{31}) = 148.5,$$

$$c^c_{66} = \tfrac{1}{3}(c'_{44} + c'_{55} + c'_{66}) = 139.8. \quad (77)$$

V. CONCLUSION

The main result of this paper is the proof that the Euclidean projection of anisotropic elastic constants onto a higher material is identical to minimizing the mean-square difference of the slowness surfaces. This provides a well-grounded acoustical basis for using the Euclidean projection as a natural way to simplify ultrasonic or acoustic data. The equivalence generalizes the original result of Fedorov for the best isotropic acoustical fit to a given set of anisotropic moduli.

ACKNOWLEDGMENT

Discussions with Dr. M. Moakher have benefited this paper and are appreciated.

APPENDIX: THE MATRIX P AND PROOF OF EQ. (41)

The tensor $\mathbb{V}_i^*$ can be expressed using Eq. (19) in terms of the totally symmetric and asymmetric parts of $\mathbb{V}_i$. Both $\mathbb{V}_i^{(s)}$ and $\mathbb{V}_i^{(a)}$ possess the same material symmetry as $\mathbb{V}_i$, and hence the symmetry is inherited by $\mathbb{V}_i^*$. Since the $\mathbb{V}_i$'s themselves form a basis for the material symmetry, it follows that $\mathbb{V}_i^*$ can be written as a linear combination of them. Let $\mathbf{P}$ be the $N\times N$ matrix which defines the * operation in terms of the basis tensors, that is,

$$\mathbb{V}_i^* = \sum_{j=1}^{N} P_{ij}\mathbb{V}_j. \tag{A1}$$

It follows from Eq. (20) that

$$\mathbb{V}_i = -\mathbb{V}_i^* + 2\sum_{j=1}^{N} P_{ij}\mathbb{V}_j^*, \tag{A2}$$

and hence the inverse of $\mathbf{P}$ is

$$\mathbf{P}^{-1} = 2\mathbf{P} - \mathbf{I}_{(N)}, \tag{A3}$$

where $\mathbf{I}_{(N)}$ is the $N\times N$ identity.

Consider the identity

$$\langle \mathbb{V}_i^*, \mathbb{V}_j\rangle = \langle \mathbb{V}_i, \mathbb{V}_j^*\rangle, \tag{A4}$$

which follows from (21). Using (A1) to eliminate $\mathbb{V}_i^*$ and $\mathbb{V}_j^*$ from the left and right members, Eq. (A4) implies

$$\mathbf{P}\mathbf{\Lambda} = \mathbf{\Lambda}\mathbf{P}^t. \tag{A5}$$

It is worth noting the very specific nature of the matrix $\mathbf{P}$ that is required to satisfy Eq. (A5). This matrix essentially defines the * operator, from which the totally symmetric and asymmetric parts of a tensor can be found in terms of the basis tensors. Some examples of $\mathbf{P}$ are given in Sec. IV.

We now turn to the proof of Eq. (41). Starting with the definition of $\mathcal{P}_{\text{sym}}$ in (40), we have

$$\begin{aligned}\mathcal{P}_{\text{sym}}\mathbb{C}^* &= \sum_{i,j=1}^{N} \langle \mathbb{C}^*, \mathbb{V}_i\rangle \Lambda_{ij}^{-1}\mathbb{V}_j = \sum_{i,j=1}^{N} \langle \mathbb{C}, \mathbb{V}_i^*\rangle \Lambda_{ij}^{-1}\mathbb{V}_j \\ &= \sum_{i,j=1}^{N} \langle \mathbb{C}, \mathbb{V}_i\rangle X_{ij}\mathbb{V}_j^*,\end{aligned} \tag{A6}$$

where (21) has been used in the second line, and the matrix $\mathbf{X}$ is defined as

$$\mathbf{X} = \mathbf{P}^t\mathbf{\Lambda}^{-1}\mathbf{P}^{-1}. \tag{A7}$$

Comparison with (A5) gives

$$\mathbf{X} = \mathbf{\Lambda}^{-1}, \tag{A8}$$

and substituting from (A8) into (A6) and comparing it with (38) implies the fundamental property (41).

[1] M. J. P. Musgrave, *Crystal Acoustics* (Acoustical Society of America, New York, 2003).
[2] M. Francois, G. Geymonat, and Y. Berthaud, "Determination of the symmetries of an experimentally determined stiffness tensor, application to acoustic measurements," Int. J. Solids Struct. **35**(31–32), 4091–4106 (1998).
[3] J. T. Browaeys and S. Chevrot, "Decomposition of the elastic tensor and geophysical applications," Geophys. J. Int. **159**, 667–678 (2004).
[4] D. C. Gazis, I. Tadjbakhsh, and R. A. Toupin, "The elastic tensor of given symmetry nearest to an anisotropic elastic tensor," Acta Crystallogr. **16**, 917–922 (1963).
[5] K. Helbig, "Representation and approximation of elastic tensors," in Sixth International Workshop on Seismic Anisotropy (Trondheim, 1994), Expanded Abstracts, pp. 37–75 (1995).
[6] F. Cavallini, "The best isotropic approximation of an anisotropic Hooke's law," Bollettino di Geofisica Teorica e Applicata **40**, 1–18 (1999).
[7] A. F. Gangi, "Fourth-order elastic-moduli tensors by inspection," in *Anisotropy 2000: Fractures, converted waves and case studies. Proc. 9th International Workshop on Seismic Anisotropy (9IWSA)*, edited by L. Ikelle and A. F. Gangi, Tulsa, OK, 2000. (Society of Exploration Geophysicists, Tulsa, OK, 2000).
[8] M. Moakher, "On the averaging of symmetric positive-definite tensors," J. Elast. (in press).
[9] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache, "Fast and simple calculus on tensors in the log-Euclidean framework," in J. Duncan, and G. Gerig, editors, *Proc. 8th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention—MICCAI 2005, Part I*, Vol. **3749** of *LNCS*, Palm Springs, CA, 2005 (Springer, Berlin, 2005).
[10] F. I. Fedorov, *Theory of Elastic Waves in Crystals* (Plenum, New York, 1968).
[11] G. Backus, "A geometrical picture of anisotropic elastic tensors," Rev. Geophys. Space Phys. **8**, 633–671 (1970).
[12] A. J. M. Spencer, "A note on the decomposition of tensors into traceless symmetric tensors," Int. J. Eng. Sci. **8**(6), 475–481 (1970).
[13] R. Baerheim, "Harmonic decomposition of the anisotropic elasticity tensor," Q. J. Mech. Appl. Math. **46**, 391–418 (1993).
[14] S. Forte and M. Vianello, "Symmetry classes for elasticity tensors," J. Elast. **43**, 81–108 (1996).
[15] A. N. Norris, "On the acoustic determination of the elastic moduli of anisotropic solidsand acoustic conditions for the existence of planes of symmetry," Q. J. Mech. Appl. Math. **42**, 413–426 (1989).
[16] L. J. Walpole, "Fourth rank tensors of the thirty-two crystal classes: Multiplication tables," Proc. R. Soc. London, Ser. A **391**, 149–179 (1984).
[17] I. A. Kunin, *Elastic Media with Microstructure* (Springer, Berlin, 1982).
[18] R. J. Arts, K. Helbig, and P. N. J. Rasolofosaon, "General anisotropic elastic tensors in rocks: approximation, invariants and particular directions," in *Expanded Abstracts, 61st Annual International Meeting, Society of Exploration Geophysicists, ST2.4*, pages 1534—1537, Tulsa, OK, 1991 (Society of Exploration Geophysicists, Tulsa, OK, 1991).
[19] H. Baruh, *Analytical Dynamics* (McGraw-Hill, New York, 1999).

Model-based analysis of dispersion curves using chirplets

Helge Kuttig[a)]
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0355

Marc Niethammer
Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02215

Stefan Hurlebaus
Zachry Department of Civil Engineering, Texas A&M University, College Station, Texas 77843-3136

Laurence J. Jacobs[b)]
School of Civil and Environmental Engineering and G. W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 18 November 2005; revised 23 January 2006; accepted 24 January 2006)

Time-frequency representations, like the spectrogram or the scalogram, are widely used to characterize dispersive waves. The resulting energy distributions, however, suffer from the uncertainty principle, which complicates the allocation of energy to individual propagation modes (especially when the dispersion curves of these modes are close to each other in the time-frequency domain). This research applies the chirplet as a tool to analyze dispersive wave signals based on a dispersion model. The chirplet transform, a generalization of both the wavelet and the short-time Fourier transform, enables the extraction of components of a signal with a particular instantaneous frequency and group delay. An adaptive algorithm identifies frequency regions for which quantitative statements can be made about an individual mode's energy, and employs chirplets (locally adapted to a dispersion curve model) to extract the (proportional) energy distribution of that single mode from a multimode dispersive wave signal. The effectiveness of this algorithm is demonstrated on a multimode synthetic Lamb wave signal for which the ground-truth energy distribution is known for each mode. Finally, the robustness of this algorithm is demonstrated on real, experimentally measured Lamb wave signals by an adaption of a correlation technique developed in previous research. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2177587]

PACS number(s): 43.35.Cg, 43.60.Hj [YHB] Pages: 2122–2130

I. INTRODUCTION

Ultrasonic waves are used for material characterization in many structural health monitoring and nondestructive evaluation applications. Guided ultrasonic waves (such as Lamb waves) are particularly effective in interrogating large structural components, because guided waves propagate far distances (when compared to body waves); see Chimenti[1] for details. Unfortunately, the dispersive nature of these guided waves complicates their analysis in the time-domain, so time-frequency representations (TFRs) are often employed to better interpret a guided wave signal. Time-frequency analysis represents the frequency content of a time-domain signal over time, and is thus ideally suited for the analysis of dispersive waves. Niethammer *et al.*[2] demonstrated the effectiveness of using TFRs for Lamb waves, while Hurlebaus *et al.*[3,4] describe several applications for crack detection and localization in a plate using TFRs such as the spectrogram and the scalogram.

In 1991, Mann *et al.*[5] coined the term "chirplet transformation" for a new TFR, a generalization of the spectrogram and scalogram with additional degrees of freedom for time and frequency shear. While the chirplet transform allows for the decomposition of time-domain signals into well-localized components (time-frequency atoms) with a constant chirp rate (that is, linearly changing instantaneous frequency), the resulting representation is of dimension five. This complicates the visualization, computation, and interpretation of a transformed signal. In order to apply the chirplet transform to signals containing chirplike components, such as, Doppler radar signals, electroencephalographies (EEGs), bird voices or bat sonar signals, researchers reduce the problem's complexity by considering only subspaces[6,7] of the five-dimensional parameter space or by employing a matching pursuit algorithm to search for ridges in a coarsely discretized parameter space.[8]

Hong *et al.*[9] utilize a dispersion-based short-time Fourier transform (a chirplet transform) for the analysis of dispersion curves from guided wave signals. In contrast, the objective of the current research is to establish a model-based chirplet analysis of dispersive waves grounded in the theory developed by Mann,[5] and then to investigate the quantitative behavior of the proposed method for the analysis of dispersion

[a)]Current address: Institute of Applied and Experimental Mechanics, University of Stuttgart, Pfaffenwaldring 9, 70569 Stuttgart, Germany.
[b)]Electronic mail: laurence.jacobs@ce.gatech.edu

 0001-4966/2006/119(4)/2122/9/$22.50

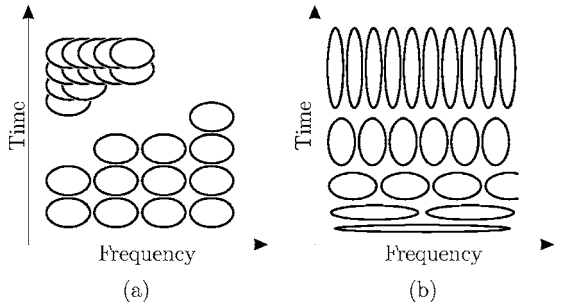


FIG. 1. Tiling of the time-frequency plane by time-frequency representations. Each ellipse represents one time-frequency atom. (a) Spectrogram. (b) Scalogram.

curves. The dispersion curve model used in this study is the Rayleigh-Lamb equations, which describes wave propagation in platelike structures. The proposed algorithm is first validated on a synthetic signal for which the ground-truth energy distribution is known for each mode. Finally, the robustness of this algorithm is demonstrated on real, experimentally measured Lamb wave signals by an adaption of the correlation technique developed in previous research.[4,10,11]

II. REVIEW OF TFRS FOR GUIDED WAVES

The dispersive nature of Lamb waves (guided ultrasonic waves in plates) means that the frequency content of a measured Lamb wave signal varies as a function of time. A simple time- or frequency-domain representation of such a dispersive signal does not (individually) localize the energy content of a Lamb wave in the time-frequency domain. So, more advanced tools such as TFRs, are needed for the comprehensive analysis of these dispersive signals. One such TFR, the short-time Fourier transform (STFT), C_x^{stft}, is calculated for a signal $x(t)$ as follows:[12]

$$C_x^{\mathrm{stft}}(t,\omega) = \int_{-\infty}^{+\infty} x(\tau) h(\tau - t) e^{-i\omega\tau} d\tau, \tag{1}$$

where $h(t)$ is a window function. The square modulus of the STFT is called the spectrogram $P^{\mathrm{sp}}(t,\omega) = |C_x^{\mathrm{stft}}(t,\omega)|^2$ and yields the energy density of the signal x in the time-frequency domain.

All TFRs suffer from the uncertainty principle,[13] so the resulting time-frequency values do not necessarily correspond to the exact energy distribution of a signal in the time-frequency domain; TFRs based on smoothing kernels (like the spectrogram and scalogram) result in smoothed versions of the exact energy distribution.[14] The extent of the smoothing depends on the smoothing kernel (the window function) employed. For a spectrogram using a Gaussian window function

$$h(t) = \frac{1}{\sqrt[4]{\pi s^2}} \exp\left[-\frac{1}{2}\left(\frac{t - t_0}{s} \right)^2 \right],$$

this smoothing (averaging) effect is minimized, and the regions of averaging can be approximately visualized by ellipses of the same size and shape as shown in Fig. 1(a).

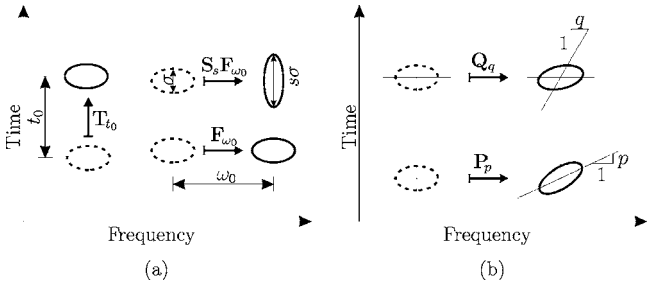


FIG. 2. Representation of time-frequency atoms of a TFR. (a) The generalization of the STFT and wavelet transforms by introducing the operators $\mathbf{T}_{t_0}$, $\mathbf{F}_{\omega_0}$, and $\mathbf{S}_s$ which allows arbitrary time and frequency shifts, and scale. (b) Modulation with a chirp in the time domain ($\mathbf{Q}_q$) shears the time-frequency atom in the frequency direction, modulation with a chirp in the frequency-domain ($\mathbf{P}_p$) shears in the time direction. The shear parameters p and q determine the slope of the semiaxes of the chirplet atom ellipses.

Other TFRs result in different tilings of the time-frequency domain; consider, for example, the wavelet transform, C_x^{wt} of a signal $x(t)$,

$$C_x^{\mathrm{wt}}(t,s) = \int_{-\infty}^{+\infty} x(\tau) \frac{1}{\sqrt{s}} \psi^*\left(\frac{\tau - t}{s} \right) d\tau, \tag{2}$$

where s is a dilatation and t is a shift parameter, which scale and shift the mother wavelet $\psi(t)$ and $(\cdot)^*$ denotes the complex conjugate. The energy density $P^{\mathrm{wt}}(t,s) = |C_x^{\mathrm{wt}}(t,s)|^2$ is called the scalogram and again localizes the energy of a signal in the time-frequency domain under the restrictions of the uncertainty principle, yielding to a tiling of the time-frequency domain as shown in Fig. 1(b).

III. THE CHIRPLET TRANSFORM

Both the STFT and the wavelet transform result from the inner multiplication of the time function under consideration with a time-frequency atom, $g(t)$

$$C_x = \int_{-\infty}^{+\infty} x(\tau) g^*(\tau) d\tau. \tag{3}$$

This time-frequency atom is obtained from the window function by either shifting it in the time and frequency directions (STFT), or with a time shift and scale (wavelet transformation); a comparison is shown in Fig. 2(a).

By introducing the time and frequency shift operators $\mathbf{T}_{t_0}$ and $\mathbf{F}_{\omega_0}$ and the scale operator $\mathbf{S}_s$ as defined in Table I, a generalized three-dimensional TFR is written as

$$C_x = \int_{-\infty}^{+\infty} x(\tau) g^*_{t_0,\omega_0,s}(\tau) d\tau \tag{4}$$

with

$$g_{t_0,\omega_0,s}(t) = \mathbf{T}_{t_0} \mathbf{F}_{\omega_0} \mathbf{S}_s h(t). \tag{5}$$

Mann[7] uses this framework and expands it by two additional operators to formulate the chirplet transform.

(1) Frequency shear: The time-frequency atom is multiplied by a harmonic function of linearly changing frequency (a "chirp" function)

TABLE I. Operations for time-frequency atoms for the chirplet transform, k_p is a complex scaling factor.

Operation		Corresponding coordinate transformation $(t,\omega)\rightarrow(\bar{t},\bar{\omega})$	
Time shift $\mathbf{T}_{t_0}$	$\mathbf{T}_{t_0}h(t)=h(t-t_0)$	$\bar{t}=t-t_0$	$\bar{\omega}=\omega$
Frequency shift $\mathbf{F}_{\omega_0}$	$\mathbf{F}_{\omega_0}h(t)=e^{i\omega_0 t}h(t)$	$\bar{t}=t$	$\bar{\omega}=\omega-\omega_0$
Scaling $\mathbf{S}_s$	$\mathbf{S}_s h(t)=\frac{1}{\sqrt{s}}h\left(\frac{t}{s}\right)$	$\bar{t}=\frac{t}{s}$	$\bar{\omega}=s\omega$
Frequency shear $\mathbf{Q}_q$	$\mathbf{Q}_q h(t)=h(t)\exp[iq/2t^2]$	$\bar{t}=t$	$\bar{\omega}=\omega-qt$
Time shear $\mathbf{P}_p$	$\mathbf{P}_p h(t)=k_p\exp\left[-i\frac{1}{2q}t^2\right]\times h(t)$	$\bar{t}=t-p\omega$	$\bar{\omega}=\omega$

$$g_q(t)=\mathbf{Q}_q h(t)=h(t)\exp\left[i\frac{q}{2}t^2\right]. \tag{6}$$

This modulation with a time-dependent frequency results in a shear $\mathbf{Q}_q$ of the time-frequency atom in the frequency direction as shown in Fig. 2(b). The time-frequency atom is now concentrated along a line with slope $1/q$.

(2) Time shear: The corresponding multiplication of a chirp function with the time-frequency atom in the frequency-domain (according to a convolution in the time-domain) constitutes the dual operation to the frequency shear, the shear $\mathbf{P}_p$ in the time direction

$$g_p(t)=\mathbf{P}_p h(t)=\sqrt{\frac{2\pi}{p}}\exp\left[i\left(-\frac{1}{2q}t^2+\frac{\pi}{4}\right)\right]\times h(t). \tag{7}$$

The effect of this operation is shown in Fig. 2(b); one semiaxis of the time-frequency atom has the slope p. The complex scaling factor $\sqrt{(2\pi/p)}\exp[i\pi/4]$ results from the transformation from the frequency domain.

Using these five operations, time-frequency atoms can be translated, scaled and sheared in the time-frequency domain. Each operation specifies a linear or affine coordinate transformation in the time-frequency domain. The mapping equations for the transformations are listed in Table I. A time-frequency atom transformed like this is called a "chirplet atom" or, in short, "chirplet" $g_{t_0,\omega_0,s,q,p}(t)$

$$g_{t_0,\omega_0,s,q,p}(t)=\mathbf{T}_{t_0}\mathbf{F}_{\omega_0}\mathbf{S}_s\mathbf{Q}_q\mathbf{P}_p h(t). \tag{8}$$

The chirplet transform C_x^{ct} as defined by Mann *et al.*[5] is obtained by the inner multiplication of a chirplet atom with a time-domain signal

$$C_x^{\mathrm{ct}}(t_0,\omega_0,s,q,p)=\int_{-\infty}^{+\infty}x(\tau)g^*_{t_0,\omega_0,s,q,p}(\tau)d\tau. \tag{9}$$

The operators just described are not commutative, and interchanging two of them generally results in a phase shift of the chirplet. This will not be discussed further, as only the magnitudes of chirplet transforms are processed in this research.

Signal components which are concentrated in the time-frequency domain at a location specified by the parameters t_0, ω_0, s, q, and p can be extracted by performing a chirplet transform. Analogous to the spectrogram and scalogram, one defines the energy distribution

$$P^{\mathrm{ct}}(t_0,\omega_0,s,q,p)=|C_x^{\mathrm{ct}}(t_0,\omega_0,s,q,p)|^2. \tag{10}$$

For a given basic window, this results in a five-dimensional representation which is difficult to visualize. So instead of calculating the entire five-dimensional space, one often confines oneself to a plane spanned by two parameters in order to visualize the time-frequency content (like (t_0,ω_0) for the spectrogram).

Note that another approach is to decompose a signal into the chirplet's base, e.g., an algorithm to find local maxima in the five-dimensional space.[8] Finally, an alternative formulation of chirplets that is based on the Wigner-Ville distribution is presented by Baraniuk *et al.*[6]

IV. LAMB WAVE ENERGY ASSIGNMENT TO MODES

Lamb waves are multimode signals, so it is of great practical interest to develop a quantitative description of the energy distribution of the signal amongst the individual Lamb modes. The dispersion relationship (the relationship between wavenumber and frequency, which can be converted to the time-frequency domain) of guided Lamb waves propagating in a homogenous, isotropic plate of thickness $2h$ is described by the Rayleigh-Lamb equations[15]

$$\frac{\tan(\tilde{q}h)}{\tan(\tilde{p}h)}=-\frac{4k^2\tilde{p}\tilde{q}}{(\tilde{q}^2-k^2)^2}, \tag{11a}$$

$$\frac{\tan(\tilde{q}h)}{\tan(\tilde{p}h)}=-\frac{(\tilde{q}^2-k^2)^2}{4k^2\tilde{p}\tilde{q}}, \tag{11b}$$

where k is the wavenumber ($k=2\pi/\lambda$, where λ is wavelength), ω is angular frequency ($\omega=2\pi f$, where f is frequency in Hz), and

$$\tilde{p}^2=\frac{\omega^2}{c_{\mathrm{L}}^2}-k^2,\quad \tilde{q}^2=\frac{\omega^2}{c_{\mathrm{T}}^2}-k^2. \tag{12}$$

c_L and c_T are the longitudinal and transverse wave speeds, respectively.

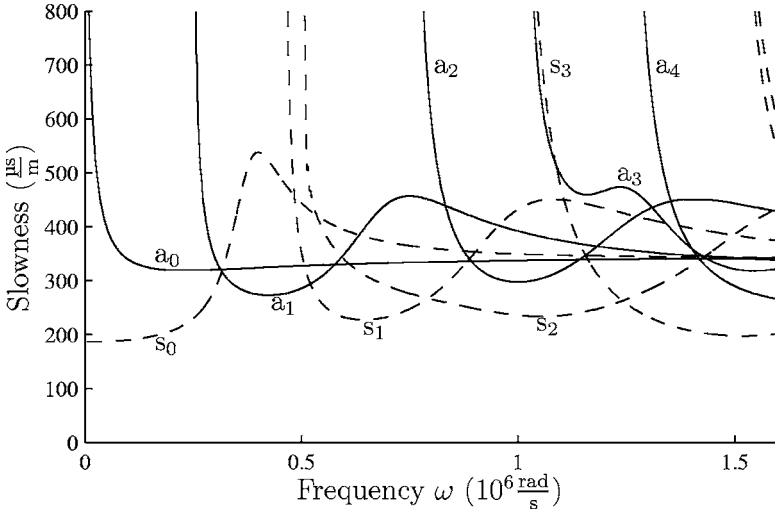


FIG. 3. Theoretical solution of Raleigh-Lamb equation for an aluminum 3003 plate of thickness 0.99 mm in the slowness-frequency domain. s_i denotes symmetric, a_i denotes antisymmetric Lamb modes.

Solutions of these equations, the dispersion curves, are plotted in Fig. 3 for an aluminum plate of thickness 0.99 mm. Each propagation mode in Fig. 3 corresponds to a single line, with modes numbered in order of the lowest frequency as they appear, and s_i and a_i denotes modes with symmetric and antisymmetric displacement profiles (with respect to the axis of propagation), respectively. These results are shown in the slowness-frequency domain, where slowness (sl_e) is defined as the reciprocal value of the group velocity. This representation is a normalization to the propagation (source-to-receiver) distance (d), where $sl_e = t/d$.

In typical structural health and nondestructive evaluation applications of guided ultrasonic waves, the displacement or velocity of a particle on the surface of a structural component is measured. In general, this displacement (or particle velocity) contains energy contributions from all the propagating modes (a superposition of all these modes) and it is difficult (if not impossible) to separate the contributions from each individual mode in a measured signal. The ability to accurately separate these individual modal energy contributions from a measured, multimode Lamb wave would significantly enhance current structural health and nondestructive evaluation applications. This research uses the chirplet to accurately separate these individual modal energy contributions from a multimode Lamb wave signal by using the Rayleigh-Lamb dispersion model given by Eq. (11)—the dispersion relationship shown in Fig. 3 is assumed to be known.

Consider a synthetically generated Lamb wave signal (developed following Refs. 16 and 17) where the exact energy distribution for each individual mode is known. This synthetic time-domain signal is an "ideal measurement" of a guided Lamb wave in an aluminum plate of thickness 0.99 mm. This synthetic signal represents ground-truth; the excitability of each mode is known, so it is possible to absolutely calculate the energy content of each of the propagation modes.

Note that the energy of a Lamb wave mode is proportional to the square of its displacement, but the constant of proportionality is mode and frequency dependent. Therefore, it is not prudent to develop an absolute ratio of the energies of the individual Lamb modes. Instead, consider the ratio of the surface displacements of two Lamb wave modes; the square root of the chirplet energy distribution defined by Eq. (10) is proportional to displacement (as is the square root of the spectrogram). This paper concentrates on the s_0 and a_0-modes—these two lowest modes carry most of the energy for the example under consideration—and will present the ratio of the out-of-plane displacements of mode s_0 over a_0, which will be referred to as the "mode displacement ratio." While not equal to a mode energy ratio, this representation will allow for the source-independent presentation of the relative "energy" distributions of each individual mode, which can be compared to the corresponding theoretical (ground-truth) values.

A. Calculating the mode displacement ratio using the spectrogram

First, consider the results when using the spectrogram to calculate the mode displacement ratio of this synthetic (time-domain) Lamb wave signal. A synthetic Lamb wave is generated for a propagation distance of 90 mm and a sampling frequency of 100 MHz (shown in the time-domain in Fig. 4) and is then operated on with a 1024-point Hanning window to calculate the spectrogram shown in Fig. 5 [plus the theoretical dispersion curves calculated using Eq. (11)]. Note that the time-domain signal in Fig. 4 contains the ten lowest modes, represents out-of-plane displacement, and has been highpass filtered at 0.3 MHz. Most of the energy of the spectrogram is localized near the theoretical dispersion curves, and it is a reasonable next step to calculate the (proportional) energy of the individual modes by simply using the magnitude values of the spectrogram—the amplitudes in the contour plot of Fig. 5 at each slowness-frequency position associated with the theoretical dispersion curves. The results of this operation (mode displacement ratio of s_0 over a_0) are plotted as a dotted line in Fig. 6, along with the known theoretical curve (as a solid gray line). The mode displacement ratio values calculated with the spectrogram are only close to the theoretical values in certain frequency regions; this poor agreement is primarily due to the uncertainty principle, since the spectrogram calculates the energy at a particular slowness-frequency (normalized time-frequency) location by taking an average over a certain region in the

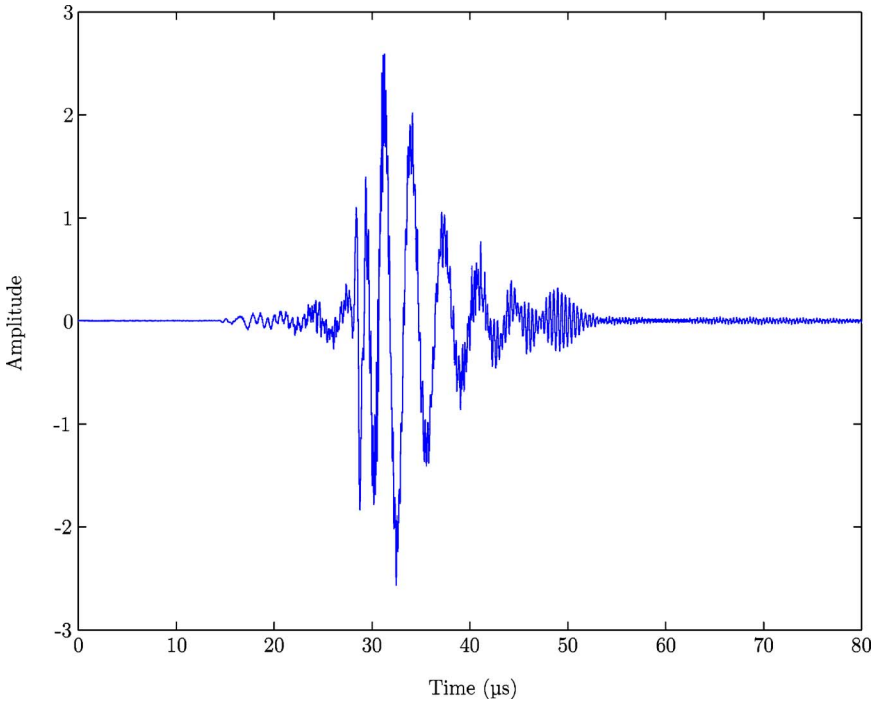


FIG. 4. (Color online) Synthetic time-domain signal, out-of-plane displacement, propagation distance $d=90$ mm.

slowness-frequency domain. This behavior is illustrated in Fig. 7(a), and note that when evaluating the spectrogram for one dispersion curve, contributions from its neighboring or intersecting dispersion curves are often inadvertently included.

B. Calculating the mode displacement ratio using chirplets

It is possible to improve the spectrogram results presented in Fig. 6 by instead using the additional degrees of freedom of the chirplet transform, combined with the dispersion model of Eq. (11). By using this dispersion relationship model, it is possible to adapt the shape of a chirplet to fit the known group delay of each propagation mode. This group delay describes which frequencies are present in a Lamb wave signal at a particular time, and is proportional to the slowness representation of a Lamb wave's dispersion curves. Using this procedure, the (proportional) energy of a particular propagation mode is extracted from a multimode signal with chirplets which are parametrized to be concentrated on the mode's dispersion curve in the slowness-frequency domain. This is accomplished by shifting chirplets onto the dispersion curve, and then shearing them to locally match the slope of the dispersion curve. The chirplet parameters t_0 and ω_0 are determined by the coordinates of the dispersion curve, and the parameter p equals the slope of the dispersion curve at the point (t_0, ω_0).

While the time-frequency atoms of the spectrogram-based energy analysis only match the position of the dispersion curve, a chirplet is concentrated on a straight line with the same rate as the dispersion curve. This is visualized in Fig. 7(b). As a result, the chirplet-based energy analysis matches the group delay of the signal as a first-order approximation (in contrast to the zeroth-order approximation of the spectrogram). Either a time or a frequency shear operator can be used to realize this matching of the chirplet onto the dis-

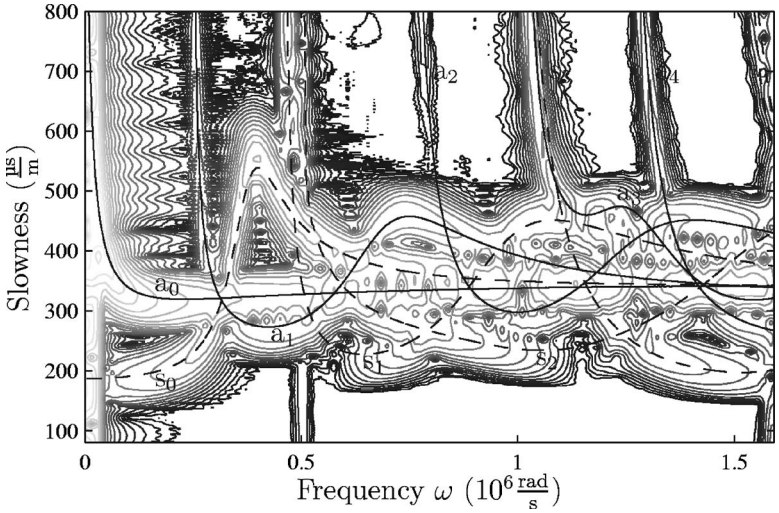


FIG. 5. Spectrogram of the synthetic time signal with theoretical dispersion curves.

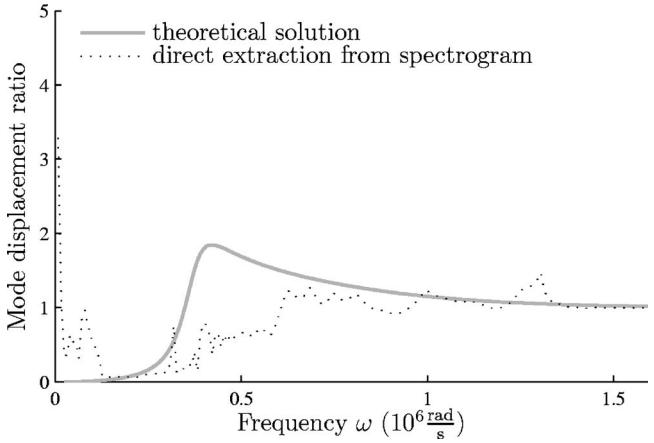


FIG. 6. Mode displacement ratio of mode s_0 over mode a_0, obtained by directly evaluating the spectrogram at the dispersion curves. Compare to the theoretical displacement ratio (solid gray line).

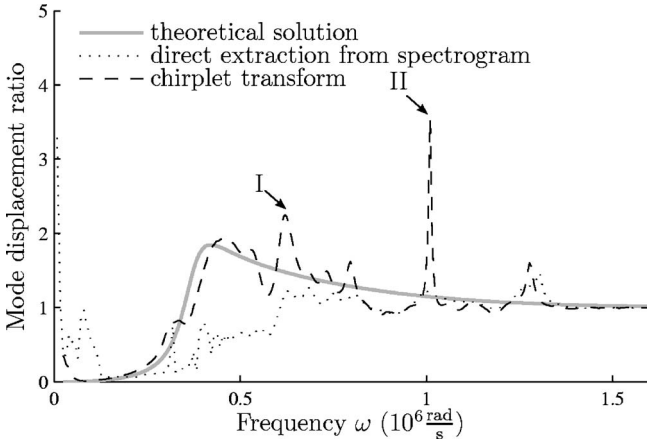


FIG. 8. Mode displacement ratio of mode s_0 over mode a_0, obtained by using time sheared chirplets which are concentrated along the dispersion curves of the modes. Compare to theoretical displacement ratio (solid gray line). (I) and (II) denote large discrepancies of the chirplet calculations and the theoretical displacement ratio, caused by the intersection of the dispersion curves of modes s_0 and a_1 (I), and by the proximity of modes s_0 and a_0 (II), respectively.

persion curve, since each operation shears one of the semiaxes of the time-frequency region of averaging. Since the dispersion curves are functions of frequency, only the time shear operation is used in this research to avoid singularities. An extension of the chirplet transform to shear operations of higher order allows for an adaption of a chirplet to dispersion curves of order higher than one, but does not show significant improvements in the accuracy of the mode displacement ratio calculations,[18] so this research only applies first order shear, and does not use one of the five possible chirplet operations (degrees of freedom).

This chirplet transformation (using chirplets which are concentrated along the dispersion curves by shearing them in the time direction) is used to calculate the mode displacement ratio of s_0 over a_0, which is compared to the theoretical and spectrogram-based results in Fig. 8.

It is important to note that the main advantage of the chirplet over the spectrogram (and scalogram) is this additional control over the shape, location, and orientation of the time-frequency atoms. All these TFRs will provide an average energy over a region in the time-frequency domain, but the chirplet provides additional degrees of freedom to control this averaging region. As a result, the chirplet should be better in resolving closely spaced modes.

C. Adaptive chirplets

Inspection of the chirplet transform results in Fig. 8 shows that the recovered mode displacement ratio is a better match to the theoretical solution than the spectrogram-based results. However, there is poor agreement between the chirplet-based and theoretical solution in the frequency ranges where two dispersion curves intersect (compare to the dispersion curves of the plate shown in Figs. 3 and 10). The regions of greatest discrepancy in Fig. 8 are indicated by (I) and (II), and occur at frequencies where the s_0-mode's dispersion curve intersects with the dispersion curve of the a_1 mode, or where the s_0- and a_0-modes are close to each other. A large portion of those modes' energy must have been improperly assigned to the s_0-mode in these frequency ranges using this procedure.

This error can be addressed by exploiting another (a fourth) degree of freedom of the chirplet transform—the scale s. An improved, adaptive chirplet analysis algorithm can also account for a mode's intersections (and closeness) to its neighboring modes, in addition to its own dispersion curve model. As in the previous chirplet algorithm, chirplets

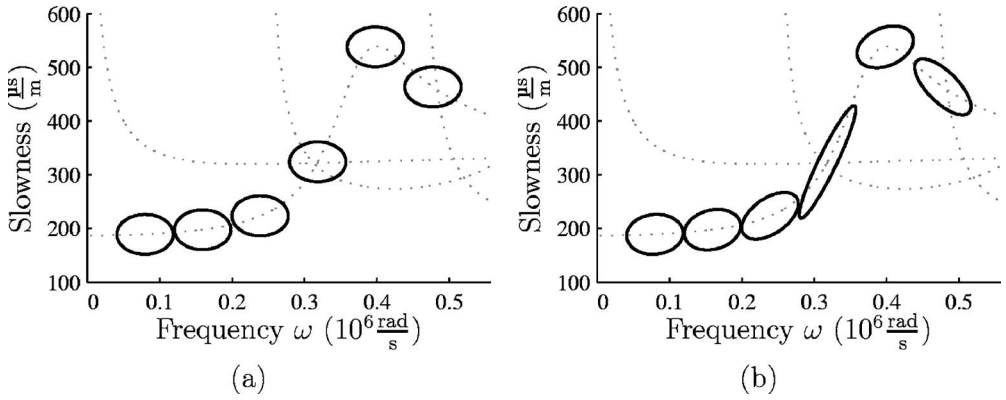


FIG. 7. (a) Regions over which time-frequency content is averaged when evaluating the spectrogram on the dispersion curve of mode s_0. (b) Time-sheared chirplets, concentrated along the dispersion curve of the s_0 mode.

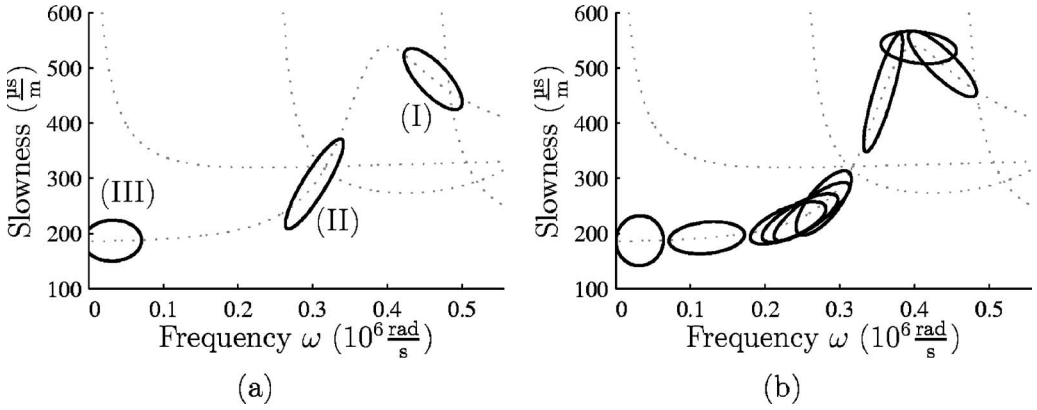


FIG. 9. Region of influence of chirplets concentrated along the s_0 mode. (a) Chirplets interfering (I) with an intersecting dispersion curve, (II) with neighboring and intersecting dispersion curves, and (III) out-of-range chirplet. (b) Adaptive chirplets to avoid interferences.

are generated from a Gaussian window centered on a point on the dispersion curve of the mode under consideration (s_0 or a_0), and time sheared as to be concentrated along the curve. A standard scale s is then chosen analogous to the window length of a STFT. For this work $s=1.3$ rad. Next, the chirplets generated under this procedure are examined to determine whether they interfere with other dispersion curves. As all window functions have an infinite extent (width) in either the frequency or time dimension (or both, like the Gaussian window used in this work), a threshold is applied to the window, under which its value is assumed to be 0. In this research, the finite extent (area) of the Gaussian window is assumed to be 2.8 times the standard deviation, so that 99.9% of the energy of the window is located within this area. With this criteria, an interference is defined as another dispersion curve lying within this 99.9% area.

If the resulting chirplet or its region of averaging interferes with a dispersion curve of another mode, its scaling factor is varied to avoid interferences. This scaling stretches the chirplet in one direction (either time or frequency) and shrinks it reciprocally in the other; this is demonstrated in Fig. 9. No scaling factor, however, can avoid interference for points near the intersection of two dispersion curves, or on portions of a dispersion curve where other dispersion curves are close. In these instances, the energy distribution cannot be calculated properly for the corresponding frequencies, so the chirplet analysis is not performed for these frequencies. Figure 10 indicates the regions of each dispersion curve where these adaptive, interference-free chirplets can be designed. Finally, note that these limitations are directly related to a wave's propagation distance. For a larger propagation distance d, a window with a certain time and frequency extent (area) has a smaller slowness dimension—as slowness is inversely proportional to propagation distance. This implies that, for larger propagation distances, it is easier to fit chirplets on dispersion curves while avoiding interferences. Figure 10 shows these frequency intervals for propagation distances of $d=90$ mm (black lines), and $d=900$ mm (gray lines).

The adaptive chirplet algorithm is applied to the same synthetic Lamb wave signal, and the results are shown in Fig. 11. Since this s_0-mode distribution is normalized to mode a_0, mode displacement ratio values are only calculated in frequency regions where both the s_0 and a_0 energies can be calculated without interference. Figure 11 shows that the chirplet transform is capable of accurately extracting the (proportional) energy distribution of an individual mode from a multimode signal. Normalization to the a_0-mode em-

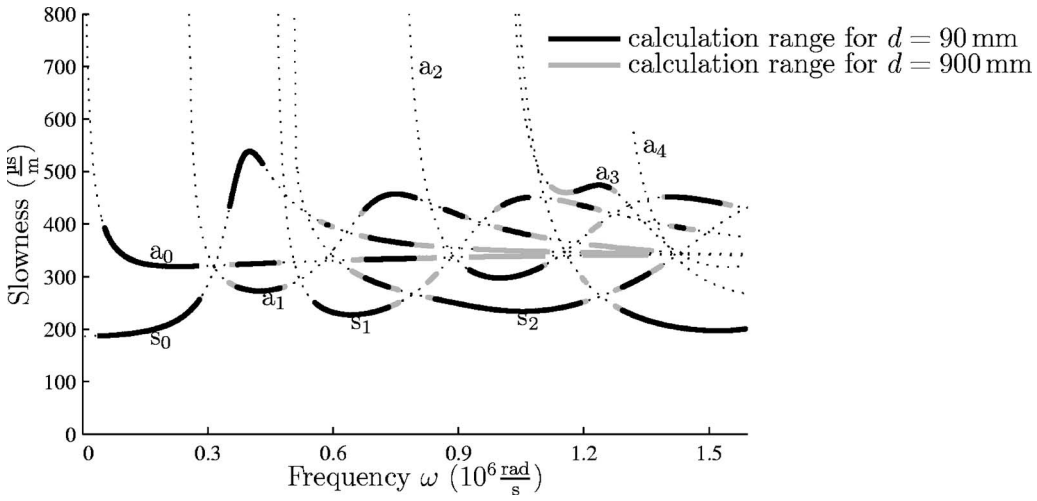


FIG. 10. Dispersion curves (dotted lines), indicating the regions where adaptive chirplets can be applied to the synthetic time signal without interferences with other modes for propagation distance $d=90$ mm (black lines) and $d=900$ mm (gray lines).

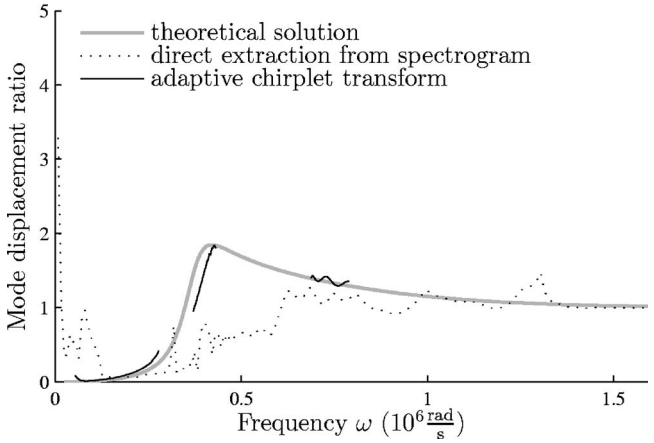


FIG. 11. Mode displacement ratio of mode s_0 over mode a_0, obtained by using adaptive chirplets (solid black line). Results only exist where both a_0 and s_0 modes can be evaluated. Compare to theoretical displacement ratio (solid gray line), and the previously obtained results by directly evaluating the spectrogram (dotted line).

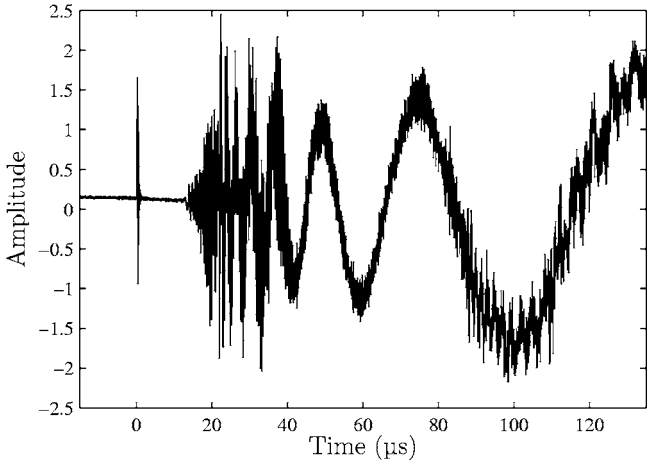


FIG. 12. Typical time-domain signal (out-of-plane particle velocity) measured in the perfect plate.

phasizes small discrepancies, but the computed mode displacement values are very close to the exact (theoretical) values.

The adaptive chirplet algorithm procedure yields a quantitative prediction of the mode energy, and uses the uncertainty principle to identify the slowness-frequency regions where interference will occur. This research takes a conservative approach to identifying these regions, and an advantage of the proposed adaptive algorithm is that the user can quantitatively define the interference criteria for a particular application. For example, a user could specify that certain modes do not exist (or at least do not carry significant energy) for higher frequencies, so no interference will occur with these modes. This would lead to an expanded frequency region in which to calculate energy content; note that this adds credibility to the fact that the energy content for higher frequencies in Fig. 8 is actually correct (but excluded in Fig. 11).

V. APPLICATION TO EXPERIMENTALLY MEASURED SIGNALS

As a final step, the accuracy and robustness of the adaptive chirplet algorithm is demonstrated on real, experimentally measured Lamb wave signals with an application of the correlation technique developed in previous research.[4,10] Benz *et al.*[10] formulated this correlation technique using the reflected Lamb wave field to locate a notch in a 0.99 mm thick, 3003 aluminum plate. These Lamb wave signals are generated with a pulse laser (noncontact, pointlike, broad band generation), and detected with a dual probe laser interferometer (noncontact, high fidelity, pointlike detection) in both a notched and a perfect plate. Figure 12 shows a typical time-domain signal (out-of-plane particle velocity) measured in the perfect plate; note that the spike at $t=0$ is spurious, and corresponds to the firing of the pulse laser. The correlation procedure of Benz *et al.*[10] uses the spectrogram to calculate the energy reflected by the notch in a certain frequency band (300 kHz–2 MHz).

In brief, their technique examines a correlation of the spectrograms measured in a notched plate (known location of this notch is $\Delta d_0' = 29.2$ mm) and in a perfect plate—each calculated with different, assumed propagation distances, $\Delta d_0'$. Reflections (from the notch and any plate edges) will introduce local maxima in the correlation curves at certain propagation distances. These local maxima occur when the reflected modes (within a spectrogram) coincide with the incident modes, which in turn provide a notch (or plate edge) location distance. A ratio of the notched to perfect plate correlations removes the maxima due to plate edges (the edges are present in both notched and plate specimens) and will show maxima only at "reflecting" features that are present in the notched plate, but not in the perfect plate. The notch is the only such feature in these experiments.

Benz[19] notes that the a_0-mode carries a large portion of the energy of a Lamb wave and is less affected by mode conversion at the notch. This implies that correlation calculations should focus on that particular mode, and this goal is achieved in Ref. 10 by restricting the spectrogram calculation to a limited frequency band (namely, 300 kHz–2 MHz).

In contrast, the chirplet transform enables the calculation of the energy of each mode individually, which will be used to formulate an alternative notch localization technique. In a fashion similar to Ref. 10, the adaptive chirplet transform is applied to an experimentally measured time-domain signal for a number of assumed propagation distances, $\Delta d_0'$. In other words, adaptive chirplet analysis is performed on the a_0-dispersion curve in slowness-frequency coordinates according to these propagation distances. This results in energy distributions over frequency for each distance. Instead of correlating spectrogram images (as in Ref. 10), the correlation of energy distributions of mode a_0 for varying assumed propagation distances is calculated.

Figure 13 shows two ratio curves (ratio of the correlations of the notched to perfect plates). One ratio curve is determined with the adaptive chirplet algorithm developed in this research, while the other curve is based on the reassigned spectrograms as described by Benz *et al.*[10] Both algorithms lead to very accurate notch localization—they both contain peaks at the correct notch location distance of $\Delta d_0'$

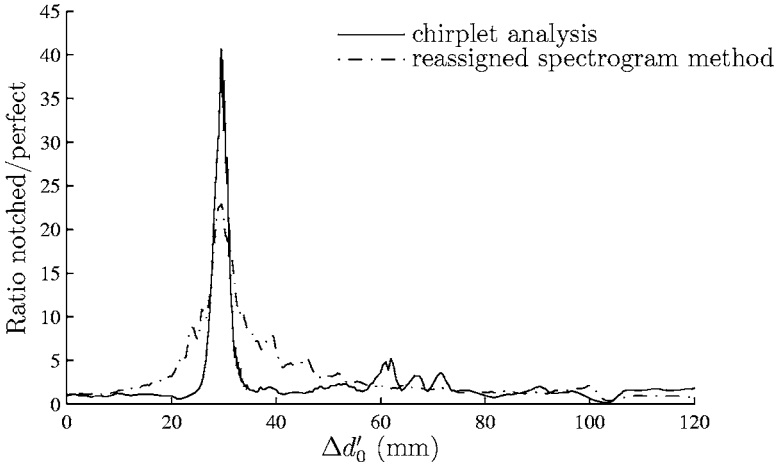


FIG. 13. Ratio curves for notch localization performed with a correlation method based on the reassigned spectrogram and the adaptive chirplet algorithm, respectively.

=29.2 mm to within ±0.4 mm, an error of 1%. However, the ratio curve calculated with the adaptive chirplet algorithm has a peak that is much sharper and higher, providing a more definitive indication of the notch location.

VI. CONCLUSION

This research demonstrates the effectiveness of using the chirplet transform as a signal processing tool to quantitatively determine the energy distribution of the individual modes of a multimode, dispersive Lamb wave. The proposed adaptive chirplet algorithm is grounded in the theory developed by Mann,[5] and is model-based; this work uses a model of the dispersion relationship (the Rayleigh-Lamb equations) to assign wave energy to the individual propagation modes by fitting chirplets to these known modelines. An adaptive algorithm then quantitatively determines the frequency regions (for each mode) in which reliable calculations of mode energy are possible, and employs chirplets (locally adapted) to extract the (proportional) energy distribution of that mode from a dispersive, time-domain wave signal.

This adaptive chirplet algorithm is first validated on a synthetic Lamb wave signal (time-domain, multimode, dispersive), and shows its effectiveness and accuracy for the quantitative measurement of mode energy content.

The robustness of this adaptive chirplet algorithm is then demonstrated on experimentally measured Lamb wave signals. This application enables a demonstration of another advantage of the proposed adaptive chirplet algorithm—the ability to tailor the chirplet atoms to a particular mode, contained within a multimode signal. The correlation procedure developed by Benz *et al.*[10] is modified for the chirplet algorithm and used to locate a notch in a plate with a high degree of accuracy and confidence.

ACKNOWLEDGMENTS

The Deutscher Akademischer Austausch Dienst (DAAD) provided partial support to Helge Kuttig. The National Science Foundation provided partial support to Laurence J. Jacobs under Grant No. CMS-0201283.

[1]D. E. Chimenti, "Guided waves in plates and their use in materials characterization," Appl. Mech. Rev. **50**, 247–284 (1997).
[2]M. Niethammer, L. J. Jacobs, J. Qu, and J. Jarzynski, "Time-frequency representation of Lamb waves," J. Acoust. Soc. Am. **109**, 1841–1847 (2001).
[3]S. Hurlebaus, L. Gaul, and L. J. Jacobs, "Localization of a 'synthetic' acoustic emission source on the surface of a fatigue specimen," Res. Nondestruct. Eval. **13**, 105–117 (2001).
[4]S. Hurlebaus, M. Niethammer, L. J. Jacobs, and C. Valle, "Automated methodology to locate notches with Lamb waves," ARLO **2**, 97–102 (2001).
[5]S. Mann and S. Haykin, "The chirplet transform: A generalization of Gabor's logon transform," Vision Interface **91**, 205–212 (1991).
[6]R. G. Baraniuk and D. L. Jones, "Wigner-based formulation of the chirplet transform," IEEE Trans. Signal Process. **44**, 3129–3135 (1996).
[7]S. Mann and S. Haykin, "The chirplet transform: Physical considerations," IEEE Trans. Signal Process. **43**, 2745–2761 (1995).
[8]R. Gribonval, "Fast matching pursuit with a multiscale dictionary of Gaussian chirps," IEEE Trans. Signal Process. **49**, 994–1001 (2001).
[9]J.-C. Hong, K. H. Sun, and Y. Y. Kim, "Dispersion-based short-time Fourier transform applied to dispersive wave analysis," J. Acoust. Soc. Am. **117**, 2949–2960 (2005).
[10]R. Benz, M. Niethammer, S. Hurlebaus, and L. J. Jacobs, "Localization of notches with Lamb waves," J. Acoust. Soc. Am. **114**, 677–685 (2003).
[11]O. Kotte, M. Niethammer, and L. J. Jacobs, "Lamb wave characterization by differential reassignment and nonlinear anisotropic diffusion," NDT & E Int. **39**, 96–105 (2006).
[12]S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1999).
[13]L. Cohen, *Time-Frequency Analysis* (Prentice-Hall, New Jersey, 1995).
[14]The Wigner-Ville distribution does not result in any smoothing of the time-frequency representation, but is not guaranteed to be positive and suffers from signal interferences.
[15]J. D. Achenbach, *Wave Propagation in Elastic Solids* (Elsevier, New York, 1973).
[16]R. L. Weaver and Y. H. Pao, "Axisymmetric elastic waves excited by a point source in a plate," Trans. ASME, J. Appl. Mech. **49**, 821–836 (1982).
[17]P. Wilcox, "Modelling the excitation of Lamb and SH waves by point and line sources," in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti, AIP Conf. Proc. (AIP, New York, 2004), Vol. **23A**, pp. 206–213.
[18]H. Kuttig, "Model-based signal processing of dispersive waves with chirplets," Diplomarbeit thesis, Institute A for Mechanics, University of Stuttgart (2005).
[19]R. Benz, "Localization of notches with Lamb waves," Master thesis, School of Civil and Environmental Engineering, Georgia Institute of Technology (2002).

Active vibration control using an inertial actuator with internal damping

Christoph Paulitsch,[a] Paolo Gardonio, and Stephen J. Elliott
Institute of Sound and Vibration Research, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

(Received 20 April 2005; revised 14 October 2005; accepted 2 November 2005)

Collocated direct velocity feedback with ideal point force actuators mounted on structures is unconditionally stable and generates active damping. When inertial actuators are used to generate the control force, the system can become unstable even for moderate velocity feedback gains due to an additional −180° phase lag introduced by the fundamental axial resonant mode of the inertial actuator. In this study a relative velocity sensor is used to implement an inner velocity feedback loop that generates internal damping in a lightweight, electrodynamic, inertial actuator. Simulation results for a model problem with the actuator mounted on a clamped plate show that, when internal relative velocity feedback is used in addition to a conventional external velocity feedback loop, there is an optimum combination of internal and external velocity feedback gains, which, for a given gain margin, maximizes vibration reduction. These predictions are validated in experiments with a specially built lightweight inertial actuator. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2141228]

PACS number(s): 43.40.Vn, 43.38.Dv [KAC] Pages: 2131–2140

I. INTRODUCTION

Much recent research work on active structural acoustic control has been devoted to the development of smart panels with arrays of piezoelectric sensor and actuator transducers.[1–7] However, in some cases active vibroacoustic control with point force actuators may be more efficient than using piezoelectric patch actuators.[7] In practical applications point forces are normally generated by either reactive or inertial actuators.[8] Reactive actuators require a base structure to react off, which is usually available in very few practical applications. Inertial actuators are free from this requirement and thus present a much more appealing solution. However, the control force that can be generated is strictly linked to the inertial mass, which has to be small in order to reduce the weight of the whole system.[9] Moreover the phase shift associated with the fundamental resonance frequency of these actuators causes instability when high-gain velocity feedback is implemented. In order to reduce this problem, internal velocity feedback has previously been used to stabilize inertial actuators[10] and increase the gain margin.[11,12] Recent studies have underlined that this additional feedback loop is also beneficial in that it adds damping to the flexible structure to be controlled.[13] However, for high gains in the internal velocity feedback loop, the inertial actuator simply acts like an attached mass. Hence, there is an optimum internal velocity feedback gain for which maximum vibration reduction is obtained.[10] Also, when both the external and internal feedback loops are implemented there is an optimum combination of control gains for which maximum vibroacoustic control is achieved.[14]

This study presents a lightweight, inertial actuator[9] which has been mounted on a clamped plate in order to implement both relative velocity feedback, using an internal velocity sensor, and plate velocity feedback, using a closely located external sensor. Both feedback loops are used in such a way as to reduce plate vibrations at low resonance frequencies of the plate by adding active damping. The internal feedback loop is also meant to increase the stability properties of the external loop. Both simulations and experiments of this model problem are presented.

Section II presents a lumped parameter model of the inertial actuator attached to a clamped plate. Also, a reduced two degrees of freedom (DOF) model for the actuator and plate is considered in such a way as to provide a better insight into the beneficial effects of internal damping for inertial actuators which are used to implement velocity feedback on structures. Section III A compares experiments and simulations using external velocity feedback only while Sec. III B shows the effect of internal velocity feedback on the external open velocity feedback loop. Finally Sec. III C shows with simulations and experiments that there is a best combination of external and internal velocity feedback for maximum vibration reduction of each resonant mode of the plate at the example of the first resonance.

II. MODEL

The steady state flexural response of a rectangular panel clamped along the four edges which, as shown in Fig. 1, is excited by a harmonic primary force, will be used as model problem to illustrate the active damping effects produced by an electrodynamic inertial actuator with internal and external velocity feedback control loops. As schematically shown in Fig. 2, the inertial actuator is equipped with a double coil system such that the "primary" coil is used to generate the control force and the "secondary" coil is used to detect the relative vibration between the base of the actuator and the

[a]Electronic mail: cpaulits@gmx.de

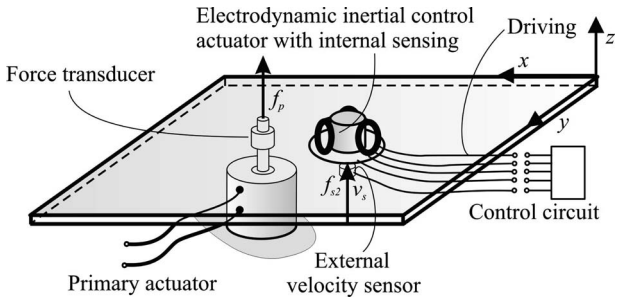


FIG. 1. Sketch of the model problem where a test panel is connected to an electrodynamic, inertial actuator with internal velocity sensor which is set to control the vibrations generated by a primary force disturbance using either external, internal or combined velocity feedback.

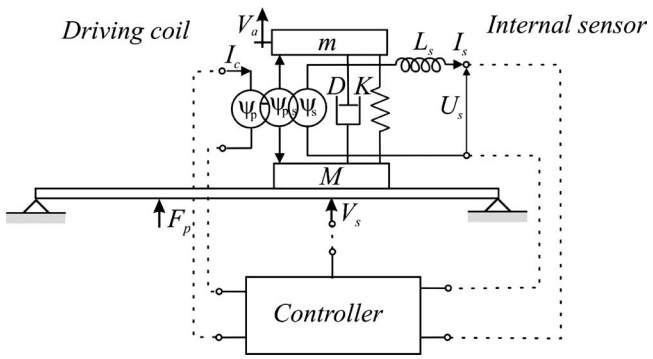


FIG. 2. Model of the electrodynamic actuator connected to the plate and reacting against a moving mass.

inertial mass, which mainly consists of the permanent magnet. The material and geometrical properties of the aluminium panel test rig and control actuator are summarized in Table I. The steady state response of the panel has been derived assuming the primary force disturbance to be harmonic, with time dependence of the form $\mathrm{Re}\{\exp(j\omega t)\}$ where ω is the circular frequency and $j=\sqrt{-1}$. The mechanical and electrical functions in the model have therefore been taken to be the real part of anticlockwise rotating complex vectors, i.e., phasors, given in the form $X(\omega)e^{j\omega t}$, where $X(\omega)$ is the phasor at $t=0$.

The phasor of the complex velocity at the error sensor, $V_s(\omega)$, can be expressed in terms of the phasors of the primary, $F_p(\omega)$, and secondary, $F_{s2}(\omega)$, force excitations with the following mobility relation;

$$V_s = Y_{sp}F_p + Y_{ss}F_{s2}, \tag{1}$$

where the two mobility functions, $Y_{sp}(\omega)$ and $Y_{ss}(\omega)$, can be expressed with the following modal expansions:

$$Y_{sp} = j\omega\sum_{n=1}^{N} \frac{\phi_n(x_s,y_s)\phi_n(x_p,y_p)}{\rho l_x l_y h[\omega_n^2 + 2j\omega\xi_n\omega_n - \omega^2]}, \tag{2a}$$

$$Y_{ss} = j\omega\sum_{n=1}^{N} \frac{\phi_n(x_s,y_s)\phi_n(x_s,y_s)}{\rho l_x l_y h[\omega_n^2 + 2j\omega\xi_n\omega_n - \omega^2]}. \tag{2b}$$

In these equations (x_p,y_p) and (x_s,y_s) are the coordinates of the primary and control positions, ρ is the density of the material, and ξ_n is the viscous damping ratio of the nth mode. Finally ω_n and $\phi_n(x,y)$ are the nth natural frequency and natural mode. The natural frequencies of a clamped plate[15] are adjusted by a factor $\sqrt{K_1/K_2}$ for nonideally clamped boundary conditions. In the experimental setup used for this study, the primary force is generated by a primary shaker and is measured with a force transducer that adds a mass M_p to the plate at position (x_p,y_p) so that the transfer mobility of the plate is modified to be

$$Y'_{sp} = \frac{Y_{sp}}{1 + j\omega M_p Y_{pp}}.$$

The secondary force at position (x_s,y_s) is generated by an electrodynamic inertial actuator[9] that reacts against a moving mass m suspended on a spring with stiffness K and damping D. Due to the mass of the shaker housing and the fixed coil

TABLE I. Parameter values used for simulations.

Parameter	Value	Description
$l_x \times l_y \times h$	$0.414 \times 0.314 \times 0.001$	Plate length × width × thickness in m
x_p, y_p	0.345 m, 0.254 m	x,y coordinates of primary force position
x_s, y_s	0.225 m, 0.12 m	x,y coordinate of secondary force position
ρ	2700 kg/m^3	Mass density of plate
K_1	1.2037	Coefficient to adjust for boundary conditions
K_2	0.7042	Coefficient to adjust for boundary conditions
ξ_n	0.002–0.3	Range of viscous damping coefficients
E	71 10^9 N/m^2	Young's modulus of plate
ν	0.33	Poisson's ratio
M_p	0.03 kg	Added mass of force transducer
M	0.008 kg	Housing mass of inertial actuator
m	0.022 kg	Moving mass of inertial actuator
K	170 N/m	Suspension stiffness of inertial actuator
D	0.3 N/m s^{-1}	Suspension damping of inertial actuator
Ψ_p	2.6 N/A	Transducer coefficient of primary driving coil
Z_s	1.22	Shunted impedance to secondary coil
L_s	0.03 mH	Secondary coil inductance
L_{sp}	0.5 mH	Mutual inductance between primary and secondary coil
Ψ_s	1.146 V s/m	Transducer coefficient of secondary coil.

M the point mobility of the plate at the control position (x_s, y_s) is modified to be

$$Y'_{ss} = \frac{Y_{ss}}{1 + j\omega M Y_{ss}}.$$

Considering the lumped parameter model of the actuator shown in Fig. 2 the velocity of the actuator mass m is given by

$$V_a = Y_m F_{s1}, \tag{3}$$

where $Y_m = 1/j\omega m$ is the mobility of the actuator mass and F_{s1} is the force between the moving mass m and the actuator suspension. The transmitted force on the plate and by the actuator mass are combined in a vector

$$\mathbf{f}_s = \begin{Bmatrix} F_{s2} \\ F_{s1} \end{Bmatrix} = -\begin{bmatrix} Z & -Z \\ -Z & Z \end{bmatrix} \begin{Bmatrix} V_s \\ V_a \end{Bmatrix} + \begin{Bmatrix} F_a \\ -F_a \end{Bmatrix} = -\mathbf{Z}\mathbf{v}_s + \mathbf{f}_a, \tag{4}$$

where $Z = (K/j\omega) + D$ is the impedance of the coil suspension that is in parallel to an actuation force vector $\mathbf{f}_a = \left\{ \begin{smallmatrix} F_a \\ -F_a \end{smallmatrix} \right\}$. Taking into account the mass and suspension effects of the inertial actuator, Eq. (1) is modified to become

$$\mathbf{v}_s = (\mathbf{I} + \mathbf{Y}_{ss}\mathbf{Z})^{-1}(\mathbf{y}_{sp}F_p + \mathbf{Y}_{ss}\mathbf{f}_a), \tag{5}$$

where the transfer and point mobility matrices are $\mathbf{y}_{sp} = \{Y'_{sp}\ 0\}^{\mathbf{T}}$, $\mathbf{Y}_{ss} = \left[\begin{smallmatrix} Y'_{ss} & 0 \\ 0 & Y_m \end{smallmatrix} \right]$, $\mathbf{v}_s = \{V_s\ V_a\}^{\mathbf{T}}$ and $\mathbf{I}$ is a 2×2 identity matrix.

The primary and the secondary coils are immersed in a constant horizontal magnetic field that is generated by the permanent magnet acting as the moving mass. When there is an electrical current I_c in the primary coil and I_s in the secondary coil, an axial actuation Lorentz's force

$$F_a = \Psi_p I_c + \Psi_s I_s \tag{6}$$

is created between the moving mass m and the housing mass M connected to the plate, where Ψ_p and Ψ_s are transducer coefficients of the primary and the secondary coil respectively.

In order to implement the external velocity feedback loop, the plate vibration V_s at the base of the actuator is measured and fed back to the primary coil. The primary coil is then driven by a current $I_c = -G_1 V_s$, where G_1 is the feedback gain. The closed loop stability of external velocity feedback is evaluated by investigating the open loop frequency response function (FRF)

$$\frac{V_s}{I_c} = \{1\ 0\}(\mathbf{I} + \mathbf{Y}_{ss}\mathbf{Z})^{-1}\mathbf{Y}_{ss}\boldsymbol{\psi}_p^{\mathbf{T}}, \tag{7}$$

where $\boldsymbol{\psi}_p = \{\Psi_p - \Psi_p\}$. The vibration reduction at the base of the actuator per unit primary excitation force is then given by the closed loop FRF,

$$\frac{V_s}{F_p} = \{1\ 0\}\left(\mathbf{I} + \mathbf{Y}_{ss}\left(\mathbf{Z} + \Psi_p \begin{bmatrix} G_1 & 0 \\ -G_1 & 0 \end{bmatrix}\right)\right)^{-1}\mathbf{y}_{sp}. \tag{8}$$

According to Biot-Savart's Law[16] the currents in the coils will also generate a time-varying magnetic B field, primarily in axial direction, so that a voltage

$$U_s = L_s j\omega I_s + L_{sp} j\omega I_c + \boldsymbol{\psi}_s \mathbf{v}_s \tag{9}$$

will be induced in the secondary coil where L_s is the self-inductance of the secondary coil, L_{sp} is the coupling or mutual inductance between the primary and the secondary coil and $\boldsymbol{\psi}_s = \{\Psi_s - \Psi_s\}$. As indicated in Fig. 2, the primary coil is used to drive the actuator with a current I_c that is controlled by a current command amplifier. The secondary coil circuit is connected to the electrical input impedance of the controller $Z_s = -U_s/I_s$ and it is used to sense the relative velocity. Using Eqs. (5), (6), and (9), with $F_p = 0$, the FRF relative to the internal open loop becomes

$$\frac{U_s}{I_c} = \frac{L_{sp} j\omega + \boldsymbol{\psi}_s \mathbf{Y}_{ss} \boldsymbol{\psi}_p^{\mathbf{T}}}{1 + \dfrac{L_s j\omega}{Z_s} + \dfrac{\boldsymbol{\psi}_s \mathbf{Y}_{ss} \boldsymbol{\psi}_s^{\mathbf{T}}}{Z_s}}. \tag{10}$$

In order to obtain the wanted relative velocity sensing, the second term in the numerator should be dominant, but the other terms exert an unwanted effect especially at higher frequencies as will be found in Sec. III. When the secondary coil voltage is fed back to the current in the driving coil $I_c = -G_2 U_s$, the stability behavior of external velocity feedback is modified so that Eq. (7) with internal feedback becomes

$$\frac{V_s}{I_c} = \left(\mathbf{I} + \mathbf{Y}_{ss}\left(\mathbf{Z} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \frac{\Psi_p \Psi_s G_2 Z_s}{Z_s + j\omega(L_s + G_2 L_{sp} Z_s)}\right)\right)^{-1} \times \mathbf{Y}_{ss} \boldsymbol{\psi}_p^{\mathbf{T}}. \tag{11}$$

Also, the vibration reduction performance is changed from Eq. (8) to the following expression:

$$\frac{V_s}{F_p} = \left(\mathbf{I} + \mathbf{Y}_{ss}\left(\mathbf{Z}' + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \frac{\Psi_p \Psi_s G_2 Z_s}{Z_s + j\omega(L_s + G_2 L_{sp} Z_s)}\right)\right)^{-1} \times \mathbf{y}_{sp}, \tag{12}$$

where $\mathbf{Z}' = \mathbf{Z} + \left[\begin{smallmatrix} \Psi_p G_1 & 0 \\ -\Psi_p G_1 & 0 \end{smallmatrix} \right]$. The stability of the external velocity feedback loop with the inertial actuator is evaluated investigating Eqs. (7) and (11) where additional internal actuator damping is added at lower frequencies using internal velocity feedback.

In order to simplify the interpretation of the results, in particular the beneficial influence of the internal actuator damping on a given mode of the panel, in the following section a simplified two DOF model will also be considered. In this model the response of the rth mode of the panel is represented by a modal mass $M_r = \rho l_x l_y h/\phi_r^2(x_s, y_s)$, modal stiffness $K_r = \omega_r^2 M_r$ and viscous damping $D_r = 2\xi_r \sqrt{K_r M_r}$ and the dynamics of the actuator are represented by its axial

FIG. 3. Photography of the experimental setup with the primary shaker below the plate in the background and the inertial actuator above the plate in the foreground.

suspension stiffness K, damping D and inertial mass m. The response of such a simplified system is derived from the second part of Eq. (5) when only the rth mode of the plate, i.e., $N=r$ and $n=r$ in Eq. (2b), and the fundamental resonance of the inertial actuator are used. In order to check the closed loop performance and stability, the open loop frequency response function (FRF) between the actuation force and the vibration velocity of this simplified model

$$\frac{V_s}{F_a} = \frac{-\omega^4/j\omega M_r}{\left(\omega_r^2 + 2\xi_r\omega_r j\omega + (\omega_0^2 + 2\xi_0\omega_0 j\omega)\frac{m}{M_r} - \omega^2\right)(\omega_0^2 + 2\xi_0\omega_0 j\omega - \omega^2) - (\omega_0^2 + 2\xi_0\omega_0 j\omega)^2\frac{m}{M_r}} \tag{13}$$

is investigated where $\xi_0 = D/(2\sqrt{Km})$ is the viscous damping ratio and $\omega_0 = \sqrt{K/m}$ the undamped natural frequency of the inertial actuator.

In the following section the effect of increased internal damping on external velocity feedback performance is studied for the model problem. Equations (7), (8), and (10)–(12) are investigated for different values of the feedback gains G_1 and G_2 and the simplified model in Eq. (13) is used to illustrate the findings by analytical results. The model parameters that give reasonable agreement with measurements at the test set-up shown in Fig. 3 are listed in Table I. Geometric dimensions x_s, y_s, x_p, y_p, l_x, l_y, and h are approximately known from the setup, mass density ρ, and Young's modulus E are usual values for aluminum and approximate actuator parameters are known from a previous identification stage[9,17] where a band limited "white noise" excitation has been used to measure several FRFs of each actuator. The viscous damping coefficients and the boundary condition adjustment factors are tuned manually as to obtain good visible agreement between measurements and simulations.

III. RESULTS

Figure 4 shows the three circuits that are used to implement (a) external velocity feedback, (b) internal velocity feedback, and (c) combined external and internal velocity feedback. A B&K 4375 accelerometer is used together with an integrating B&K 2635 charge amplifier to obtain the plate velocity and the amplification for the external velocity feedback loop. Because of the mutual inductance effect a KEMO VBF8 filter is used to appropriately cut off the secondary coil voltage signal and to adjust the internal velocity feedback gain. An OPA549 in a current command circuit together with a DC300 amplifier drives the control actuator via the primary coil.

A. External velocity feedback (type a)

Figure 5 shows the Nyquist plot of the simulated (left) and measured (right) open loop FRF between the primary coil current I_c and the plate velocity at the control actuator position V_s. Both measurements and simulations indicate that the locus lies in the positive real half-plane for all frequencies except for frequencies around the fundamental resonance of the inertial actuator. The locus intersects with the real axis at approximately the fundamental resonance frequency at about 13 Hz. This behavior is highlighted by the simplified 2 DOF analytical model of Eq. (13) that has one loop in the left real half-plane due to the fundamental resonance of the inertial actuator with resonance frequency ω_0 and a single loop in the right real half-plane due to the rth plate mode with resonance frequency ω_r.

According to the Nyquist criterion, closed loop stability is guaranteed if the locus of the open loop FRF in the real-imaginary plane does not encircle the critical point $(-1,0j)$. Defining δ_0 as the real part of the FRF where its locus intersects with the negative real axis as shown in Fig. 5, in the studied case, it is sufficient that $|\delta_0| < 1$. The amplitude of the FRF at this point $|\delta_0|$ is inversely proportional to the gain margin and hence a measure for stability.[5,18] The smaller the gain margin the closer the locus is to the critical point and more control spillover is expected in the closed loop case.

In measurements there is additional phase shift at very low frequencies due to the electronic integrator in the charge amplifier. Hence, because of the inertial actuator resonance

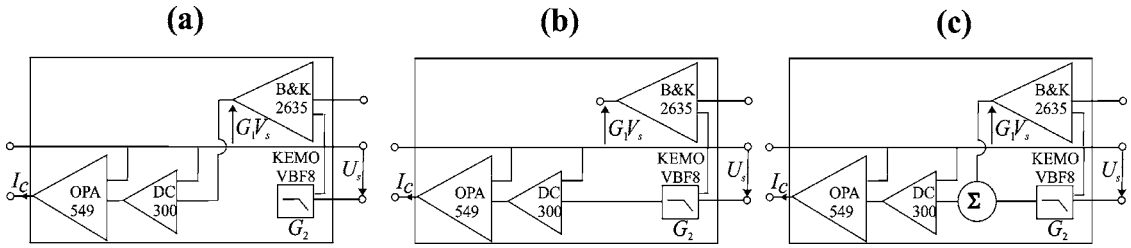


FIG. 4. Feedback circuits added between the secondary and primary coil with current command amplifier and the velocity sensor for external velocity feedback (left), internal velocity feedback (center), and combined feedback (right).

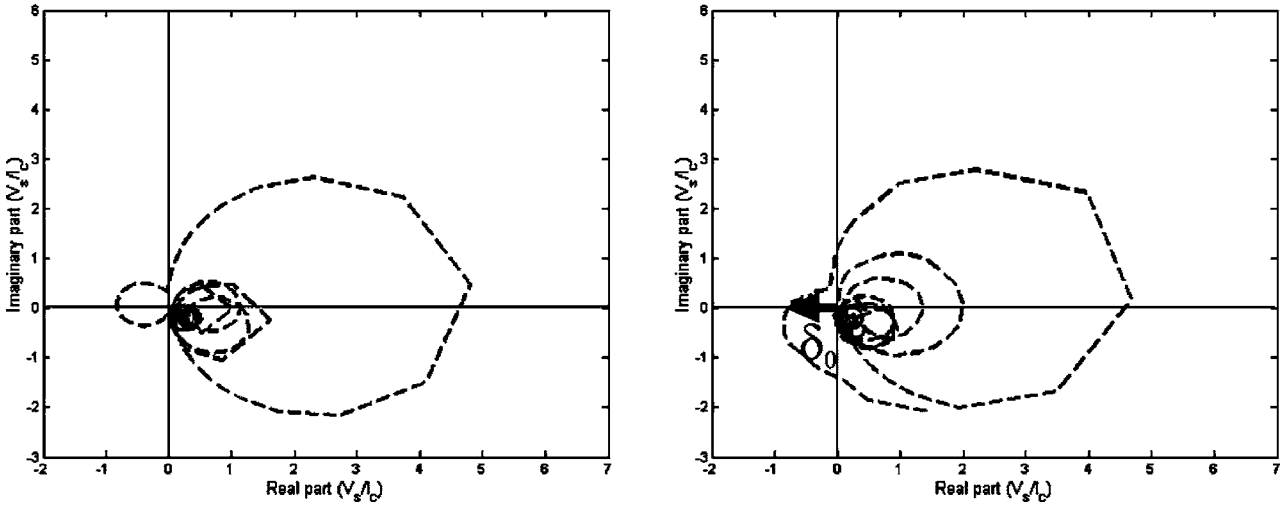


FIG. 5. Nyquist plot of the simulated (left) and measured (right) FRF between the primary coil current and the plate velocity at the actuator position, for the external feedback loop.

leading to $\delta_0<0$ and because of the charge amplifier distortion only a limited range of control gains is predicted to guarantee closed loop stability. Also, control spillover effects are anticipated in the vicinity of the fundamental resonance frequency since the locus enters the circle with radius 1 about the critical point.

In the analytical model the locus of the FRF in Eq. (13) intersects with the real axis at

$$\bar{\omega}_{r,0}^2 = 0.5\left(\left(\omega_r^2 + 4\xi_r\xi_0\omega_r\omega_0 + \omega_0^2\left(1+\frac{m}{M_r}\right)\right) \pm \sqrt{\left(\omega_r^2 + 4\xi_r\xi_0\omega_r\omega_0 + \omega_0^2\left(1+\frac{m}{M_r}\right)\right)^2 - 4\omega_0^2\omega_r^2}\right), \tag{14}$$

where $\bar{\omega}_0$ and $\bar{\omega}_r$ indicate the modified fundamental resonance frequency and the rth resonance frequency of the plate when the 2DOF model is considered. The amplitude of Eq. (13) at these intersection points is given by

$$\delta_{0,r} = \frac{\bar{\omega}_{0,r}^2}{\bar{\omega}_{0,r}^2\left(2\xi_0\omega_0\left(1+\frac{m}{M_r}\right)+2\xi_r\omega_r\right) - \omega_r^2 2\xi_0\omega_0 - \omega_0^2 2\xi_r\omega_r}. \tag{15}$$

In the special case where the plate damping is neglected, i.e., $\xi_r=0$, and the mass of the inertial actuator is much smaller than the modal mass of the plate $m/M_r \ll 1$, Eq. (15) can be simplified to the following expression:

$$\delta_0 = \frac{1}{2M_r\xi_0\omega_0\left(1-\frac{\omega_r^2}{\omega_0^2}\right)}. \tag{16}$$

The advantageous effect of internal damping is clearly observable in Eq. (16) since the gain margin increases with increasing ξ_0. Closed loop stability for external velocity feedback of V_s is guaranteed if $\delta_0>-1$ and thus

$$\omega_r^2 > \omega_0^2 + \frac{\omega_0}{M_r 2\xi_0}. \tag{17}$$

(Note that $\delta_0<0$ because of the fundamental actuator resonance.) Since ω_r and M_r are given by the structure to be controlled, the resonance frequency of the inertial actuator ω_0 should be chosen as small as possible. In particular the condition has to be fulfilled for the lowest resonance frequency of the plate ω_1. However, there are practical limitations[9] for the reduction of $\omega_0=\sqrt{K/m}$ since for example m should be kept as small as possible so that the inertial actuator is lighter than other passive means that could be used to increase the damping of the structure to be controlled. Also the K must be kept large enough in order to provide the necessary static support of the inertial mass. In conclusion the addition of internal damping ξ_0 is proposed in order to either increase the gain margin or performance at a given gain margin. In order to add internal damping, the implementation of an additional internal relative velocity feedback control loop in the actuator is investigated in Sec. III B.

Even when closed loop stability is achieved there will be spillover for frequencies where $|1+G_1V_s/F_a|<1$. For instance considering Eq. (13) for $r=1$, the cut-off frequency for spillover is found to be

$$\omega_{so} = \omega_0\sqrt{\frac{\left(\frac{\omega_1}{\omega_0}\right)^2 + \frac{\xi_1}{\xi_0}\frac{\omega_1}{\omega_0}}{\left(1+\frac{m}{M_1}\right) + \frac{\xi_1}{\xi_0}\frac{\omega_1}{\omega_0} + \frac{G_1}{4M_1\xi_0}}}. \tag{18}$$

Thus the higher is the frequency ratio ω_1/ω_0 the higher will be the spillover frequency. In general the spillover frequency can be further brought up by the internal damping effect although this effect tends to fade away as the ratio ω_1/ω_0 decreases. Because of limits in the design of low stiffness suspensions[9] small frequency ratios tend to occur in systems with lightweight, inertial actuators. For such systems, where additionally m is important relative to M_1, increased ξ_0 can

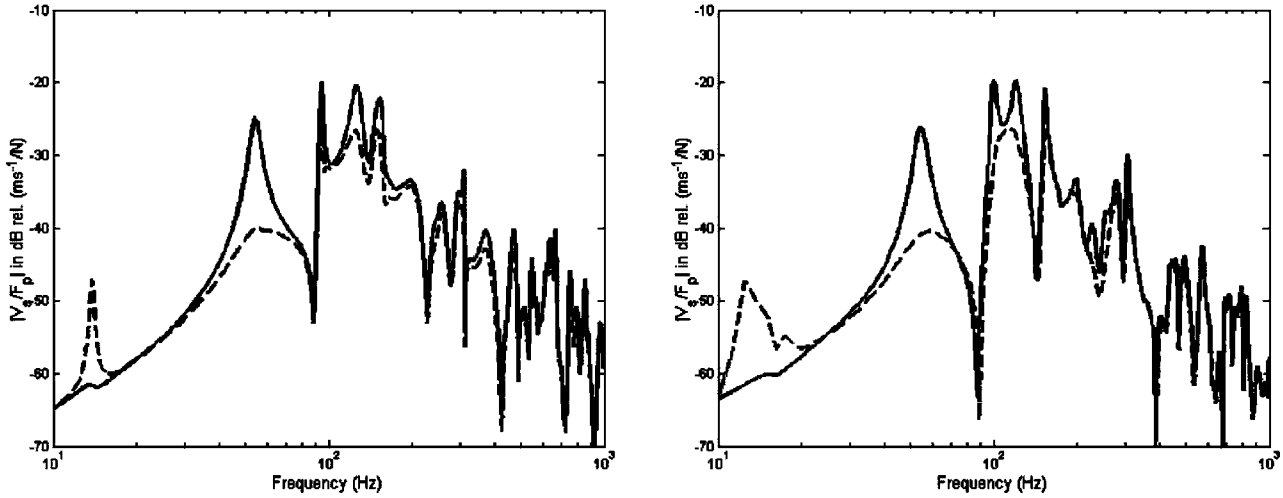


FIG. 6. Amplitude of the simulated (left) and measured (right) FRF between the primary disturbance force and the plate velocity at the actuator position in the open loop case (thick solid line) and for external velocity feedback (thick dashed line).

help to reduce ω_{so}. External velocity feedback also reduces the frequency range in which spillover occurs.

In conclusion, there will be an optimum internal velocity feedback gain for which, despite the negative spillover effects of the external velocity feedback at some frequencies, large vibration reductions are achieved at other frequencies so that an overall reduction results in the frequency band of interest. This notion of best internal velocity feedback gain will be studied in Sec. III C.

Figure 6 shows the amplitude of the simulated (left) and measured (right) response of the plate at the base of the actuator per unit primary force [Eq. (8)] in the open external loop case (solid line) and for external velocity feedback (dashed line) with the gain used to plot Fig. 5 and no additional internal damping. As expected, large spillover is visible at the fundamental resonance frequency of the inertial actuator that diminishes the average vibration reduction. Vibration reduction of about 16 dB at the first plate resonance and about 5 dB at higher resonances are measured and predicted by simulations. At higher frequencies there is no spillover, which indicates that higher order actuator resonance frequencies or non-ideal collocation of the actuation force and velocity sensing could be neglected at higher frequencies.

B. Internal velocity feedback (type b)

In this section internal velocity feedback is investigated as a means of adding internal damping to the inertial actuator, which has been found to improve stability properties of external velocity feedback control and reduce the spillover frequency under certain conditions. The stability of the internal velocity feedback is evaluated with the Nyquist plot of the simulated (left) and measured (right) FRF between the primary coil current and the secondary coil voltage shown in Fig. 7. The open loop is shown for high gains with a 240 Hz cutoff filter (solid line) and for medium gains with a 700 Hz cutoff filter (thick dashed-dotted line). The filters are used to cut-off the effect of the electromagnetic coupling between the primary and the secondary actuator coils, which would make the locus encircle the critical point $(-1,0j)$. Both measurements and simulations predict spillover in the closed loop case since for both gains the locus enters the circle with radius 1 about the critical point. For relatively high gains (thin solid line) larger number of loops due to plate resonances enter the negative real half-plane, which indicates higher control spillover effects. However, the large loop in the positive real half-plane indicates larger closed loop vibration reduction at the fundamental resonance of the actuator. Vibration increase due to spillover could be reduced by the simultaneous action of the proposed external velocity feedback if this spillover is within the bandwidth of the external sensor.

Figure 8 shows the Nyquist plot of the simulated (left) and measured (right) FRF between the primary coil current and the plate velocity at the control actuator position for high internal velocity feedback gains (thin solid line) and for medium internal velocity feedback gains (dashed-dotted line). Considering the measured FRF, the loops due to plate resonances do not significantly move into the left half-plane at high frequencies. In contrast, simulations show small spillover into the negative real half-plane because the simulated filter in the internal velocity feedback loop does not cut-off as steeply as the one used in experiments.

The beneficial effect of additional internal damping using internal relative velocity feedback is clearly visible when the plate is lightly damped since $|\delta_0|$ decreases at similar size of the loops due to plate resonances. Inversely at equal gain margin $1/|\delta_0|$, when medium internal feedback gains (dashed-dotted line) are implemented, the part of the loops due to the resonance frequencies of the plate in the positive real half-plane is bigger than in the purely external feedback case shown in Fig. 5. Since the further away the locus in the positive real half-plane is from the critical point the more vibration reduction is achieved,[5,18] larger vibration reduction is predicted at the rth resonance of the plate. For high internal relative velocity feedback (thin solid line) the size of the loops due to the resonance frequencies of the plate in the positive real half-plane decreases at equal gain margin. This type of behavior indicates the existence of an optimum con-

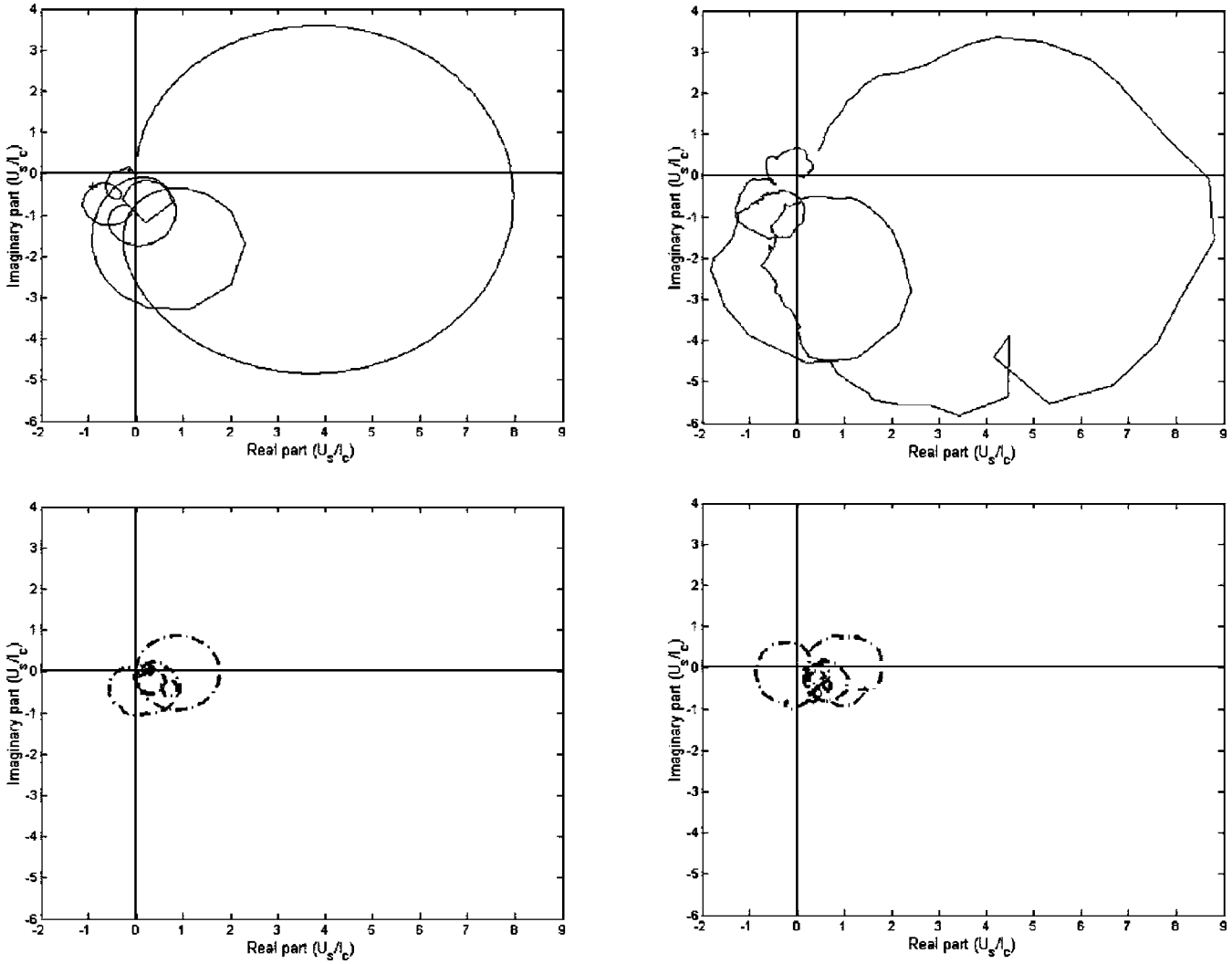


FIG. 7. Nyquist plot of the simulated (left) and measured (right) FRF between the primary coil current and the secondary coil voltage for a 240 Hz cutoff filter and high gain (solid line) and for a 700 Hz filter and medium gain (thick dashed line), for the internal feedback loop.

trol gain that can not be predicted by Eq. (16) because it neglects plate damping and the mass effect of the inertial actuator. When there is plate damping the vibration reduction generated by the external velocity feedback loop at the *r*th resonance of the plate is approximated by the amplitude of the FRF locus at the intersection point with the positive real axis δ_r which is indicated by δ_1 in Fig. 8 at the first plate mode. The smaller the ratio $|\delta_0|/\delta_r$ the larger the closed loop vibration reduction predicted at the *r*th plate mode for a given gain margin or the less spillover is predicted at an expected vibration attenuation. Since this criterion does not consider the smallest distance between the locus and the critical point, it also does not guarantee robustness.

Figure 9 shows the measured (dots) and simulated (lines) ratios $|\delta_0|/\delta_1$ in the case with coupling inductance between the primary and the secondary coils (solid line) and for ideal internal relative velocity feedback (dashed line). Both measurements (dots) and simulations (solid line) have a parabolic shape, which indicates that there is a best gain for which this ratio becomes minimum, i.e., maximum vibration reduction is achieved at the first plate resonance at a given gain margin $1/|\delta_0|$. As mentioned before the gain margin also determines the amount of spillover at the fundamental resonance frequency. Simulations only follow the general trend of measurements because they do not account for the distortion effect by the nonideal integrator in the charge amplifier. In the simulated case without coupling inductance between the primary and the secondary coils (dashed line), but with plate damping, this minimum is smaller; it lies at much higher gains and is much shallower. For increased vibration reduction hence the coupling inductance should be compensated.

C. Internal and external velocity feedback (type c)

Figure 10 shows the amplitude of the simulated (left) and measured (right) FRF between the primary disturbance force F_p and the plate velocity V_s in the open loop case (thick solid line), for high internal/medium external (thin solid line) and for medium internal/high external velocity feedback gains (thick dashed line).

Although for a relatively high internal feedback gain these plots show discrepancies between predictions and measurements, as also shown in Fig. 7, the trend for significant

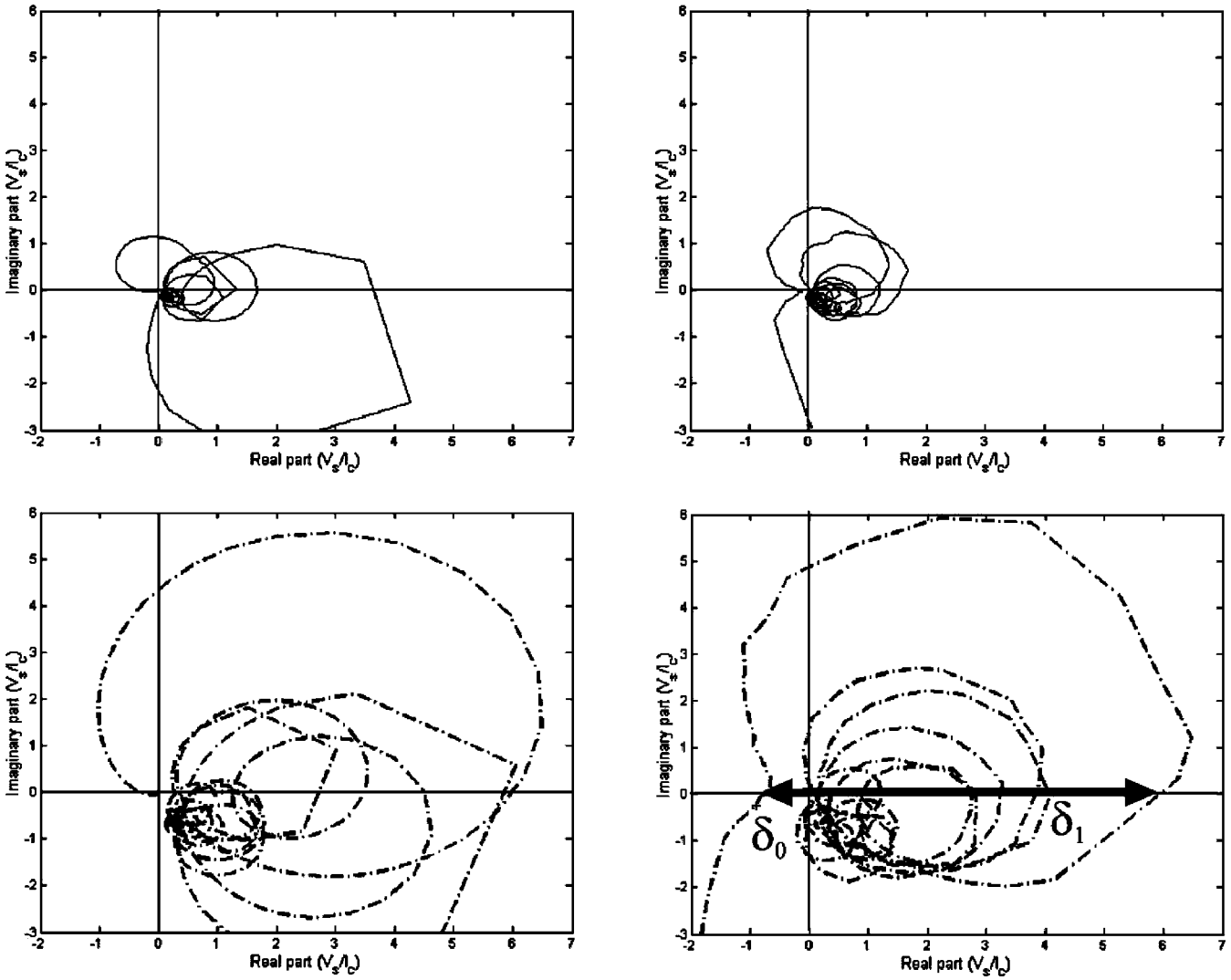


FIG. 8. Nyquist plot of the simulated (left) and measured (right) FRF between the primary coil current and the plate velocity at the secondary actuator position, for the external loop with high (thin solid line) and medium (thick dash dotted line) internal velocity feedback gains.

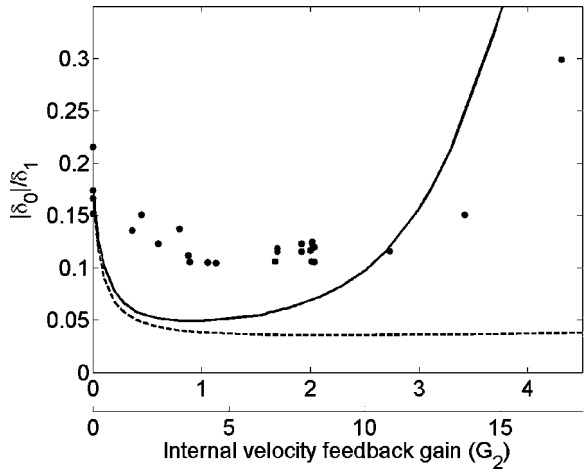


FIG. 9. Simulated (lines) and measured (dots) ratio between δ_1 and δ_2 in Fig. 7 for different internal velocity feedback gains with secondary coil voltage feedback (solid line and upper x-axis) and relative velocity feedback (dashed line and lower x-axis).

spillover at frequencies below the first plate resonance is clearly visible. This trend would even be enhanced when larger external velocity feedback gains were implemented. Larger vibration reductions of up to 21 dB and up to 10 dB at higher plate resonances are possible even with only a medium internal and a high external velocity feedback gain. Higher external velocity feedback gains cannot be implemented because of stability limits due to the non-ideal integrator of the charge amplifier. Also, in this case where $m < M_1$ and $\omega_1 > 4\omega_0$ increasing internal velocity feedback apparently does not reduce the frequency at which spillover ω_{so} occurs.

Figure 11 shows the simulated (left) and measured (right) vibration reduction at the first plate resonance for different internal and external velocity feedback gains. The measurements in Fig. 10 correspond to the highest and lowest vibration reduction implemented experimentally with simultaneous internal and maximum possible external velocity feedback (crosses). If no or very small external velocity feedback is used both in measurements (dots) and simulations (solid line 0.5) high internal velocity feedback gains have to be implemented to achieve about 12 dB vibration reduction.

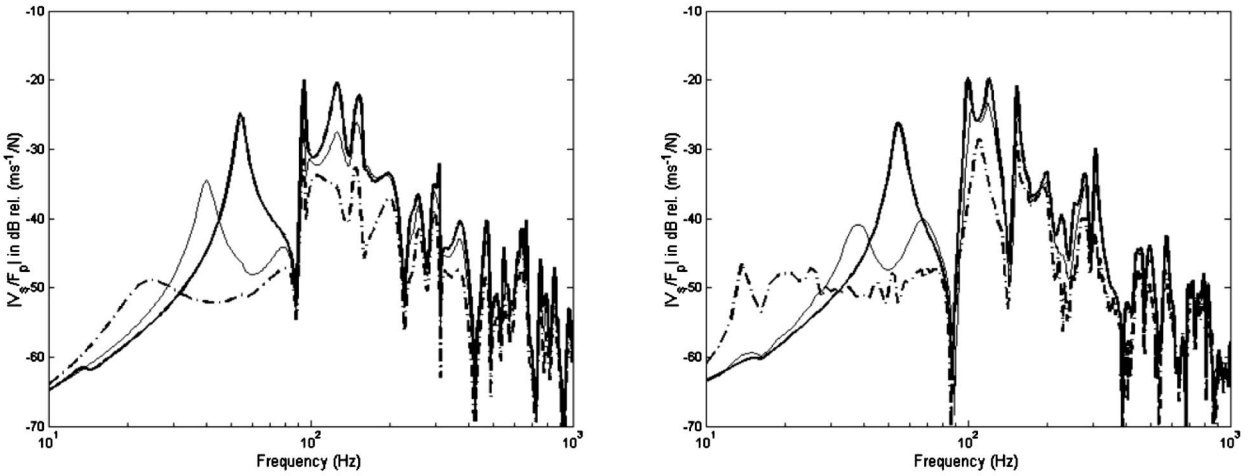


FIG. 10. Amplitude of the simulated (left) and measured (right) FRF between the disturbance force and the plate velocity at the secondary actuator position in the open loop case (thick solid line), internal high gain and external velocity feedback (thin solid line) and internal medium gain and external velocity feedback (thick dashed line).

For higher internal velocity feedback gains, vibration reduction levels off or is even reduced since the spillover frequency approaches the resonance frequency of the targeted first plate mode. If only external velocity feedback is implemented in measurements (cross on the y-axis) vibration reduction of about 16 dB is possible before the actuator becomes unstable. Internal actuator damping is increased by applying internal relative velocity feedback so that higher external velocity feedback gains can be implemented. If not enough internal damping is added at a given external velocity feedback gain, the closed loop will become unstable. If too much internal damping is added, following Eq. (18) if $\xi_0 \gg \xi_1\omega_1/\omega_0$ and $1+m/M_1+\xi_1/\xi_0\omega_1/\omega_0+G_1/4/M_1/\xi_0 \rightarrow 1$, spillover appears at frequencies close to the resonance frequency of the first plate mode so that vibration reduction decreases again. Hence, as indicated in Fig. 11, both in measurements and simulations there is a best combination of internal and external velocity feedback gains for which maximum vibration reduction is achieved when the first resonance of the plate is considered. Measurements only show the general trend of simulations because of the additional phase shift due to the nonideal integrator in the charge amplifier of the accelerometer that is used to sense the plate vibration. For the same reason the maximum external velocity feedback gain has not been implemented experimentally. Even for the high external velocity feedback gain measurements in Fig. 10 (dashed line) very low frequency vibrations are present that reduce measurement coherence due to stroke saturation.

IV. CONCLUSIONS

This study investigates the effect of internal velocity feedback on an inertial actuator used for active vibration control of a plate. A simplified two degrees of freedom model shows that internal velocity feedback increases the range of stable external velocity feedback gains, but in general it also increases the frequency range in which spillover occurs. Measurements and simulations of conventional velocity feedback using an external velocity sensor on the plate show that vibration reductions of about 16 dB at the first flexible plate mode and of about 5 dB at higher plate modes can be achieved. Internal velocity feedback adds internal damping to the actuator, although in practice the vibration reduction bandwidth is limited by the coupling effect between the primary and secondary coil used to measure the internal relative velocity. When combined internal and external velocity feedback is implemented, both measurements and simulations show vibration reduction of about 21 dB at the first flexible plate mode and about 10 dB at higher plate modes. Maximum reduction at a chosen resonance can only be achieved when an appropriate relative velocity feedback

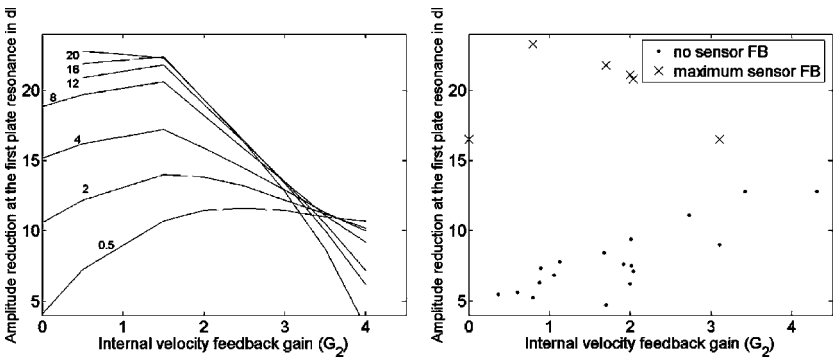


FIG. 11. Simulated (left) and measured (right) vibration reduction at the resonance frequency of the first plate resonance as function of the internal velocity feedback gain for different external velocity feedback gains.

TABLE II. Approximate vibration reduction in dB at different plate resonances.

External feedback gain	Internal feedback gain	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
medium	no	16	6	6	⋯	3	3	3	2	7	⋯	1	1
medium	high	16	4	3	5	3	5	2	3	5	⋯	1	1
high	medium	21	10	10	8	3	6	4	7	12	2	9	3

gain is implemented in order to stabilize the fundamental actuator resonance. This is explained by a characteristic modification of the open loop FRF for external velocity feedback due to internal velocity feedback. Vibration reduction performance is now limited by the sensor dynamics and, as expected, spillover into a lower frequency range is visible. Table II summarizes the approximate vibration reductions measured at the first twelve plate resonances in the case of (a) medium external and no internal velocity feedback, (b) medium external and high internal velocity feedback, and (c) high external and medium internal velocity feedback. The use of internal velocity feedback in an inertial actuator clearly improves the performance of collocated direct velocity feedback systems.

ACKNOWLEDGMENTS

The lightweight, electrodynamic, inertial actuator has been designed and built at the PMA department of the KU Leuven, Belgium in collaboration with Professor Paul Sas and Dr. Rene Boonen within the "European Doctorate in Sound and Vibration Studies" which is funded by the "Marie Curie Training Site" European Commission programme.

[1]P. Gardonio, "Review of active techniques for aerospace vibro-acoustice control," J. Aircr. **39**(2), 206–214 (2002).
[2]P. Gardonio, E. Bianchi, and S. J. Elliott, "Smart panel with multiple decentralized units for the control of sound transmission. Part I: theoretical predictions," J. Sound Vib. **274**(1–2), 163–192 (2004).
[3]B. Petitjean and I. Legrain, "Feedback controllers for active vibration suppression," Journal of Structural Control **3**(1–2), 111–127 (1996).
[4]B. Petitjean, I. Legrain, F. Simon, and S. Pauzin, "Active control experiments for acoustic radiation reduction of a sandwich panel: Feedback and feedforward investigations," J. Sound Vib. **252**(1), 19–36 (2002).
[5]A. Preumont, *Vibration Control of Active Structures* (Kluwer Academic, Dordrecht, 2002).
[6]C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration* (Academic, London, 1997).
[7]S. J. Elliott, P. Gardonio, T. C. Sors, and M. J. Brennan, "Active vibroacoustic control with multiple local feedback loops," J. Acoust. Soc. Am. **111**(2), 908–915 (2002).
[8]S. J. Elliott, M. Serrand, and P. Gardonio, "Feedback stability limits for active isolation systems with reactive and inertial actuators," Trans. ASME, J. Vib. Acoust. **123**, 250–261 (2001).
[9]C. Paulitsch, P. Gardonio, S. J. Elliott, P. Sas, and R. Boonen, "Design of a lightweight, electrodynamic, inertial actuator with integrated velocity sensor for active vibration control of a thin lightly-damped panel," Proceedings of ISMA2004, 2004, pp. 239–253.
[10]D. J. Inman, "Control/structure interaction: Effects of actuator dynamics," in *Mechanics and Control of Large Structures*, edited by J. L. Junkins (American Institute of Aeronautics and Astronautics, Reston 1990), Chap. 20, pp. 507–533.
[11]D. W. Miller and E. F. Crawley, "Theoretical and experimental investigation of space-realizable inertial actuation for passive and active structural control," J. Guid. Control **11**(5), 449–458 (1988).
[12]L. Benassi and S. J. Elliott, "Global control of a vibrating plate using a feedback-controlled inertial actuator," J. Sound Vib. **283**, 69–90 (2005).
[13]C. Paulitsch, P. Gardonio, and S. J. Elliott, "Active vibration damping using an electrodynamic actuator with internal velocity sensor," Proc. SPIE **5760**, 305–316 (2005).
[14]C. Paulitsch, P. Gardonio, and S. J. Elliott, "Internal velocity feedback for stabilisation of inertial actuators for active vibration control," in *IUTAM Symposium on Vibration of Nonlinear Mechanisms and Structures*, edited by H. Ulbrich and W. Gynther (Springer, Berlin, 2005) pp. 243–254
[15]P. Gardonio and M. J. Brennan, "Mobility and impedance methods in structural dynamics," in *Advanced Applications in Noise and Vibration*, edited by F. Fahy and J. Walker (E & FN Spon, London, 2004), pp. 389–447.
[16]U. S. Inan and A. S. Inan, *Engineering Electromagnetics* (Addison-Wesley, Menlo Park, 1998).
[17]C. Paulitsch, P. Gardonio, and S. J. Elliott, "Active vibration damping using a self-sensing electrodynamic actuator," Proc. SPIE **5386**, 282–293 (2004).
[18]L. Meirovitch, *Dynamics and Control of Structures* (Wiley, New York, 1989).

Near-irreversibility in a conservative linear structure with singularity points in its modal density

A. Carcaterra[a)]
Department of Mechanics and Aeronautics, University of Rome, "La Sapienza" Via Eudossiana, 18, 00184, Rome, Italy

A. Akay and I. M. Koç
Department of Mechanical Engineering, Carnegie Mellon University Pittsburgh, Pennsylvania 15213

(Received 13 June 2005; revised 3 February 2006; accepted 3 February 2006)

Through two complementary approaches, using modal response and wave propagation, the analyses presented here show the conditions under which a decaying impulse response, or a nearly irreversible energy trapping, takes place in a linear conservative continuous system. The results show that the basic foundation of near-irreversibility or apparent damping rests upon the presence of singularity points in the modal density of dynamic systems or, analogously, in the wave-stopping properties associated with these singularities. To illustrate the concept of apparent damping in detail, a simple undamped beam is modified to introduce a singularity point in its modal density distribution. Simulations show that a suitable application of a compressive axial force to an undamped beam placed on an elastic foundation attenuates its impulse response with time and develops the characteristics of a nearly irreversible energy trap. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2179747]

PACS number(s): 43.40.At, 43.40.Kd, 43.40.Cw, 4340.Jc [RLW] Pages: 2141–2149

I. NEARLY IRREVERSIBLE ENERGY TRAPPING IN UNDAMPED CONTINUOUS FINITE STRUCTURES

Irreversible energy processes in conservative systems that consist of multiple coupled resonators are generally ascribed to nonlinear interactions through which ordered energy at a macroscopic scale converts into disordered energy at a microscopic scale. A prominent example of such a case is the anharmonic vibrations of atoms that describe dissipation in solids. Although significant efforts by the scientific community continue to advance our understanding of irreversibility and oscillations at the atomic level, the underlying principles that relate irreversibility of a large population of oscillators and nonlinearity of their mutual interaction is a reasonably well-established concept.[1]

Generally, irreversibility is not thought to be intrinsic to the nature of conservative linear systems. Typically their response exhibits the recurrence of energy transfer: energy injected by an external source into the system is cyclically absorbed and released back to the source.[2,3] Although recurrence is a common characteristic of undamped linear (and nonlinear) systems, recent analyses show how a nearly irreversible energy transfer from a master structure to an attached set of oscillators can be obtained by an optimal selection of the natural frequencies within the set.[4,5]

This paper brings a new perspective to the phenomenon of apparent damping[6–12] manifested in the response of undamped continuous linear structures. The general relationship between the concept of irreversibility, apparent damping, and the natural frequency distribution of a vibrating continuous structure is investigated and examined for the special case of a beam. In the next section we present a brief discussion about the concept and meaning of near-irreversibility in the context of continuous systems considering two complementary approaches: a modal interpretation and wave propagation, respectively.

II. NEAR-IRREVERSIBILITY IN CONTINUOUS SYSTEMS

A. Singularity in the modal density

To illustrate the concept of near-irreversibility in the context of modal analysis, consider a one-dimensional continuous linear undamped structure S subject to an impulse at point x_0. The governing equation,

$$\Lambda[w(x,t)] + m' \frac{\partial^2 w(x,t)}{\partial t^2} = 0,$$

also satisfies the initial conditions $w(x,0)=0$, $\dot{w}(x,0) = V_0 L \delta(x-x_0)$, where V_0 is the initial velocity at $x=x_0$, L is the length of the structure, δ is the Dirac's distribution, $\Lambda[\ ]$, and $w(x,t)$, m' represent the structural operator, the displacement, and the mass density per unit length, respectively.

Representing with $\phi_i(x)$ and $q_i(t)$ the orthonormal modes and the principal coordinates of S, respectively, the vibration field is described by

$$w(x,t) = \sum_{i=1}^{M} \phi_i(x) q_i(t)$$

[a)]Corresponding author: Antonio Carcaterra, Dipartimento di Meccanica ed Aeronautica, Università di Roma, "La Sapienza," Via Eudossiana, 18, 00184, Roma. Electronic mail: a.carcaterra@dma.ing.uniroma1.it

$$q_i(t) = A_i \sin \omega_i t, \quad A_i = \frac{1}{\omega_i}\int_S m' \dot{w}_0 \phi_i \, dS = \frac{m'}{\omega_i} V_0 L \varphi_i(x_0),$$

where dS is the volume element of S. Thus, the impulse response $h(t)$ of the structure at x_0 becomes

$$h(t) = \sum_{i=1}^{M} G(\omega_i) \sin \omega_i t, \tag{1}$$

where $G(\omega_i) = (m'/\omega_i) V_0 L \phi_i^2(x0)$. In the absence of damping, the impulse response of such a structure consists of a combination of sine functions, and exhibits an almost-periodic behavior. An asymptotically decaying trend of $h(t)$ is normally expected only in the presence of energy dissipation or in rare cases of nonlinearities and is intrinsically related to and describes irreversible energy transfer.

However, under particular conditions, generalized from results given in Ref. 3, a decaying impulse response can be observed, even in a conservative linear system, which we refer here as apparent damping with near-irreversibility. Because energy does not dissipate in the classical sense but is confined or trapped spatially, it appears like damping. For the time scales used in the computations presented in this paper, recurrence is not observed, which would have attenuated in the presence of any amount of dissipation in the system, thus making the energy transfer nearly irreversible.

Characteristic properties of $h(t)$ in Eq. (1) follow those of its continuous counterpart:

$$h_{\text{int}}(t) = \int_0^1 G[\omega(\xi)] \sin \omega(\xi) t \, d\xi,$$

where ξ is a dummy variable. As shown in Ref. 3, $h_{\text{int}}(t)$ has the following asymptotic properties:

$$\text{If } \frac{d\omega(\xi)}{d\xi} \neq 0, \quad \forall\, \xi \in [0,1],$$

$$\text{then } h_{\text{int}} \propto \left(\frac{1}{t}\right) \to 0, \quad \text{for } t \to \infty, \tag{P1}$$

$$\text{If } \frac{d\omega}{d\xi}(\xi_0) = 0, \quad \xi_0 \in [0,1],$$

$$\text{then } h_{\text{int}} \propto \left(\frac{1}{t^{1/2}}\right) \to 0, \quad \text{for } t \to \infty. \tag{P2}$$

If the local property (P2) is extended to all points $\xi \in [0,1]$, then all frequencies in that interval are equal and $h_{\text{int}}(t)$ becomes a simple sine function and does not decay in time.

Since $h(t)$ is a discrete approximation of $h_{\text{int}}(t)$, their difference is given by the remainder term:[3]

$$\int_0^1 G[\omega(\xi)] \sin \omega(\xi) t \, d\xi = \sum_{i=1}^{M} G(\omega_i) \sin \omega_i t + \Re_M(t),$$

$$\Re_M(t) = \frac{1}{2}\Delta\xi \left.\frac{d}{d\xi}\{G[\omega(\xi)] \sin \omega(\xi) t\}\right|_{\xi=\xi_M(t)},$$

where $\Delta\xi = 1/M$ and $\xi_M(t) \in [0,1]$ depends on the number of modes M used in the modal expansion. Then

$$\Re_M(t) = h_{\text{int}}(t) - h(t) = \left.\frac{1}{2M}\left(\frac{dG}{d\omega}\frac{d\omega}{d\xi} \sin \omega t + Gt\frac{d\omega}{d\xi} \cos \omega t\right)\right|_{\xi=\xi_M(t)}$$

and

$$|\Re_M(t)| = |h_{\text{int}}(t) - h(t)| \leqslant \frac{1}{2M} \max\left(\left|\frac{d\omega}{d\xi}\right|\right)\left[\left.\left(\frac{dG}{d\omega}\right)^2\right|_{\xi=\xi_M(t)} + \left. t^2 G^2\right|_{\xi=\xi_M(t)}\right]^{1/2}.$$

The remainder in this form has a third property:

$$|\Re_M(t)| \propto \max\left(\left|\frac{d\omega}{d\xi}\right|\right), \quad \text{for } \xi \in [0,1]. \tag{P3}$$

Properties P1, P2, and P3 point to frequency distributions that lead to a decaying impulse response $h(t)$, which forms the basis of the phenomenon of near-irreversibility or apparent damping. Properties P1 and P2 show that $h_{\text{int}}(t)$ vanishes for any monotonically increasing frequency distribution or for distributions with one or more stationary points ξ_0. Property P3 proposes a criterion that brings $h(t)$ close enough to $h_{\text{int}}(t)$ to have the same properties: $\max(|d\omega/d\xi|)$ must be small within the interval $\xi \in [0,1]$. It appears that those natural frequency distributions with one or more stationary points within the interval $\xi \in [0,1]$ are the likely candidates to produce a nearly vanishing impulse response $h(t)$. In fact, in such cases $h_{\text{int}}(t)$ vanishes for large times. In addition, since around any stationary point the function $|d\omega/d\xi|$ is small, the value $\max(|d\omega/d\xi|)$ in a distribution with multiple stationary points tends to be smaller. Thus, having multiple stationary points implies that $|h_{\text{int}}(t) - h(t)|$ is small, and $h(t)$ exhibits asymptotically decaying properties close to those of $h_{\text{int}}(t)$.

The differential change $d\omega/d\xi$ relates to the modal density $\delta(\omega)$ of the structure. Considering $d\xi = (1/M)dM$, with M the total number of modes for $\xi \in [0,1]$, and dM the number of modes in the interval $[\xi, \xi + d\xi]$, $d\omega/d\xi \propto d\omega/dM = 1/\delta(\omega)$. Thus, any stationary point ξ_0 in the natural frequency distribution corresponds to a singularity point $\omega_0 = \omega(\xi_0)$ in the structural modal density:

$$\frac{d\omega}{d\xi}(\xi_0) = 0; \quad \text{then } \delta(\omega_0) \to \infty. \tag{2}$$

Consequently, singularity in modal density leads to a decaying behavior of the impulse response function, even in the absence of damping in linear systems.

B. Vanishing group velocity

The physics of the same behavior can also be described in the context of wave propagation: In a one-dimensional undamped waveguide that is initially at rest, transient energy introduced by an impulse force propagates away from the excitation point toward the boundaries of the structure. As a consequence, the response of the waveguide at the excitation point initially decreases as the energy travels away, in the

structure. Once the injected energy reaches the boundaries and reflects, it travels back toward the excitation point. Since the waveguide is undamped, energy reflects without losses, and the response of the waveguide at the excitation point starts to build up. This example is typical of the reversibility observed in undamped linear systems, where, in general, energy continuously redistributes itself spatially and does not stay at location permanently.

Normally, the return of energy to the excitation point does not take place if either the medium is infinitely large or possesses large damping. Just as an infinitely long waveguide does not have energy returning from its boundaries, effectively leading to an irreversible energy transfer, the same is true for a finite waveguide that has large enough damping, which can dissipate the radiated energy before it returns to the excitation point. In this section we address the conditions for producing a nearly irreversible energy transfer in a finite, undamped structure using a different approach, as discussed above.

A wave group with frequencies around ω_0 has corresponding wave numbers around $k(\omega_0)$, in accordance with the dispersion relationship $k=k(\omega)$ of the structure. The energy of the wave group radiated from the excitation point into the waveguide propagates with the group velocity $c_g(\omega)=d\omega/dk$, evaluated at the frequency ω_0. The time it takes for this energy to reflect back from the boundaries essentially depends on its group velocity. The set of natural frequencies of the structure, ω_n, have corresponding wave numbers $k_n=2\pi/\lambda_n$, where λ_n are the natural wavelengths. In general, at high frequencies, $\lambda_n \propto L/n$ and $k_n \propto n$, and thus $dk \propto dn \propto d\xi$. It follows that the group velocity depends on the distribution of natural frequencies as before $c_g(\omega)=d\omega/dk \propto d\omega/d\xi$ and

$$\frac{d\omega}{d\xi}(\xi_0)=0 \rightarrow \quad c_g(\omega_0)=\left.\frac{d\omega}{dk}\right|_{\omega=\omega_0}=0. \tag{3}$$

Thus, at a condensation point ξ_0, group velocity vanishes and propagation of the group of waves around ω_0 is inhibited, with a consequent attenuation of reflection from the boundaries, virtually preventing a complete energy return. Thus, the presence of a condensation point in the frequency distribution is to filter part of the impulse energy by a stop-band effect. The group velocity around ω_0 does not propagate toward the boundaries but remains localized near the excitation point. As a result its energy remains stored in the structure, without contributing to the reflected energy that returns to the excitation point. This filtering effect is broadened to include multiple frequency regimes if more than one stationary point appears in the natural frequency distribution of the structure.

As a result of the filtering effect, the response of the structure decreases at the excitation point at early times, followed by an almost constant amplitude vibration for late times, exhibiting an apparently damped behavior at the excitation point.

Apparent damping, or irreversible radiation of energy from the excitation point around ω_0, manifests itself as a decaying impulse response. Consequently, the presence of condensation points in the natural frequency distribution presents a strategy for producing nearly irreversible energy trapping in undamped continuous finite structures.

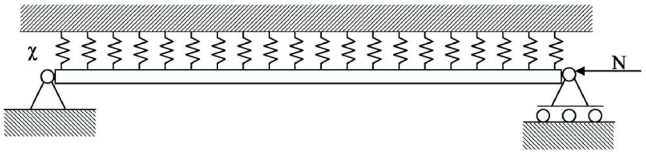


FIG. 1. Sketch of the modified beam.

In the next section this strategy is applied to the case of a simply supported beam by suitably modifying it to match the requirements given by Eqs. (2) and (3).

III. AN APPLICATION: SIMPLY SUPPORTED BEAM

Since the natural frequencies ω_n of a simply supported beam are proportional to n^2, their distribution does not possess any condensation points and its modal density decreases with increasing frequency. Accordingly, on the basis of the arguments given in the previous section, such a beam cannot inhibit propagation of a wave group and reduce its reflection from boundaries.

However, as shown in the following sections, a singularity can be introduced to the modal density of a simply-supported beam, by placing it on an elastic foundation and by applying to it a constant axial compressive force. Such a modification introduces the capability of producing a nearly irreversible energy process.

A. Condensation of natural frequencies: Modal approach

The motion of a Bernoulli beam suspended on an elastic foundation, characterized by a stiffness, χ, per unit length and compressed by an axial constant force N (see Fig. 1) is described by

$$B\frac{\partial^4 w(x,t)}{\partial x^4}+N\frac{\partial^2 w(x,t)}{\partial x^2}+\chi w(x,t)+\rho A\frac{\partial^2 w(x,t)}{\partial t^2}=0,$$

where $w(x,t), B, \rho, A$ represent, respectively, the transverse displacement, the bending stiffness, the material mass density, and the area of the cross section of the beam. The normalized modes for a beam of length L hinged at each end are

$$\Phi_n(x)=\sqrt{\frac{2}{\rho AL}}\sin k_n x, \quad k_n=\frac{\pi n}{L}, \quad n=1,2,3,\ldots .$$

The corresponding natural frequencies,

$$\omega_n=\frac{1}{\sqrt{\rho A}}\sqrt{Bk_n^4-Nk_n^2+\chi}, \tag{4}$$

reduce to that of a simply supported beam for $N=0$, $\chi=0$, with a frequency distribution $\omega_n=(\pi n/L)^2\sqrt{B/\rho A}$ that has a parabolic trend and a corresponding modal density $\delta=(1/2n)\sqrt{\rho A/B}(L/\pi)^2$ that monotonically decreases with the mode number.

The singularity in the modal density of the beam can be obtained from the condensation requirement in Eq. (2) as

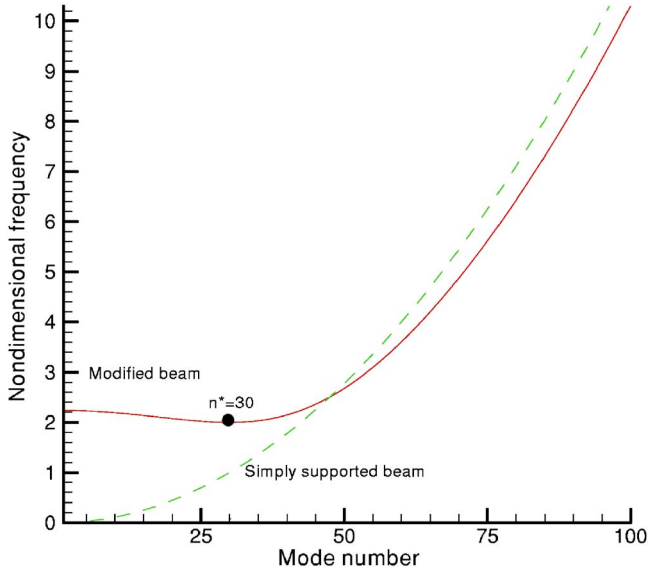


FIG. 2. (Color online) Frequency distribution: simply supported beam versus modified beam.

$$\delta_{n\to n^*} = \left.\frac{1}{d\omega_n/dn}\right|_{n\to n^*} = \infty,$$

$$\frac{d\omega_n}{dn} = \frac{1}{\sqrt{\rho A}}\frac{2Bk_n^3 - Nk_n}{\sqrt{Bk_n^4 - Nk_n^2 + \chi}}.$$

Thus, the singularity point corresponding to a particular mode $n=n^*$ is found from

$$2Bk_{n^*}^2 = N, \tag{5}$$

where $k_{n^*}=\pi n^*/L$. The substitution for N in Eq. (4) produces the natural frequency distribution of the beam that now has a singularity in its modal density:

$$\omega_n = \frac{1}{\sqrt{\rho A}}\sqrt{Bk_n^4 - 2Bk_{n^*}^2k_n^2 + \chi}. \tag{6}$$

Introducing as a reference frequency the corresponding n^*th natural frequency of the simply supported beam (with $N=0$, $\chi=0$), $\omega_{n^*}=k_{n^*}^2\sqrt{B/\rho A}$, a convenient nondimensional form of the frequency distribution in Eq. (6) results as

$$\Omega_n = \frac{\omega_n}{\omega_{n^*}} = \sqrt{\left(\frac{n}{n^*}\right)^4 - 2\left(\frac{n}{n^*}\right)^2 + \sigma^2}, \tag{7}$$

$$\sigma^2 = \frac{\chi}{\rho A\omega_{n^*}^2}, \quad n = 1,2,3,\ldots .$$

As an illustration, the distribution (7) (for values of $n^*=30$, $\sigma^2=5$), compared with that of the simply supported beam in Fig. 2 shows how the frequency distribution of the modified beam now possesses a condensation point at $n=n^*$, where the curve $\omega_n(n)$ exhibits a relative minimum. In addition, the same distribution would present a second condensation point at $n=0$, that, however, is not an allowed value for n in Eq. (7) (minimum is $n=1$).

Note that the frequencies that correspond to the branch of the curve (for the modified beam) within the band bounded by the points $n=1$ and $n=n^*$, fall approximately within the interval $[\sqrt{\sigma^2-1},\sigma]$, leaving the distribution between the two points $n=1$ and $n=n^*$ with a nearly constant value within the band $[1,n^*]$. Accordingly, the modified frequency distribution has a flat region within $[1,n^*]$ that satisfies the requirements that lead to near-irreversibility or apparent damping, as described in Sec. II.

An additional requirement relates to potential buckling of the beam under the axial load N. In order to avoid buckling of the beam, the parameters n^* and σ upon which the frequency distribution (7) depends, must also satisfy the constraint (see the Appendix)

$$n^* < \frac{L}{\pi}\sqrt[4]{\frac{\chi}{B}}, \quad \sigma > 1.$$

As an example, consider the free motion of the beam excited by an initial velocity impulse at $x=x_0$ is represented by

$$w(x,t) = \sum_{i=1}^{\infty}\Phi_i(x)(\alpha_i \sin\omega_i t + \beta_i\cos\omega_i t),$$

with $\alpha_i=1/\omega_i\Phi_i(x_0)V_0m$, $\beta_i=0$ obtained by imposing the initial conditions

$$w(x,0) = 0, \quad \dot{w}(x,0) = V_0L\delta(x-x_0), \tag{8}$$

where V_0 represents the initial velocity at $x=x_0$ and m the mass of the beam. The response can be expressed in nondimensional form as

$$W(X,T) = \sum_{n=1}^{\infty}\frac{1}{\Omega_n}\sin(\pi nX)\sin(\pi nX_0)\sin 2\pi\Omega_nT, \tag{9}$$

with $X=x/L$, $T=t\omega^*/2\pi$, $W=w\omega^*/2V_0$.

The drive point frequency response function at X_0 follows as

$$H(\Omega) = \sum_{n=1}^{\infty}\frac{\sin^2(\pi nX_0)}{\Omega_n^2-\Omega^2}. \tag{10}$$

The response of the simply supported beam at $X=X_0=1/2$ and shown in Fig. 3 includes 140 modes (and $n^*=1$), and illustrates how such a beam exhibits a reversible response and maintains a constant-amplitude displacement.

The corresponding response for the modified beam ($n^*=80$, $\sigma^2=10$) shown in Figs. 4 (short time scale) and 5 (long time range) clearly show the development of apparent damping. In this case, the response very rapidly decreases to a level much lower than its initial value and remains there. The response does not gain its initial value, even for very long-time simulations.

To compare with the point excitation examined above, the following example considers the effects of a spatially distributed velocity perturbation:

$$w_0(x,0) = 0, \quad \dot{w}_0(x,0) = V_0H(x-x_0),$$

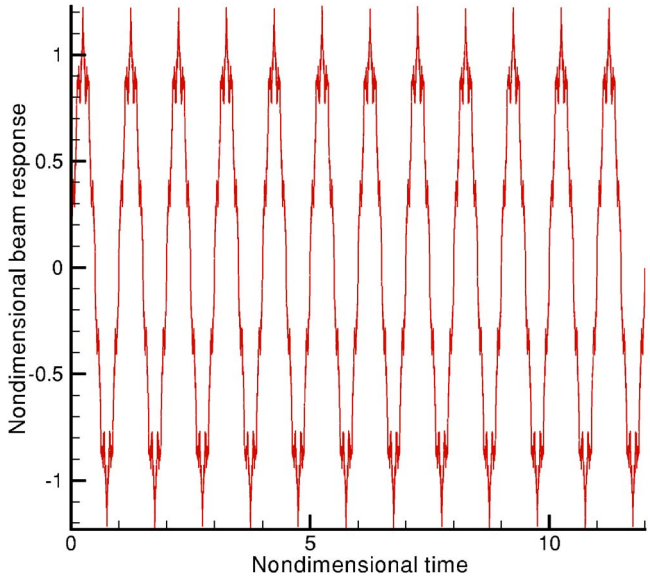


FIG. 3. (Color online) Response to an impulse of the simply supported beam.

$$H(x) = \begin{cases} 1 - \cos\dfrac{2\pi}{\lambda}(x + \lambda/2), & \text{for } |x| \leqslant \dfrac{\lambda}{2}, \\ 0, \text{ elsewhere}, \end{cases}$$

where H is the Hanning window, and the initial velocity is distributed within a region of length λ centered around x_0 with maximum value $2V_0$ at x_0. The corresponding beam response has the form

$$W(X,T) = \sum_{n=1}^{\infty} \frac{Q_n}{\Omega_n} \sin(\pi n X) \sin 2\pi\Omega_n T,$$

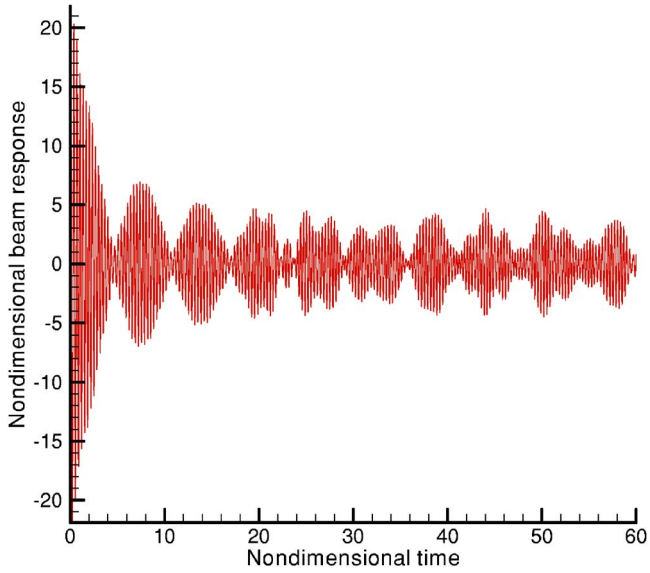


FIG. 4. (Color online) Response to an impulse of the modified beam: short time scale.

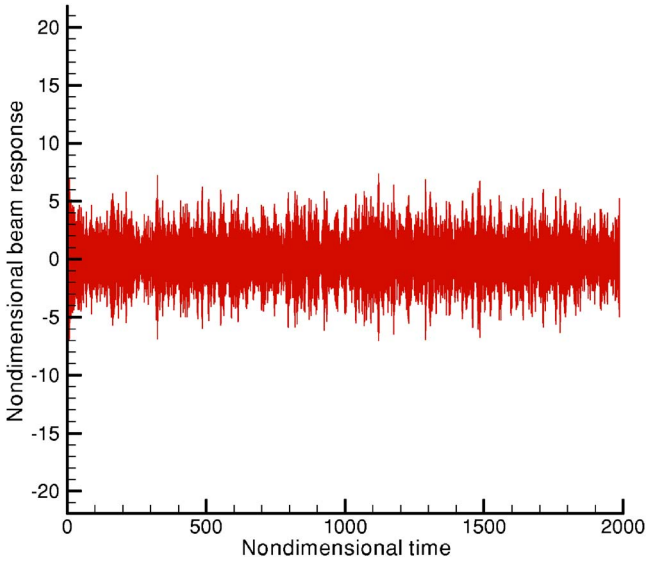


FIG. 5. (Color online) Response to an impulse of the modified beam: long time range.

$$Q_n = \frac{1}{L}\int_{x_0-\lambda/2}^{x_0+\lambda/2} H(x - x_0)\sin k_n x \, dx.$$

Figures 6–9 show the response of the same modified beam for $\lambda/L=0.01$, $\lambda/L=0.05$, $\lambda/L=0.1$, $\lambda/L=0.25$, respectively. These results suggest that the apparent-damping effect depends on the spatial distribution of the excitation and is stronger for spatially small distributions, $\lambda/L \ll 1$, such that the initial condition excites a wider spectrum of wave numbers, thus involving the participation of a larger number of modes in the beam response.

Figures 10 and 11 compare the frequency response functions of the simply supported beam and of the modified beam calculated by expressions (7) and (10), where $n^*=30$, $\sigma^2=5$. The modifications through the axial force and the stiffness of the platform increases the natural frequencies in the spectrum below the condensation point ω_{n^*} and compresses them

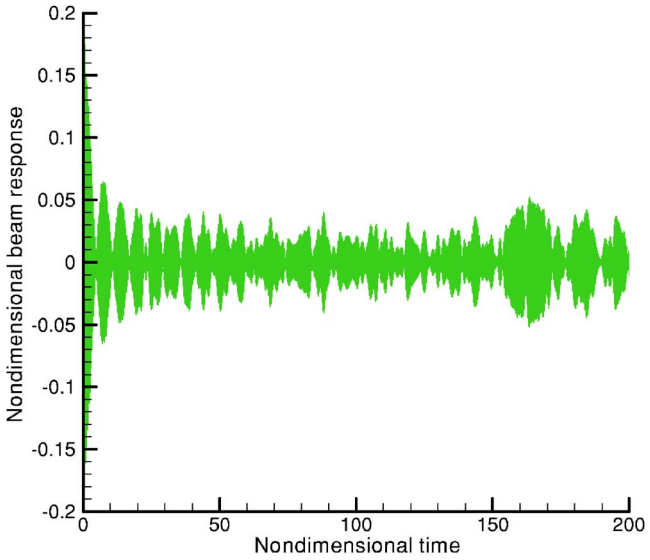


FIG. 6. (Color online) Response to an impulse of the modified beam: $\lambda/L=0.01$.

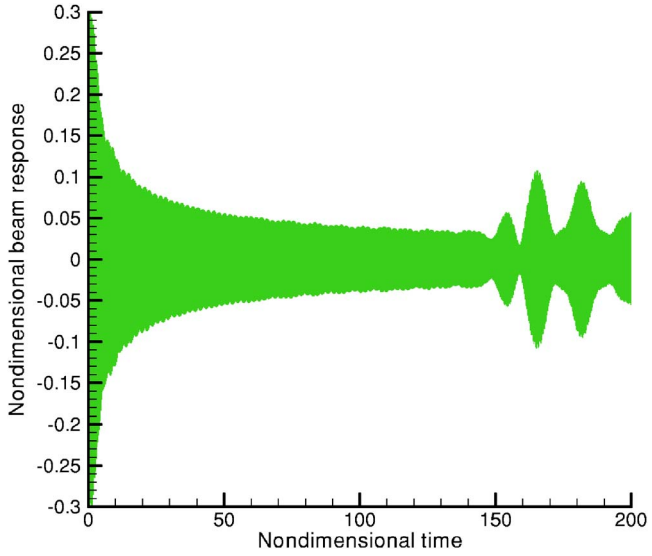


FIG. 7. (Color online) Response to an impulse of the modified beam: $\lambda/L=0.05$.

around it, leaving the lower part of the spectrum devoid of resonances. As discussed in the next section, an inspection of the dispersion relationship of the modified beam suggests that this frequency range is characterized by a cutoff phenomenon.

B. Inhibition of group wave propagation

In this section we examine the implications of Eq. (3) that produces the stop-band effect in energy propagation along the beam. The dispersion relationship of the modified beam is expressed as

$$Bk^4 - Nk^2 + \chi - \rho A\omega^2 = 0, \quad \omega = \frac{1}{\sqrt{\rho A}}\sqrt{Bk^4 - Nk^2 + \chi}.$$

The corresponding group velocity c_g along the beam follows as

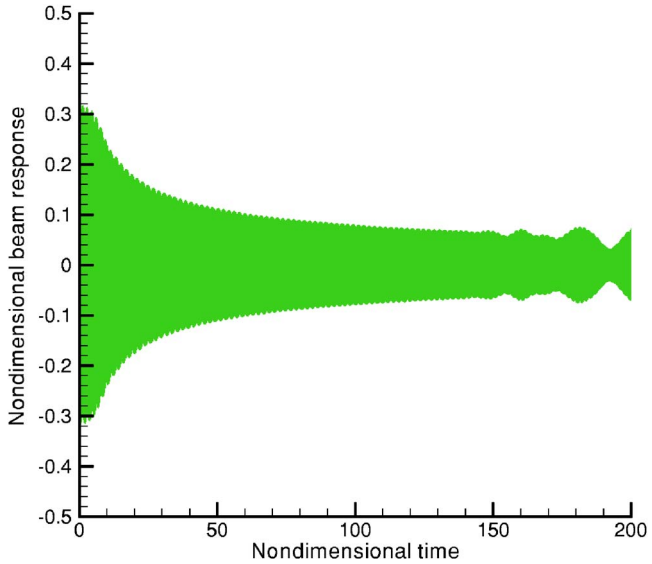


FIG. 8. (Color online) Response to an impulse of the modified beam: $\lambda/L=0.1$.

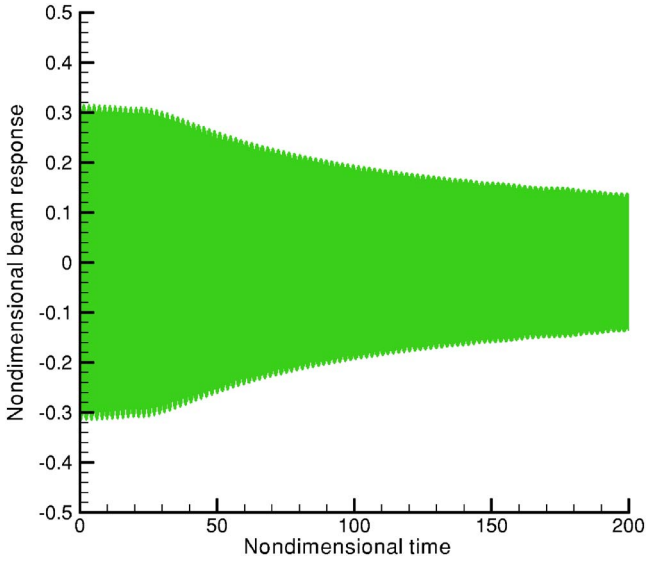


FIG. 9. (Color online) Response to an impulse of the modified beam: $\lambda/L=0.25$.

$$c_g(\omega) = \frac{d\omega}{dk} = \frac{1}{\sqrt{\rho A}}\frac{2Bk^3 - Nk}{\sqrt{Bk^4 - Nk^2 + \chi}}.$$

Following Eq. (3), requiring that the group velocity of the beam at a frequency ω^* be zero, implies that

$$c_g(\omega) = \left.\frac{d\omega}{dk}\right|_{k_{n^*}} = \frac{1}{\sqrt{\rho A}}\frac{2Bk_{n^*}^3 - Nk_{n^*}}{\sqrt{Bk_{n^*}^4 - Nk_{n^*}^2 + \chi}} = 0,$$

$$\rightarrow 2Bk_{n^*}^2 = N. \tag{11}$$

The resulting condition coincides with that in Eq. (5) and implies that energy associated with a group of waves having wave numbers around k_{n^*}, or frequencies close to ω_{n^*}, cannot propagate along the beam.

Following an initial impulse applied the beam, all its modes respond through which energy is introduced to the

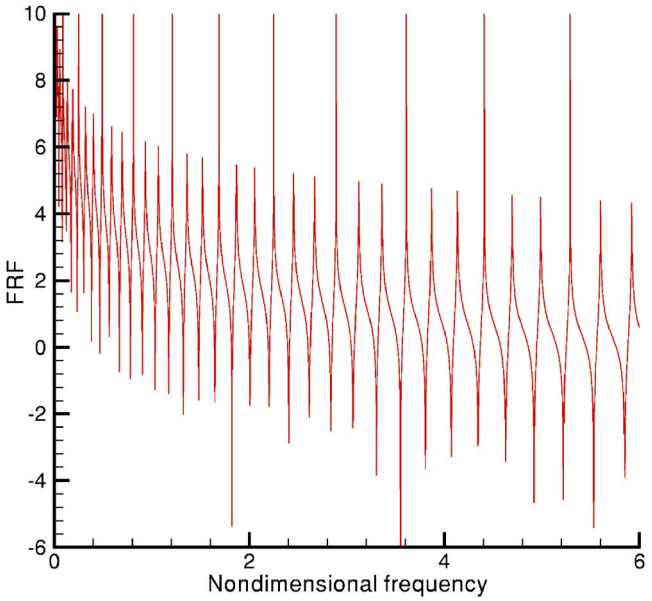


FIG. 10. (Color online) FRF of the simply supported beam.

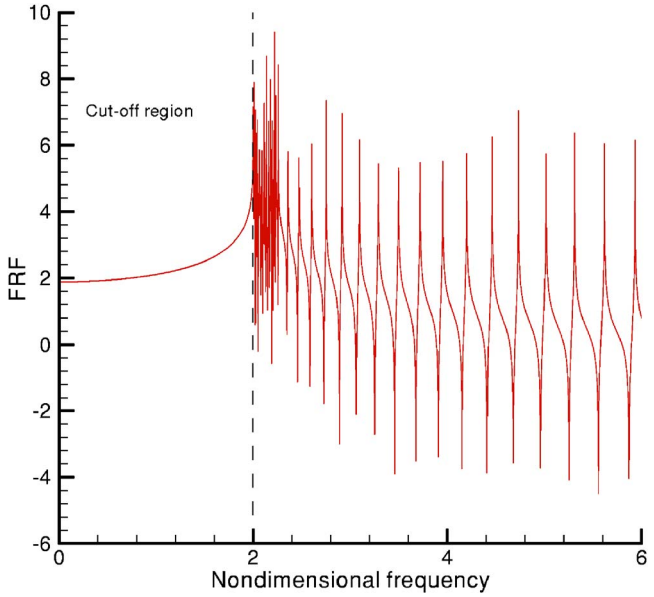


FIG. 11. (Color online) FRF of the modified beam.

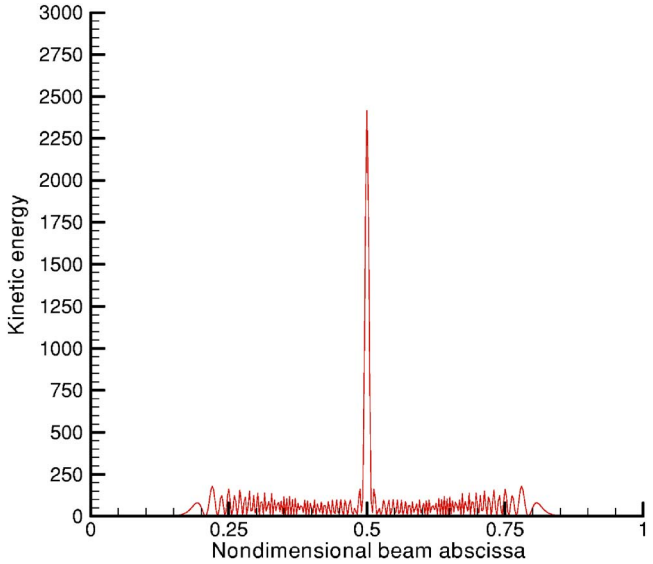


FIG. 13. (Color online) Kinetic energy distribution along the beam $T=1.25$.

beam at the excitation point, including those modes near ω^*. Because $c_g = d\omega/dk|_{k_{n^*}}=0$ in the neighborhood of ω^*, the propagation of the energy associated with these modes away from the excitation is prevented, producing an energy localization that leads to the near-irreversibility, as described in Sec. II.

Figures 12–17 show the nondimensional kinetic energy distribution along the beam at different times, illustrating the development of energy localization produced by the stop band effect. The beam response is obtained using Eq. (9) with 400 modal terms in the series and the frequency distribution (7) with $n^*=80$, $\sigma^2=10$. At early times (Fig. 12, $T=0.5$ and Fig. 13, $T=1.25$), the peak at the excitation point ($X=0.5$) is still observable. The faster energy components depart from the excitation point and reach the boundaries (Fig. 14, $T=2$) and then start to reflect (Fig. 15, $T=2.5$). However, these faster energy components have negligible amplitudes compared with the energy at the excitation point, which does not travel along the beam, but remains localized near the excitation point with a speed that approaches zero (Fig. 16, $T=3.5$ and Fig. 17, $T=5$).

IV. CONCLUDING REMARKS

Irreversibility characterizes systems that convert mechanical energy into thermal energy. This conversion phenomenon is responsible for the decay of the impulse response of any structure that has dissipation. In this paper we show that a similar response, called apparent damping or near-irreversibility can be achieved in a linear conservative system when its modal density has singularity points.

The two complementary approaches given support this conclusion. The first is a modal-based consideration, which shows that the presence of singularity points in the modal density produces an impulse response with a sharply decay-

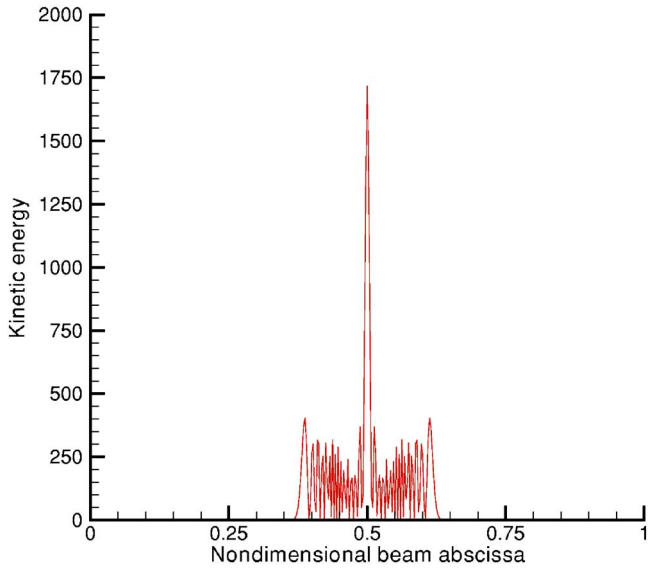


FIG. 12. (Color online) Kinetic energy distribution along the beam $T=0.5$.

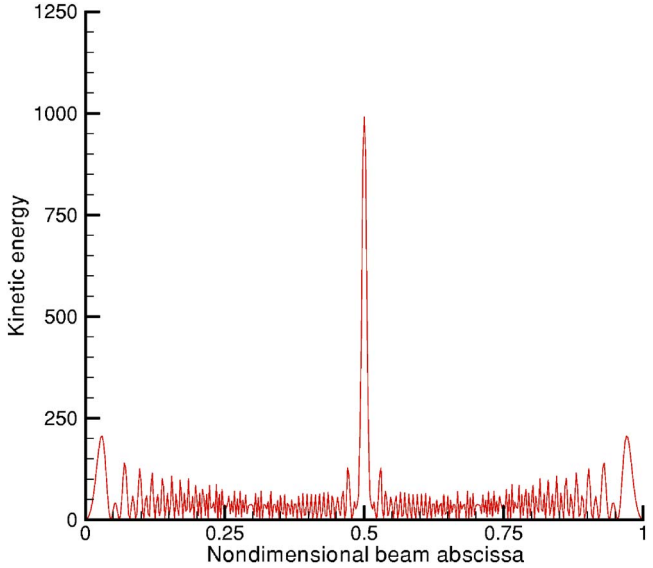


FIG. 14. (Color online) Kinetic energy distribution along the beam $T=2$.

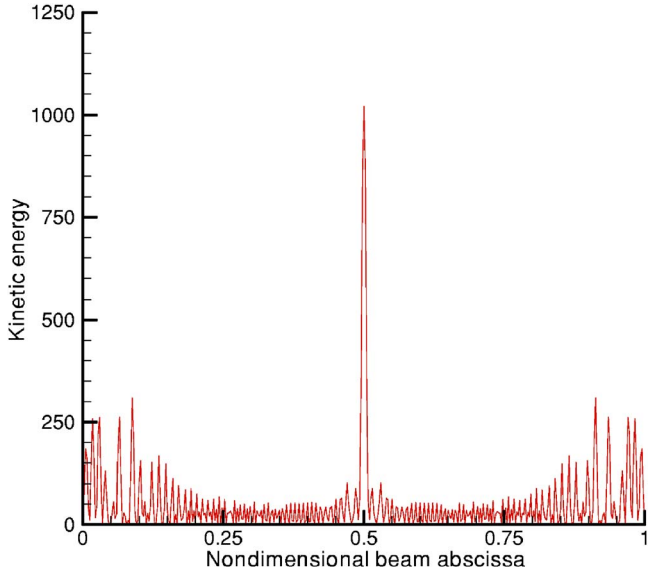


FIG. 15. (Color online) Kinetic energy distribution along the beam $T=2.5$.

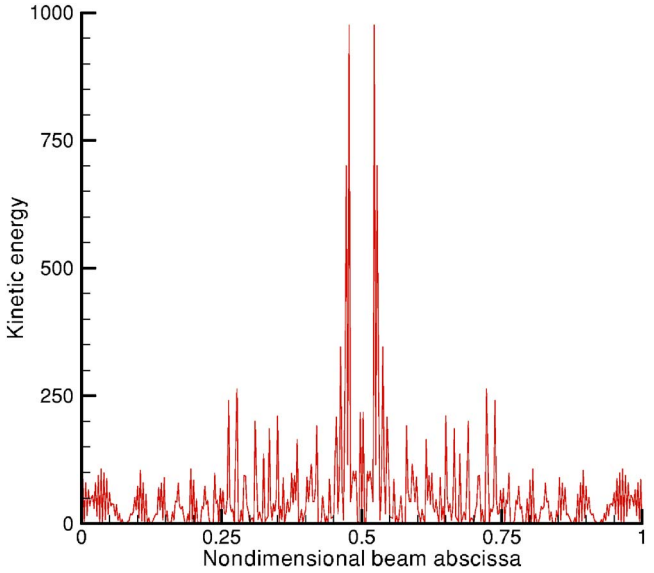


FIG. 17. (Color online) Kinetic energy distribution along the beam $T=5$.

ing property. The second argument relies on the concept of wave propagation related to the modal density singularities. In this case, wave filtering along the structure is produced, inhibiting the propagation of select wave groups in the waveguide. This effect produces again an impulse response, characterized by a sharp decay in time. The concept of apparent damping and near irreversibility was illustrated by means of an example based on the impulse response of a simply supported beam using both approaches.

APPENDIX: CONDITION TO PREVENT INSTABILITY

The compression force N applied to the beam on an elastic foundation can produce instability conditions defined by a critical load:[13]

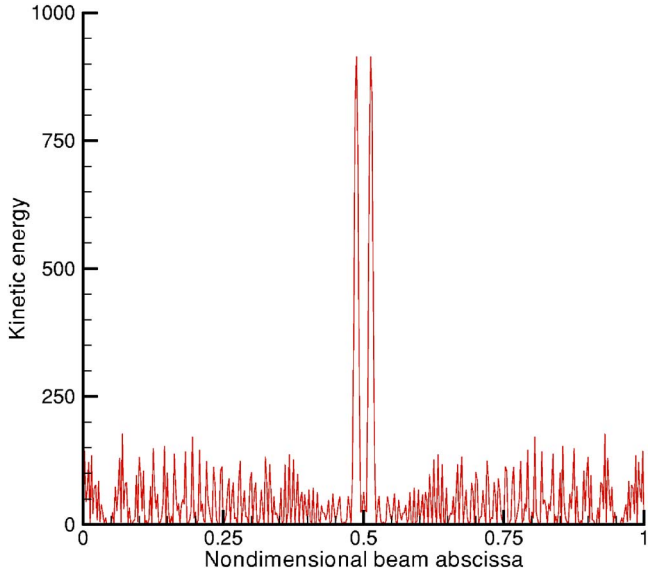


FIG. 16. (Color online) Kinetic energy distribution along the beam $T=3.5$.

$$N_{\mathrm{CR}} = \frac{\pi^2 B}{L^2}\left(m^2 + \frac{\chi L^4}{m^2 \pi^2 B}\right), \tag{A1}$$

where m is an integer that makes N_{CR} minimum. Its expression follows as

$$m^2(m+1)^2 = \frac{\chi L^4}{\pi^2 B}.$$

For $m \gg 1$, i.e. for $\chi L^4/\pi^2 B \gg 4$, this expression simplifies as $m^4 \approx \chi L^4/\pi^2 B$, that when substituted in Eq. (A1) produces

$$N_{\mathrm{CR}} \approx 2B\sqrt{\frac{\chi}{B}}.$$

In a comparison with Eq. (4), the relationship $2Bk_{n^*}^2 = N$ yields the following condition for preventing instability:

$$N < N_{\mathrm{CR}} \rightarrow k_{n^*}^2 < \sqrt{\frac{\chi}{B}} \rightarrow n^* < \frac{L}{\pi}\sqrt[4]{\frac{\chi}{B}}. \tag{A2}$$

Moreover, since $\sigma^2 = \chi/\rho A \omega^{*2} = \chi/Bk_{n^*}^4 = (1/n^{*4})(\chi L^4/\pi^4 B)$, and because of inequality (A2), the condition $\sigma > 1$ is implied.

[1] C. E. Celik and A. Akay, "Dissipation in solids: thermal oscillations of atoms," J. Acoust. Soc. Am. **108**, 184–191 (2000).
[2] R. L. Weaver, "The effect of an undamped finite degree of freedom 'fuzzy' substructure: numerical solution and theoretical discussion," J. Acoust. Soc. Am. **101**, 3159–3164 (1996).
[3] A. Carcaterra and A. Akay, "Transient energy exchange between a primary structure and a set of oscillators: return time and apparent damping," J. Acoust. Soc. Am. **115**, 683–696 (2004).
[4] I. Murat Koç, A. Carcaterra, Zhaoshun Xu, and A. Akay, "Energy sinks: vibration absorption by an optimal set of undamped oscillators," J. Acoust. Soc. Am. **118**, 3031–3042 (2005).
[5] A. Akay, Zhaoshun Xu, A. Carcaterra, and I. Murat Koç, "Experiments on vibration absorption using energy sinks," J. Acoust. Soc. Am. **118**, 3043–3049 (2005).
[6] A. D. Pierce, V. W. Sparrow, and D. A. Russel, "Foundamental structural-acoustic idealization for structure with fuzzy internals," ASME J. Vib., Acoust., Stress, Reliab. Des. **117**, 339–348 (1995).
[7] M. Strasberg and D. Feit, "Vibration damping of large structures induced by attached small resonant structures," J. Acoust. Soc. Am. **99**, 335–344

(1996).
[8]R. J. Nagem, I. Veljkovic, and G. Sandri, "Vibration damping by a continuous distribution of undamped oscillators," J. Sound Vib. **207**, 429–434 (1997).
[9]R. L. Weaver, "Mean and mean square responses of a prototypical master/fuzzy system," J. Acoust. Soc. Am. **101**, 1441–1449 (1997).
[10]R. L. Weaver, "Equipartition and mean square response in large undamped structures," J. Acoust. Soc. Am. **110**, 894–903 (2001).
[11]G. Maidanik, "Induced damping by a nearly continuous distribution of a nearly undamped oscillators: linear analysis," J. Sound Vib. **240**, 717–731 (2001).
[12]M. V. Drexel and J. H. Ginsberg, "Modal overlap and dissipation effects of a cantilever beam with multiple attached oscillators," ASME J. Vib., Acoust., Stress, Reliab. Des. **123**, 181–187 (2001).
[13]S. P. Timoshenko and J. M. Gere, *Theory of Elastic Stability* (McGraw-Hill, New York, 1988).

The modeling of the radiation and response Green's function of a fluid-loaded cylindrical shell with an external compliant layer

J. M. Cuschieri
Lockheed Martin MS2, Perry Technologies, Riviera Beach, Florida 33404

(Received 5 July 2005; revised 20 November 2005; accepted 10 January 2006)

Scattering and radiation of acoustic waves from a fluid-loaded cylindrical shell with an external compliant layer are of interest. The compliant layer can be modeled by a normally reacting impedance layer, which has the advantage that complex compliant layer geometries, such as partial compliant layers, can be considered. A question may, however, arise as to the accuracy of this approach. A more rigorous approach is to model the elastic shell and compliant layer using a multilayer shell theory, which has the disadvantage that it cannot be extended to consider partial layers. In this paper scattering results from the normally reacting compliant layer model are compared to those from the multilayer shell model to show that the two approaches produce similar results, except for thickness resonances of the compliant layer. Having established the consistency between the two approaches, results for the far-field acoustic radiation as a function of frequency and radiation angle for a fluid-loaded shell with an external compliant layer excited by an internal ring force are obtained using the normally reacting impedance layer model. These results clearly show the reduction in the far field radiation due to the presence of the compliant layer. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2173068]

PACS number(s): 43.40.Fz, 43.40.Rj [SFW] Pages: 2150–2169

I. INTRODUCTION

The problem of acoustic radiation and scattering from fluid-loaded cylindrical shells with an external compliant layer has in the past been generally analytically modeled using multilayer shell theory.[1–3] An advantage of these models is that the number of layers is unlimited and the thickness of each layer can be varied and not restricted by a thin shell requirement. Each layer is described by its thickness and longitudinal and shear wave speeds. There is, however, interest in considering the influence of the compliant layer on the radiation and scattering when the cylinder is only partially covered by the compliant layer. The multilayer shell approach does not readily extend to shells with partial compliant layers. An alternative to the multilayer approach is to model the compliant layer as a normally reacting surface[4] that can handle partial layers.

Modeling the external compliant layer by a normally reacting impedance surface is not as rigorous as using the multilayer approach, leading to the question of accuracy of the modeling approach. Available in the literature for multilayer shells are scattering results, therefore, scattering results from the two approaches are generated and compared. The accuracy of the multilayer approaches in the literature[1–3] have been verified by setting the material properties of all layers to be identical and then compare the scattering results to those for a fluid-loaded shell of the same total thickness. Reference to experimental results is also made in Ref. 3 as a basis to verify the approach.

Using the multilayer shell theory, the case of a two-layered shell, where the outside layer is acoustically compliant and absorptive is considered. The same problem is solved by modeling the compliant layer by a normally reacting mass/stiffness impedance layer on the external surface of the shell. Scattering results are obtained from both approaches and good agreement is shown, provided the compliant layer is thin relative to the acoustic wavelength. The scattering results are generated for plane wave input, which is of relatively long wavelength compared to the thickness of the compliant layer, and thus it would be expected that good matching is obtained between the two approaches when solving for the scattering problem provided the thickness of the compliant layer is much less than the surrounding acoustic wavelength. In the case of far-field radiation, relatively long wavelengths also dominate and therefore the validation based on scattering can be extended to the case of far-field radiation.

Having established the acceptability of using the normally reacting surface approach to model an external compliant layer, results for the far-field acoustic radiation for a shell, fully covered by an external compliant layer and excited by an internal ring load are generated and compared to the case of a bare shell without a compliant layer. Also included are results for the response Green's function even though no direct validation is provided for these results. The inclusion of the response Green's functions results help understand the behavior of the shell with the compliant layer.

II. MULTILAYER SHELL THEORY

Doolittle and Uberall[1] developed a formulation for an infinite, multilayer cylindrical shell, with fluid inside and outside and excited by an incident plane acoustic wave. The coupled equations of motion are written in terms of the elastic equations for the shell layers and the acoustic pressures in the interior and exterior acoustic fields. The solution for the

 0001-4966/2006/119(4)/2150/20/$22.50

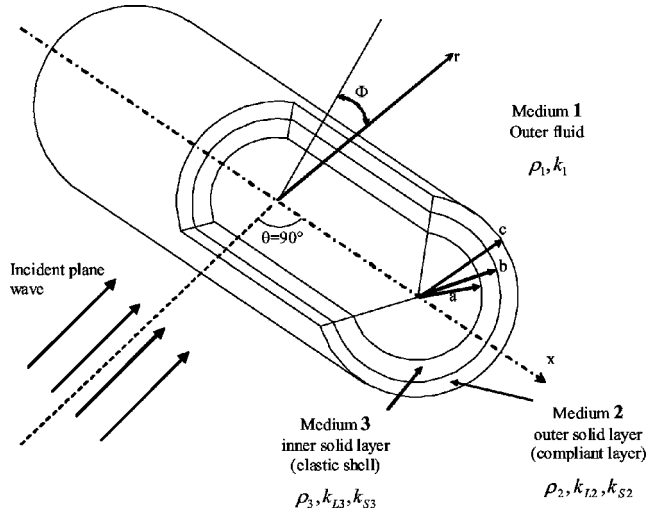


FIG. 1. Geometry of the infinite circular two-layered cylindrical shell, with an acoustic plane wave incident perpendicular to the surface of the shell.

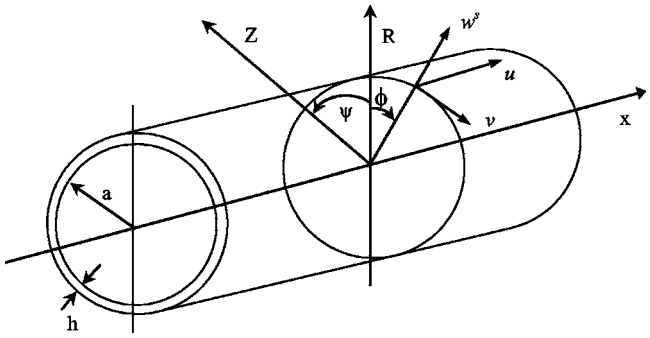


FIG. 2. Geometry and coordinate system for the cylindrical shell with the normally reacting compliant layer.

scattered pressure is obtained in terms of potential functions whose coefficients are determined from continuity of stresses and displacements at the interfaces between the layers and the outside and inside surfaces of the shell and the acoustic media.

Gaunaurd[2] used the approach developed in Ref. 1 to solve the problem of an infinite hollow elastic cylinder covered with an ideal layer of viscoelastic material or sound-absorbing material. It is shown that because of the viscoelastic nature of the compliant layer, some of the incident acoustical energy is absorbed by the compliant layer. The viscoelastic behavior of the compliant layer is controlled by the Kelvin-Voigt model, and the motion of the compliant layer is described by the exact three-dimensional (3D) Navier equations. The solution is in terms of potential functions whose coefficients are again determined from the boundary conditions at the interfaces. Flax[3] used a similar approach as in Refs. 1 and 2 to obtain the scattering of plane waves by a two layered cylindrical shell in water. Either or both of the two layers of elastic material may be dissipative. The solution as a function of frequency is examined for several combinations of inner and outer layers, and reference is made to consistency of the results with experimental data.

In this paper the multilayered shell is composed of two layers, with the outer layer representing the compliant layer and can be absorptive. The inner layer is considered to be a lossless elastic layer. The shell is excited by an incident plane wave, with an angle of incidence $\theta=90°$ to the shell axial direction. A solution is derived in terms of potential functions and solved for the scattered pressure in the fluid surrounding the shell. The configuration of the shell model with the surrounding fluid is as shown in Fig. 1, where a is the radius of the inner core, b is the inner radius and c is the outside radius of the compliant layer, respectively. Each layer is defined by the material density and wavenumber components—in the case of elastic solids, the wavenumber components are k_L and k_S, respectively, the longitudinal and the shear wavenumbers.

For an isotropic elastic body, the components of displacement must satisfy the exact Navier equation of motion. Expressing the displacement fields (u_x, u_r, u_θ) in terms of two potentials:

$$u = -\nabla\Phi + \nabla \times \Psi, \tag{1}$$

where Φ is a scalar potential that describes compressional waves, and Ψ a vector potential that describes shear waves. For an infinite shell with excitation by an axially uniform plane wave, the displacement fields are independent of the longitudinal axis of the shell (x), such that $\Psi_r=\Psi_\theta=0$ (which leads to $u_x=0$). The displacement equations of motion are satisfied if the two potentials functions $\Phi(r,\theta)$ and $\Psi_X(r,\theta)$ satisfy the two uncoupled Helmholtz equations.[1–3]

$$(\nabla^2 + k_L^2)\Phi(r,\theta) = 0, \tag{2}$$

$$(\nabla^2 + k_S^2)\Psi_X(r,\theta) = 0. \tag{3}$$

In the exterior fluid (medium 1), the pressure field is governed by the Helmholtz equation. The total pressure is the sum of the incident and scattered field components

$$p_1 = p_i + p_s, \tag{4}$$

$$p_i = p_0 \sum_{n=0}^{\infty} \in_n i^n J_n(k_1 r)\cos(n\phi), \tag{5}$$

$$p_s = p_0 \sum_{n=0}^{\infty} \in_n i^n C_n H_n^{(1)}(k_1 r)\cos(n\phi), \tag{6}$$

where p_0 is the incident pressure amplitude, $\in_n=2-\delta_{n0}$ is the Newman factor, $H_n^{(1)}$ is the Hankel function of the first kind, J_n is the Bessel function, C_n are the scattering coefficients, and n is the circumferential mode number. The external pressure satisfies the Sommerfeld radiation condition (the pressure represents outward traveling waves), and for continuity the norm of the pressure and the radial component of fluid displacement at the shell-fluid interface must satisfy the condition

$$u_r^{(1)} = \frac{1}{\rho_1 \omega^2}\frac{\partial p_1}{\partial r}. \tag{7}$$

The solutions to the potential function for the elastic layer and the compliant layer are given by

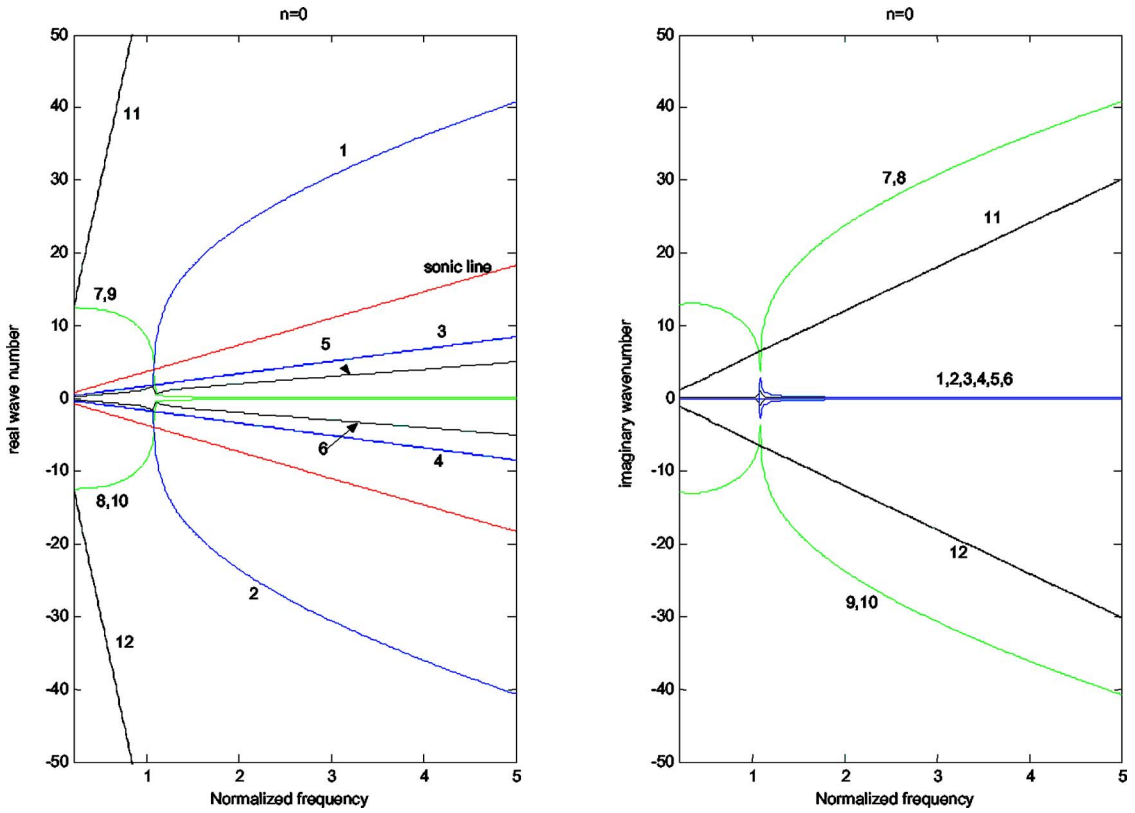


FIG. 3. (Color online) Dispersion curves for circumferential mode, $n=0$, obtained by finding the roots of the denominator of the characteristic equation (independent of response quantity) for a cylinder with an external compliant layer and fluid-loading.

$$\Phi_2 = p_0 \sum_{n=0}^{\infty} \in_n i^n [A_n J_n(k_{L2} r) + B_n Y_n(k_{L2} r)] \cos(n\phi), \quad (8)$$

$$\Psi_2 = p_0 \sum_{n=0}^{\infty} \in_n i^n [D_n J_n(k_{S2} r) + E_n Y_n(k_{S2} r)] \sin(n\phi), \quad (9)$$

$$\Phi_3 = p_0 \sum_{n=0}^{\infty} \in_n i^n [G_n J_n(k_{L3} r) + K_n Y_n(k_{L3} r)] \cos(n\phi), \quad (10)$$

$$\Psi_3 = p_0 \sum_{n=0}^{\infty} \in_n i^n [I_n J_n(k_{S3} r) + L_n Y_n(k_{S3} r)] \sin(n\phi), \quad (11)$$

where k_{L2} and k_{S2} are the longitudinal and shear wave numbers in the compliant layer and $k_{L3}=\omega/C_{L3}$, $k_{S3}=\omega/C_{S3}$, are the longitudinal and shear wave numbers in the elastic layer. The elastic layer is assumed to be lossless and thus the wavenumber components k_{L3} and k_{S3} are real valued. The compliant layer can be dissipative and hence the wavenumber components k_{L2} and k_{S2} can be complex. An *in vacuo* condition is assumed for the interior of the shell. The solution consists of a set of nine equations with nine unknown constants which are determined from the boundary conditions. Details of the solution can be found in Refs. 1–3.

III. COMPLIANT LAYER AS A NORMALLY REACTING LAYER

Cuschieri and Feit[4] formulated an approach based on modeling the compliant layer by a normally reacting layer between the shell and the surrounding fluid. The compliant layer consists of a mass/stiffness impedance in series/parallel to the surface of the shell.[4] The compliant layer impedance is expressed as a fraction of the impedance of the surrounding fluid.

Figure 2 shows the geometry and coordinate system of the cylindrical shell of thickness h and radius a. It is assumed that the shell is thin with a large radius-to-thickness ratio ($h \ll a$). The Donnel-Mushtari[5] coupled equations of motion are used to describe the motion of the shell in the axial (u), circumferential (v), and radial direction (w_s). The radial displacement w_s is directed outward along the normal to the shell surface, the tangential displacement (v) is in the direction of increasing ϕ, and the displacement (u) is directed in the direction of the cylindrical shell generator (axial direction).

These equations of motion are solved using a spatial Fourier transform in the axial direction, and a modal decomposition in the circumferential direction.[6] The solution for the response can be obtained for specific values of n, the circumferential mode number. Writing the equations of motion for particular circumferential mode n in matrix form[6]

$$[S^n(\bar{k},n)] \begin{bmatrix} \tilde{U}_n \\ \tilde{V}_n \\ \tilde{W}_n^s \end{bmatrix} = \begin{bmatrix} -\tilde{F}_{U_n} \\ \tilde{F}_{V_n} \\ \tilde{F}_{W_n} - \tilde{P}_s^n - \tilde{P}_i \end{bmatrix}, \quad (12)$$

where $\bar{k}$ is the normalized wavenumber with respect to the acoustic wavenumber k_0, $\bar{k}=k/k_0$, and $S^n(\bar{k},\omega)$ is the system

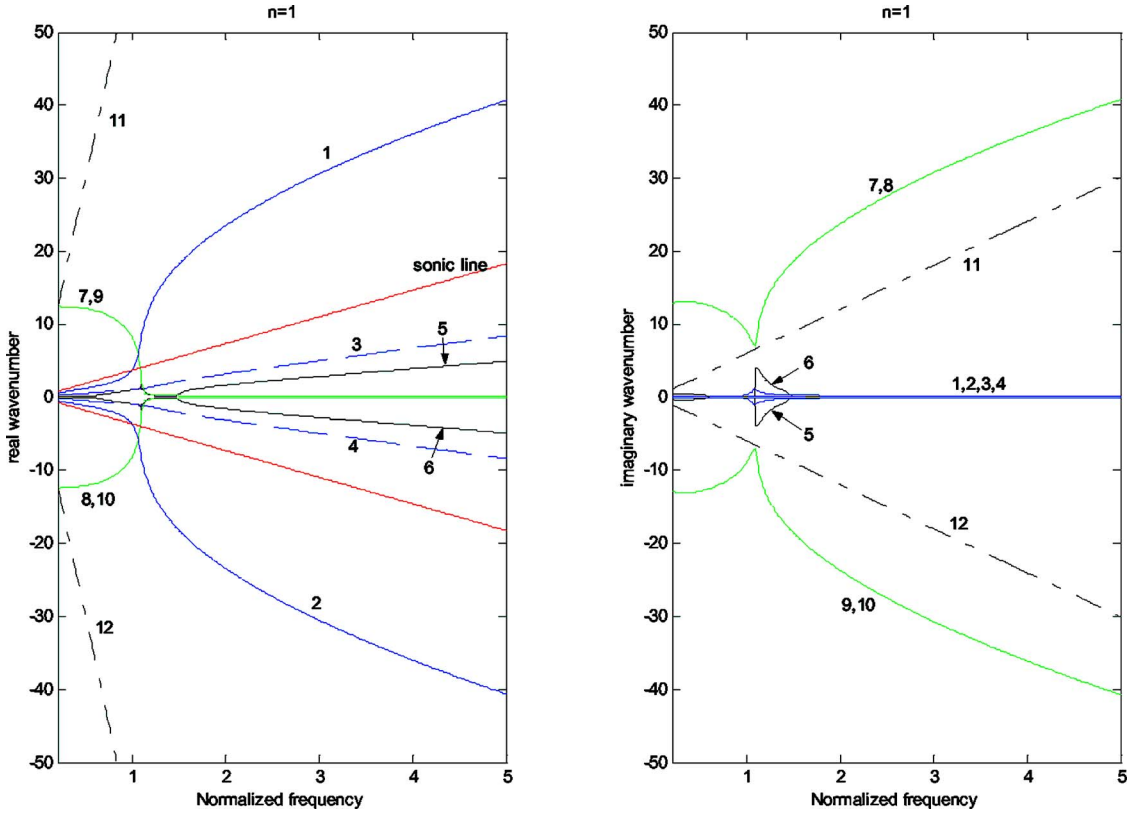


FIG. 4. (Color online) Dispersion curves for circumferential mode, $n=1$, obtained by finding the roots of the denominator of the characteristic equation (independent of response quantity) for a cylinder with an external compliant layer and fluid-loading.

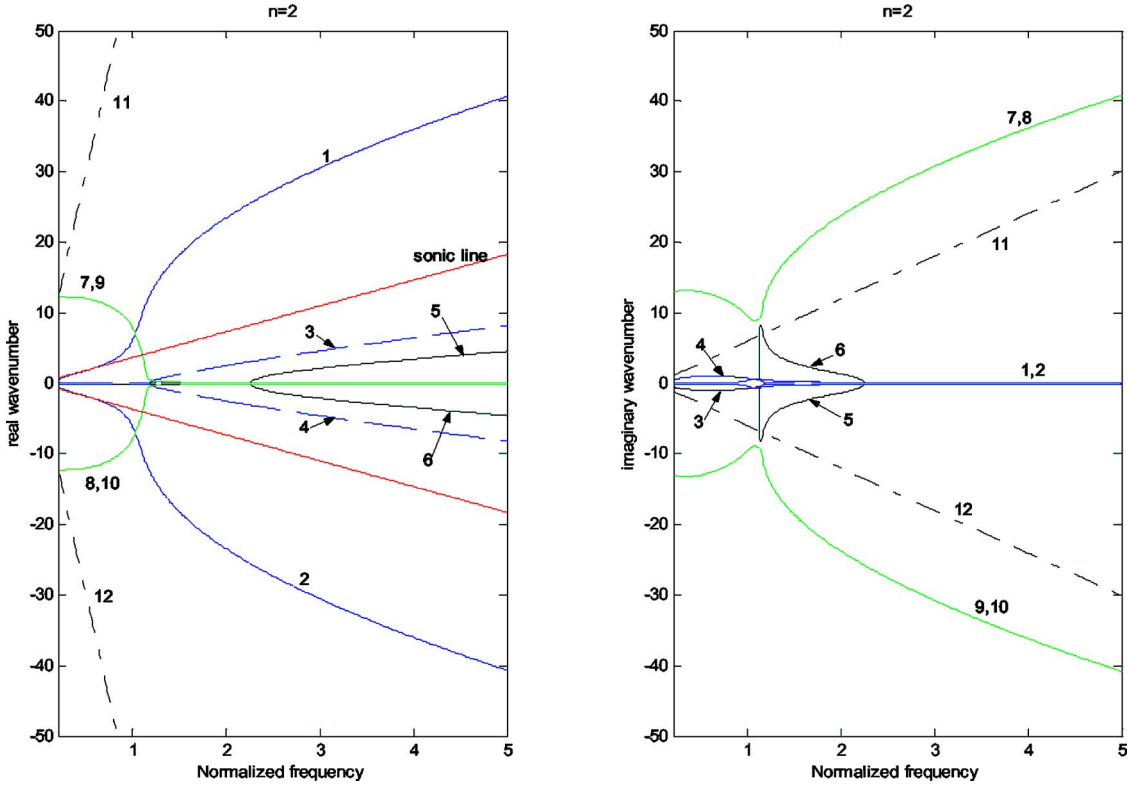


FIG. 5. (Color online) Dispersion curves for circumferential mode, $n=2$, obtained by finding the roots of the denominator of the characteristic equation (independent of response quantity) for a cylinder with an external compliant layer and fluid-loading.

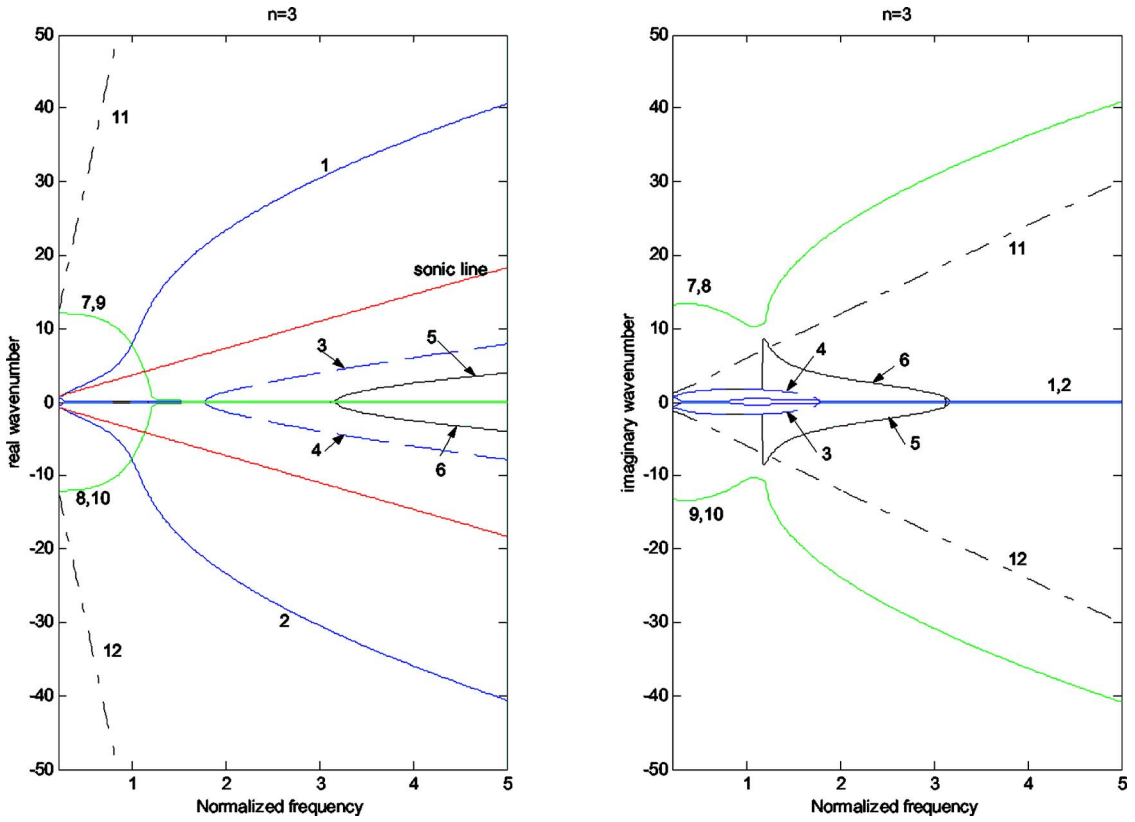


FIG. 6. (Color online) Dispersion curves for circumferential mode, $n=3$, obtained by finding the roots of the denominator of the characteristic equation (independent of response quantity) for a cylinder with an external compliant layer and fluid-loading.

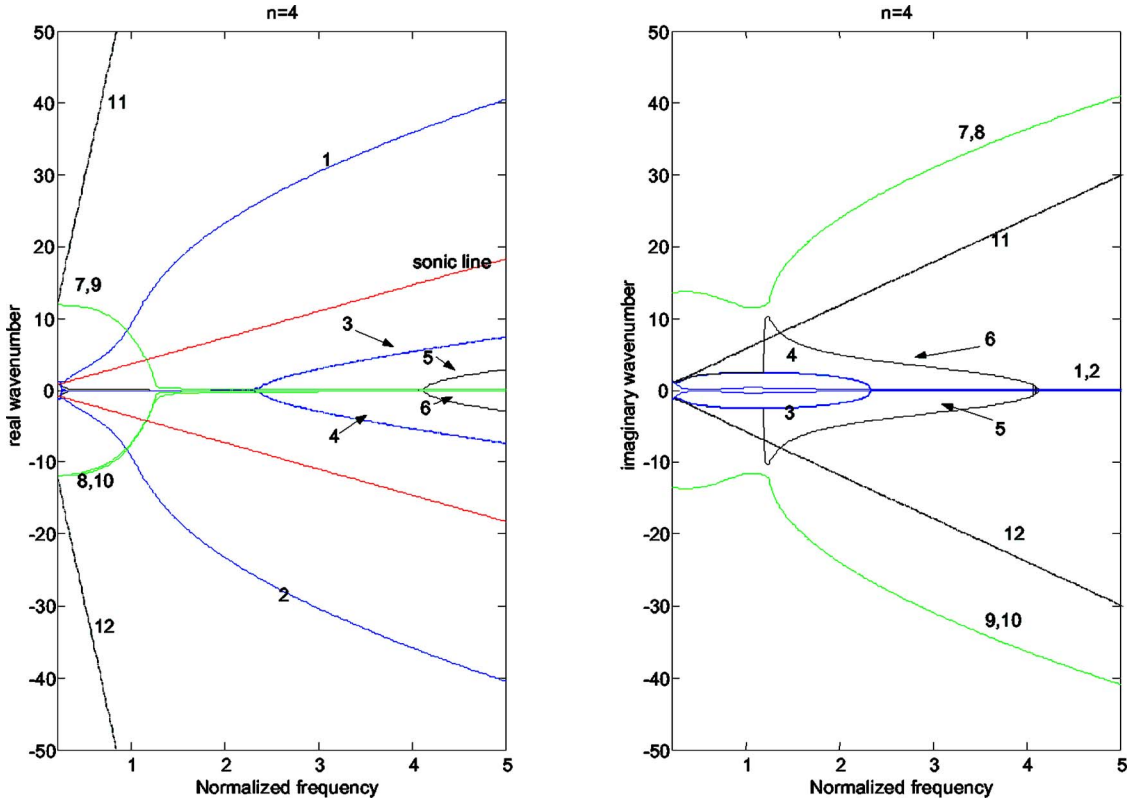


FIG. 7. (Color online) Dispersion curves for circumferential mode, $n=4$, obtained by finding the roots of the denominator of the characteristic equation (independent of response quantity) for a cylinder with an external compliant layer and fluid-loading.

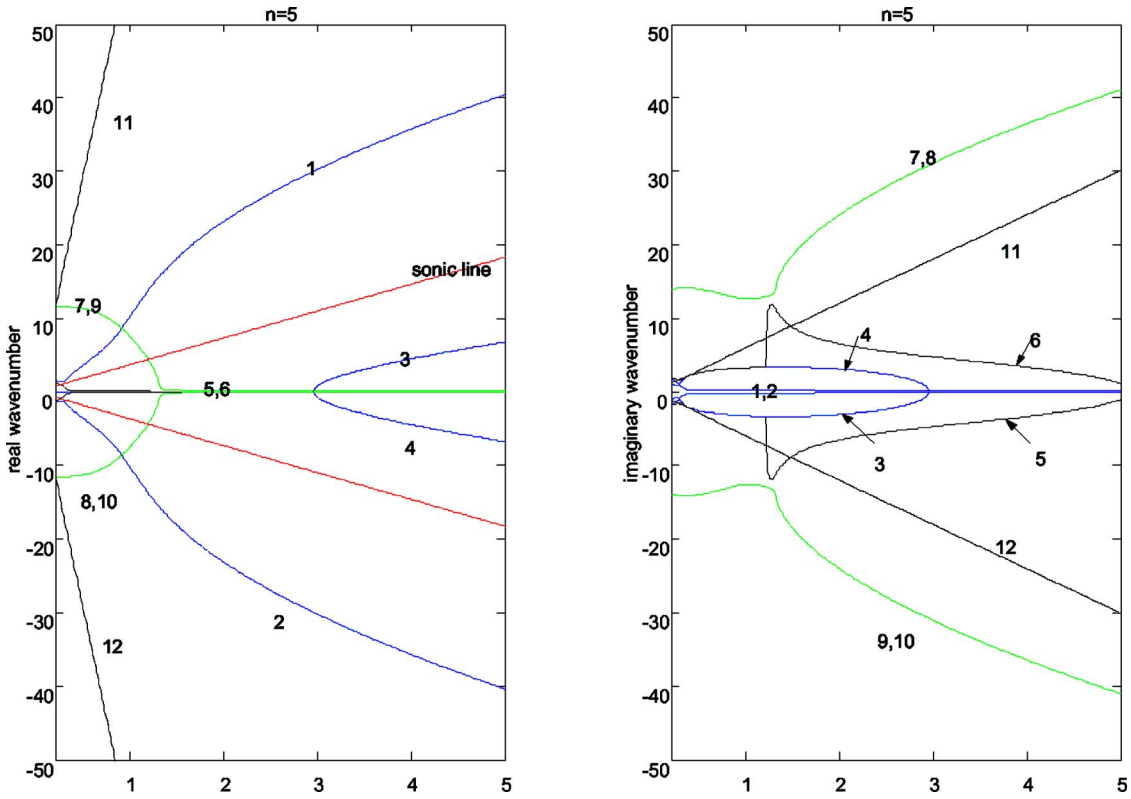


FIG. 8. (Color online) Dispersion curves for circumferential mode, $n=5$, obtained by finding the roots of the denominator of the characteristic equation (independent of response quantity) for a cylinder with an external compliant layer and fluid-loading.

response matrix definition of which can be found in Ref. 6.

Equation (12) includes excitation by incident pressure $\tilde{P}_i^n$ and ring internal loads $\tilde{F}_{U_n}, \tilde{F}_{V_n}, \tilde{F}_{W_n}$, respectively, in the axial, tangential, and radial directions. Thus Eq. (12) can be used to solve for both the scattering problem or the radiation problem.

The compliant layer impedance Z^c is given in the form

$$Z^c = -j\rho_0 c_0 \xi(1 - j\eta), \tag{13}$$

where ξ is the ratio of the compliant layer impedance to the impedance of the acoustic fluid $(\rho_0 c_0)$, ρ_0 and c_0 are the density and acoustic wave speed in the external acoustic fluid medium, and η represents the damping of the compliant layer.[7] Using the impedance function for the compliant layer, the radial displacement of the exterior surface of the compliant layer $w_a(x,\phi)$ satisfies the relationship[4]

$$j\omega(w^a(x,\phi) - w^s(x,\phi)) = -\left(\frac{p_s(x,\phi)}{Z^c} + \frac{p_i(x,\phi)}{Z^c}\right), \tag{14}$$

where p_s is the scattered or radiation pressure, p_i is the incident pressure, w^s is the displacement at the surface of the shell, w^a is the displacement at the surface of the compliant layer and x and ϕ are, respectively, the axial and angular coordinates. For a shell completely covered by a compliant layer, the functions are independent of ϕ.

Normalizing the equations of motion[4,6] and the impedance of the compliant layer (Z^c), the normalized impedance is defined by

TABLE I. Multilayer shell model material properties.

	Compliant layer	Shell layer
Ratio of Density of water to steel	$(\rho_2/\rho)(C_{L2}/c_0)=0.06$	$\rho_3/\rho=0.136$
Ratio of the longitudinal wave speed to acoustic wave speed in water		$C_{L3}/c_0=3.65$
Poisson's ratio	$\upsilon=0.4$	$\upsilon=0.3$
Ratio of the shear wave speed to acoustic wave speed in water	$C_{s2}/c_0=(C_{L2}/c_0)[1/\sqrt{2(1+\upsilon)}]$	$C_{s3}/c_0=(C_{L3}/c_0)[1/\sqrt{2(1+\upsilon)}]$
Damping	$\eta_{L2}=\eta_{S2}=0.1$	

$$\bar{Z}_c = j\omega\gamma Z_c = \left(\frac{a}{h}\right)\left(\frac{\rho_0}{\rho}\right)\frac{\Omega\xi(1-j\eta)}{M}, \tag{15}$$

where $M=c_p/c_0$, with $c_p=\sqrt{E/[\rho(1-\upsilon^2)]}$ and $\gamma=a^2(1-\upsilon^2)/(Eh)$.

Spatial Fourier transforming and substituting for $\tilde{P}_s^n$, Eq. (14) becomes

$$\tilde{W}_n^a(\bar{k}) - \tilde{W}_n^s(\bar{k}) = -\frac{F_n^L}{\bar{Z}^c}\tilde{W}_n^a(\bar{k}) - \frac{\gamma\tilde{P}_i(\bar{k})}{\bar{Z}^c}, \tag{16}$$

where F_n^L is the fluid-loading (related to the radiation impedance of the external acoustic medium), and is given by

$$F_n^L = \Omega^2\left(\frac{a}{h}\right)\left(\frac{\rho_0}{\rho}\right)\frac{H_n(\chi)}{\chi H_n'(\chi)}. \tag{17}$$

The solution to Eq. (12) for the shell response components, in the wavenumber domain, for mode n is given by

$$\begin{bmatrix} \dot{\tilde{U}}_n \\ \dot{\tilde{V}}_n \\ \dot{\tilde{W}}_n^s \end{bmatrix} = \frac{-j\Omega}{\Delta S_n\left(1+\frac{F_n^L}{\bar{Z}^c}\right)+F_n^L\Delta_{33}} \begin{bmatrix} \Delta_{11}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)+s_{22}F_n^L & \Delta_{12}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)-s_{21}F_n^L & \Delta_{31} \\ \Delta_{21}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)-s_{12}F_n^L & \Delta_{22}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)+s_{11}F_n^L & \Delta_{32} \\ \Delta_{13}\left(1+\frac{F_n^L}{\bar{Z}^c}\right) & \Delta_{23}\left(1+\frac{F_n^L}{\bar{Z}^c}\right) & \Delta_{33} \end{bmatrix} \times \begin{bmatrix} -\tilde{F}_{U_n} \\ \tilde{F}_{V_n} \\ \tilde{F}_{W_n}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)-\tilde{P}_i \end{bmatrix}, \tag{18}$$

where ΔS_n is the determinant of the 3 by 3 $[S(\bar{k},n)]$ matrix.

The bar over the velocity components indicates that these quantities are normalized with respect to $\gamma\omega_r$.

From Eq. (18) the solution for the nth mode radial velocity on the surface of the shell $\dot{\tilde{W}}_n^s(\bar{k},\Omega)$ is given by

$$\begin{aligned} \dot{\tilde{W}}_n^s(\bar{k},\Omega) &= \frac{-j\Omega}{\Delta S_n\left(1+\frac{F_n^L}{\bar{Z}^c}\right)+F_n^L\Delta_{33}} \\ &\times\left\{-\Delta_{13}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)\tilde{F}_{U_n}+\Delta_{23}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)\tilde{F}_{V_n}+\Delta_{33}\right. \\ &\left.\times\left[\tilde{F}_{W_n}\left(1+\frac{F_n^L}{\bar{Z}^c}\right)-\tilde{P}_i\right]\right\}. \end{aligned} \tag{19}$$

Substituting in the expression for the scattered or radiated pressure (depending on the values of $\tilde{P}_i^n$ and $\tilde{F}_{U_n},\tilde{F}_{V_n},\tilde{F}_{W_n}$)

$$\tilde{P}_s^n(\bar{k},R) = \frac{\Omega^2}{1+F_n^L/\bar{Z}^c}\left(\frac{a}{h}\right)\left(\frac{\rho_0}{\rho}\right)\frac{H_n(\chi R)}{\chi H_n'(\chi)} \times\left\{\tilde{W}_n^s(\bar{k})-\frac{\gamma\tilde{P}_i(\bar{k})}{\bar{Z}^c}\right\}. \tag{20}$$

In Eq. (20) $\tilde{W}_n^s(\bar{k})$ have contributions from the internal loads $\tilde{F}_{U_n},\tilde{F}_{V_n},\tilde{F}_{W_n}$ as well as the incident pressure $\tilde{P}_i^n$. Setting either the internal loads or the incident pressure to zero will result in the scattering solution or the radiation solution, respectively. After substituting for $\tilde{W}_n^s(\bar{k})$ in Eq. (20) and setting $\tilde{F}_{U_n}=\tilde{F}_{V_n}=\tilde{F}_{W_n}=0$, the resulting equation is that for scattering and is given by

$$\begin{aligned} P_{s1}^n(X,R) &= -\frac{\gamma\Omega M P_i}{\pi^2}\left(\frac{a}{h}\right) \\ &\times\left(\frac{\rho_0}{\rho}\right)\frac{H_n(\Omega M\cos\theta R)\in_n j^{n+1}}{M^2\cos^2\theta(H_n'(\Omega M\cos\theta))^2}e^{j\Omega MX\sin\theta} \\ &\times\left[\frac{1}{1+\frac{F_n^L}{\bar{Z}^c}}\left(\frac{\Delta_{33}}{\Delta S_n(1+F_n^L/\bar{Z}^c)+F_n^L\Delta_{33}}+\frac{1}{\bar{Z}^c}\right)\right]_{k=\sin\theta}, \end{aligned} \tag{21}$$

where P_{s1}^n represents the scattered pressure of an infinite fluid-loaded shell with an external compliant layer, excited by an incident plane wave. This component can be directly compared with the results of the scattered pressure from the multilayered shell theory after including the scattering from a rigid cylinder is given by

TABLE II. Impedance model material properties.

Poisson's coefficient for Steel	$\upsilon=0.3$
Ratio of the density of water to steel	$\rho_0/\rho=0.136$
Ratio of the longitudinal wave speed to acoustic wave speed in water	$M=c_p/c_o=3.65$
Damping of the compliant layer	$\eta=0.1$
Compliant layer impedance expressed as a fraction (ξ) of the impedance of the acoustic fluid	$\xi=0.06$

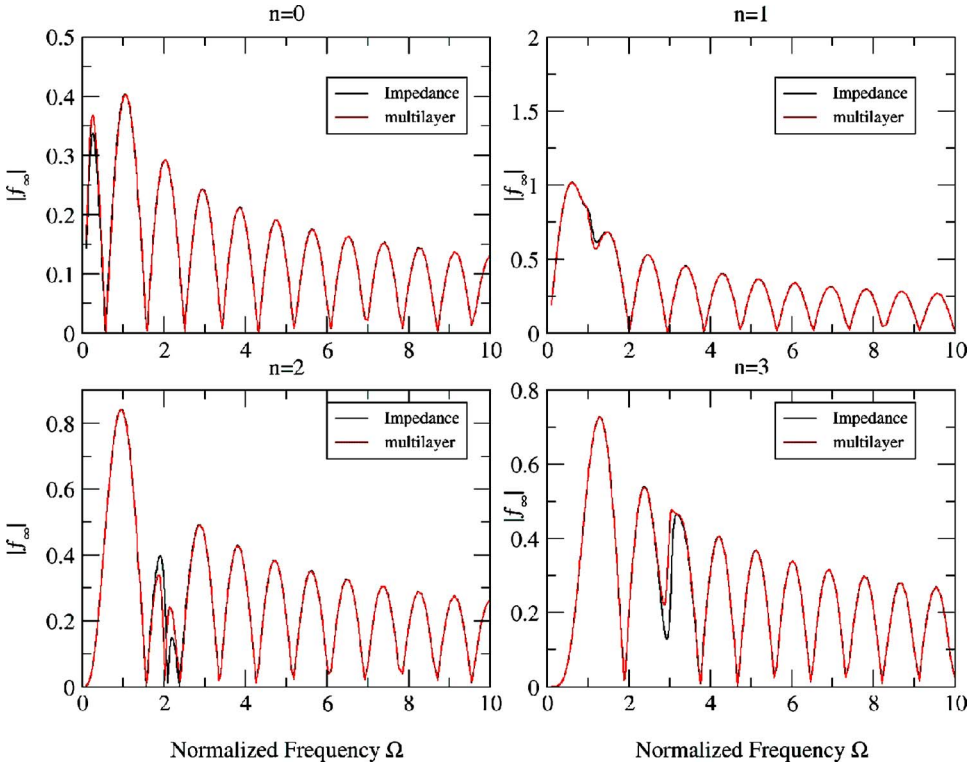


FIG. 9. (Color online) Comparison of model results for the far field backscatter form function for specific values of the circumferential modes for the bare shell.

$$p_r = P_i \sum_n \varepsilon_n \frac{j^n J'(\chi)_n H_n(\chi R)}{H_n'(\chi)} \cos(n\phi). \tag{22}$$

If the shell is excited by internal radial ring load, the radiated pressure field is given by

$$P_{s2}^n(X,R) = \frac{\gamma \Omega M}{2\pi} \int_{-\infty}^{+\infty} \Omega^2 \left(\frac{a}{h}\right)\left(\frac{\rho_0}{\rho}\right) \times \frac{H_n(\chi R)}{\chi H_n'(\chi)} \frac{\Delta_{33} \tilde{F}_{W_n}}{\Delta S_n \left(1 + \frac{F_n^L}{\bar{Z}^c}\right) + F_n^L \Delta_{33}} e^{j\Omega M \bar{k} X} d\bar{k}. \tag{23}$$

This expression can be solved either by using an inverse spatial Fourier transform or for the far field radiation by a stationary phase approach.[5] In the far-field, the radiated pressure is given by

$$P_{s2}^n(Z,\psi,\Omega) = \frac{\gamma \Omega M}{2\pi} (-j)^n \frac{2e^{j\Omega MZ}}{Z\Omega M} \left[\Omega^2 \left(\frac{a}{h}\right)\left(\frac{\rho_0}{\rho}\right) \times \frac{1}{\chi H_n'(\chi)} \frac{\Delta_{33} \tilde{F}_{W_n}}{\Delta S_n (1 + F_n^L/\bar{Z}^c) + F_n^L \Delta_{33}} \right]_{k=\sin\psi}, \tag{24}$$

where $Z = \sqrt{X^2 + R^2}$ is the normalized spherical distance from the location of the radial internal ring force and ψ is the radiation angle measured from the normal to the shell

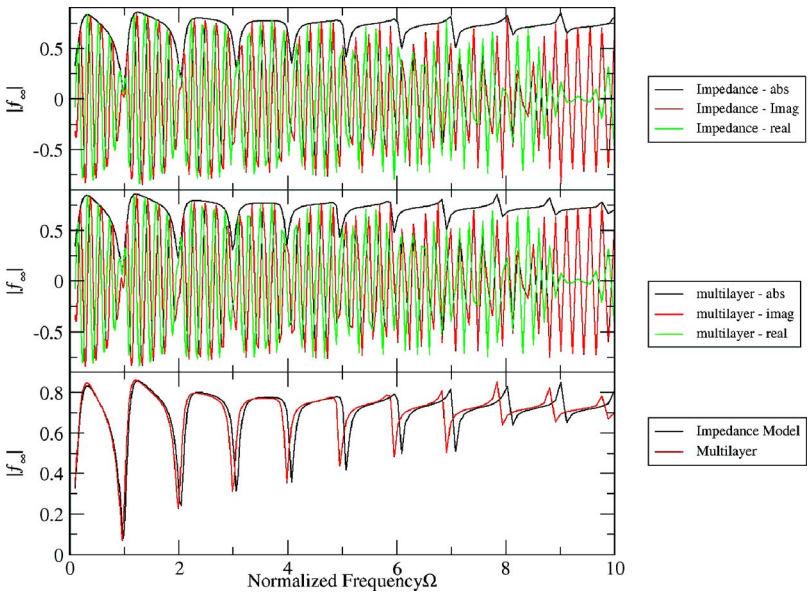


FIG. 10. (Color online) Model results for the total far field backscatter form functions for the bare shell.

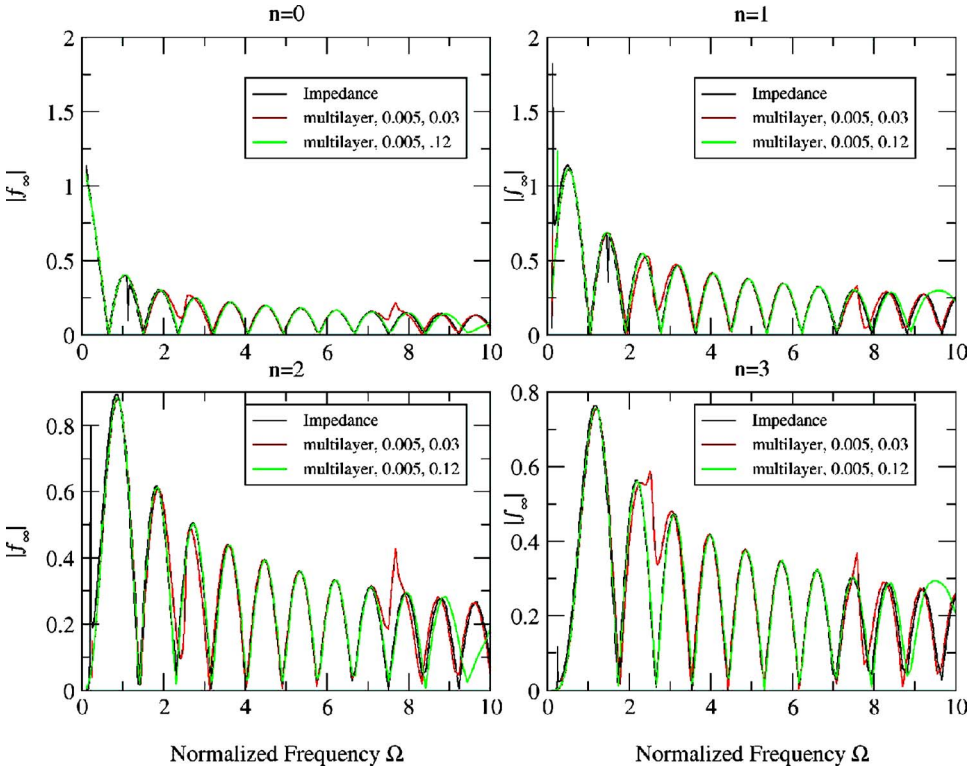


FIG. 11. (Color online) Model results for the far field backscatter form function for specific values of the circumferential modes for the shell with an external compliant layer. Compliant layer normalized thickness in the multilayer shell of 0.005 and two alternative normalized wave speeds of 0.03 and 0.12.

(Fig. 2). The response Green's function of the shell in the spatial domain is obtained from an inverse discrete fourier transform (IDFT) of the response function in the wavenumber domain [Eq. (19)]. This equation represents the response for the scattering case when $\tilde{F}_{U_n}=\tilde{F}_{V_n}=\tilde{F}_{W_n}=0$ or the response due to internal loading if $\tilde{P}_i^n$ is set to zero. The denominator of the integrand, which is the characteristic equation of the fluid-loaded shell with an external compliant layer, becomes singular at its roots. Therefore, the integrand cannot be directly evaluated using an IDFT and an alternative approach, such as the hybrid numerical/analytical method,[6,8] is required. In the hybrid method, the singularities that precluded the direct use of numerical integration are first removed from the integrand and their contribution evaluated using a contour integration. The smoothed integrand, free of any singularities, is then evaluated using an IDFT. The complete solution in the spatial domain is obtained by adding the contributions from the IDFT result and the singularities.

The denominator of the integrand

$$D(\bar{k},\Omega)=\Delta S_n\left(1+\frac{F_n^L}{\bar{Z}^c}\right)+F_n^L\Delta_{33}, \tag{25}$$

is the characteristic equation of the fluid-loaded shell with the external compliant layer. The zeros or roots of the characteristic equation represent the different type of waves. Figures 3–8 show the zeros of the shell's characteristic equation

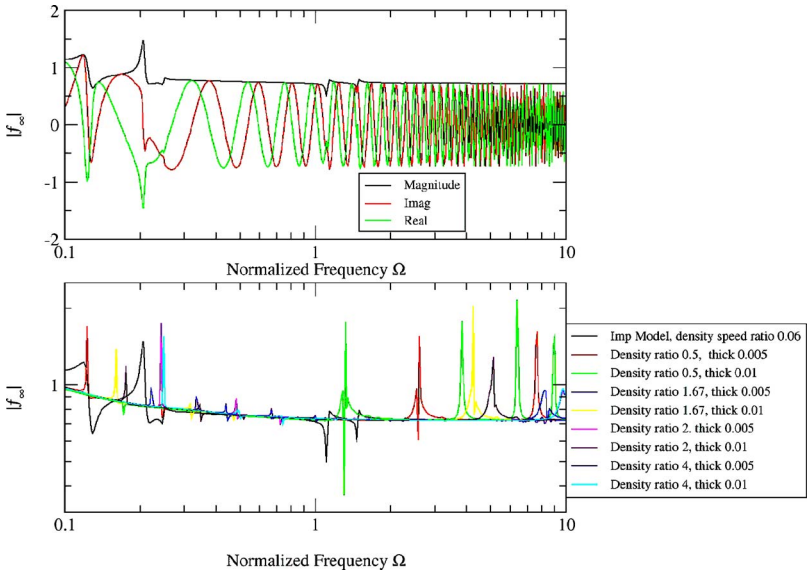


FIG. 12. (Color online) Model results for the total far field backscatter form function for shell with an external compliant layer. (a) Impedance model results; (b) impedance model and multilayer model results for the magnitude of the backscatter form function for two values of the normalized thicknesses of the multilayer external compliant layer and density ratios ranging from 0.5 to 4 for a constant ρc value of 0.06.

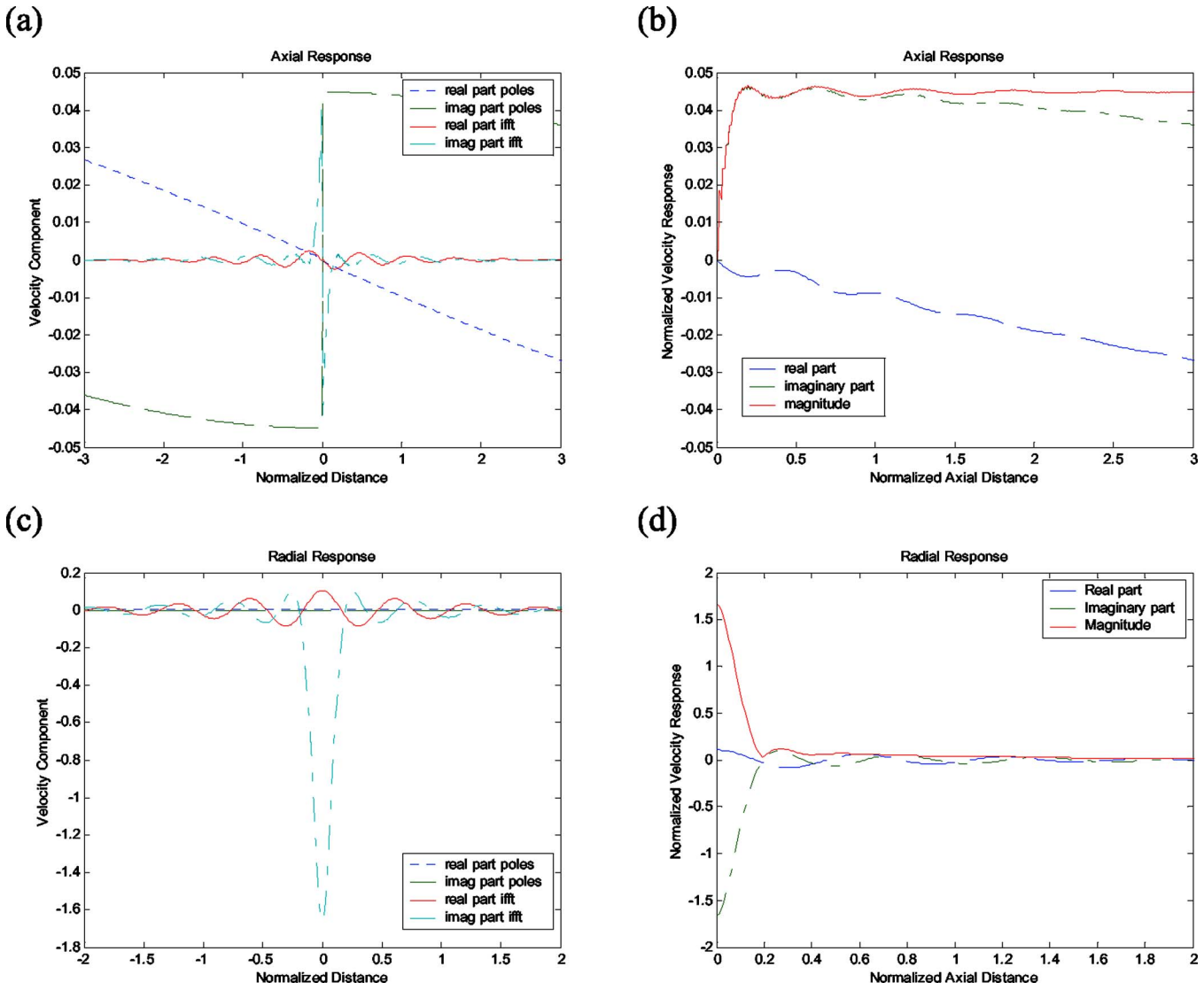


FIG. 13. (Color online) Axial and radial shell surface response for radial excitation as a function of the axial distance normalized with respect to the shell radius, for $n=0$, and $\Omega=0.2$. The response is normalized with respect to (γw_r). Axial response: (a) Contribution from the IDFT and the poles; (b) magnitude, real, and imaginary parts. Radial response: (c) Contribution from IDFT and poles; (d) magnitude, real, and imaginary parts.

for circumferential modes n equals 0, 1, 2, 3, 4, and 5 as a function of normalized frequency Ω. The real and imaginary components of the normalized wavenumber $\Omega\bar{k}M$ roots are plotted separately.

The sonic line (corresponding to waves propagating at the speed of sound in the acoustic medium) is shown in each figure. The roots above this line ($|\bar{k}|>|\bar{k}_0|$) correspond to subsonic waves, whereas the roots below this line ($|\bar{k}|<|\bar{k}_0|$) correspond to supersonic waves. The roots can be classified into different categories:

• Complex roots with small imaginary part

Two of these roots represent the quasi-flexural waves (denoted as 1 and 2 in the Figs. 3–8) which start as supersonic waves and then transition to subsonic waves. When these waves are subsonic, the imaginary component of these roots is due to the damping in the compliant layer. These quasi-flexural waves propagate along the shell with little attenuation.

The four other roots represent the torsional waves (3 and 4) and the axial longitudinal waves (5 and 6), these waves are supersonic. The four roots have small imaginary components associated with the loss of energy to acoustic radiation and/or due to the damping in the compliant layer as the waves propagate along the shell. Since this loss of energy is small, the imaginary part is also very small. It would be expected that the compliant layer does not add a significant amount of damping to these waves.

Because of their proximity to the real axis, all these roots create near discontinuities on the real axis, and have to be removed before evaluating the IDFT.

• Complex roots

Four complex roots have a real and imaginary component of the same order of magnitude (7, 8, 9, and 10). They occur as complex pairs which are the negative of each other. Combining roots with the same sign imaginary part but different sign real part produces standing waves in the vicinity of discontinuities which decay rapidly along the shell axial direction. These waves lose some energy to acoustic radia-

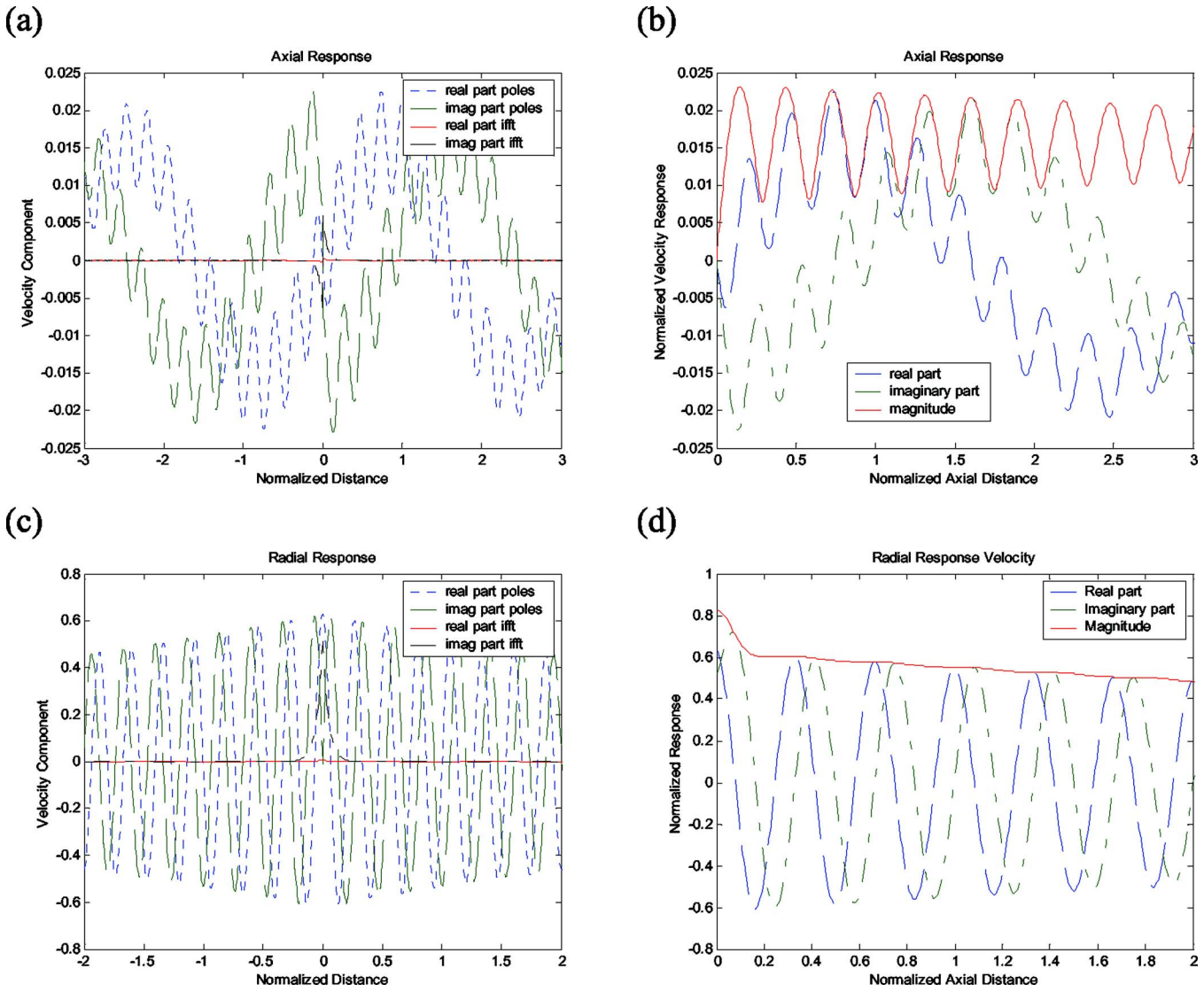


FIG. 14. (Color online) Axial and radial shell surface response for radial excitation as a function of the axial distance normalized with respect to the shell radius, for $n=0$, and $\Omega=2.0$. The response is normalized with respect to (γw_r). Axial response: (a) Contribution from the IDFT and the poles; (b) magnitude, real, and imaginary parts. Radial response: (c) Contribution from IDFT and poles; (d) magnitude, real, and imaginary parts.

tion, and are an important contributor to the acoustic pressure field in the vicinity of discontinuities.

• Compliant layer roots

Two complex roots are introduced by the compliant layer (11 and 12). These roots correspond to damped slow propagating waves. The magnitude of the imaginary part of these roots is a function of the compliant layer damping. If the compliant layer has no damping these roots would be near real.

• Purely imaginary roots

These roots represent evanescent waves below the cut-on frequency. An important aspect of the dispersion relationship is the cut-on frequency of the "near real" roots (corresponding to the tangential and axial compression waves). These waves do not exist below their cut-on frequency. For example, in the case of $n=2$ (Fig. 5), the cut-on frequency for the tangential wave is approximately $\Omega \approx 1.2$, while the cut-on frequency of the axial longitudinal wave occurs at a higher frequency: $\Omega \approx 2.2$. The cut-on frequency increases with the order of the circumferential mode. Below the cut-on frequency, the roots are purely imaginary, and thus do not propagate along the shell.

It can also be observed that in the particular case of circumferential mode 0, the dispersion curves present similarities to the ones obtained in the in-vacuo condition. The longitudinal waves have a cut-on frequency which corresponds to the ring frequency of the shell (at this frequency, radial hooplike resonance occurs). For circumferential modes greater than 0, the curves are similar to a fluid-loaded shell, except for the additional roots introduced by the compliant layer.

IV. VALIDATION RESULTS—SCATTERED PRESSURE

Results for the scattered pressure are obtained using both the multilayer approach and the normally reacting layer approach. To compare the results of both methods, it is necessary to establish parameters which will be identical for

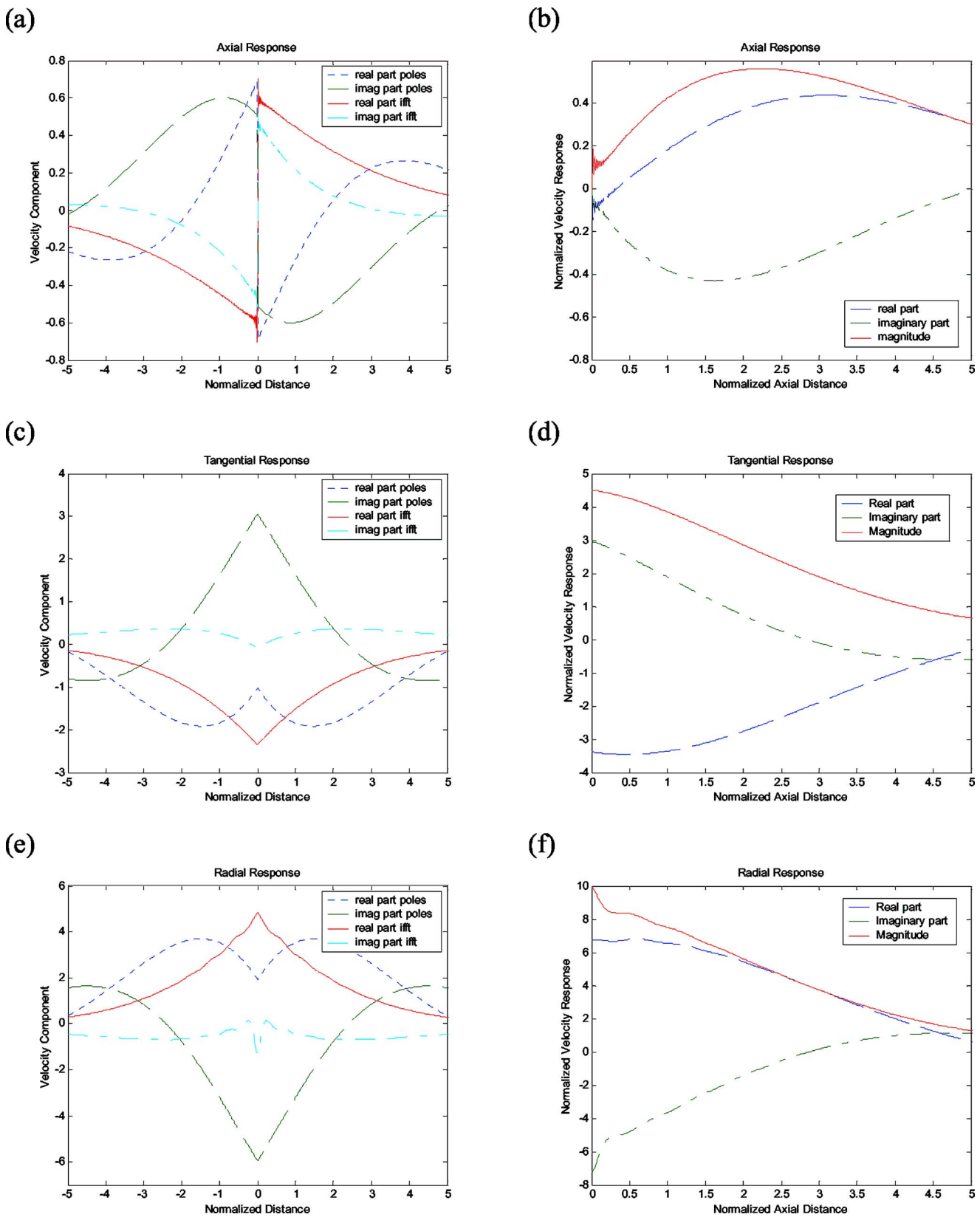


FIG. 15. (Color online) Axial, tangential and radial shell surface response for radial excitation as a function of the axial distance normalized with respect to the shell radius, for $n=2$, and $\Omega=0.2$. The response is normalized with respect to (γw_r). Axial response: (a) Contribution from the IDFT and the poles; (b) magnitude, real, and imaginary parts. Tangential response: (c) Contribution from IDFT and the poles, (d) magnitude, real, and imaginary parts. Radial response: (e) Contribution from IDFT and poles; (f) magnitude, real, and imaginary parts.

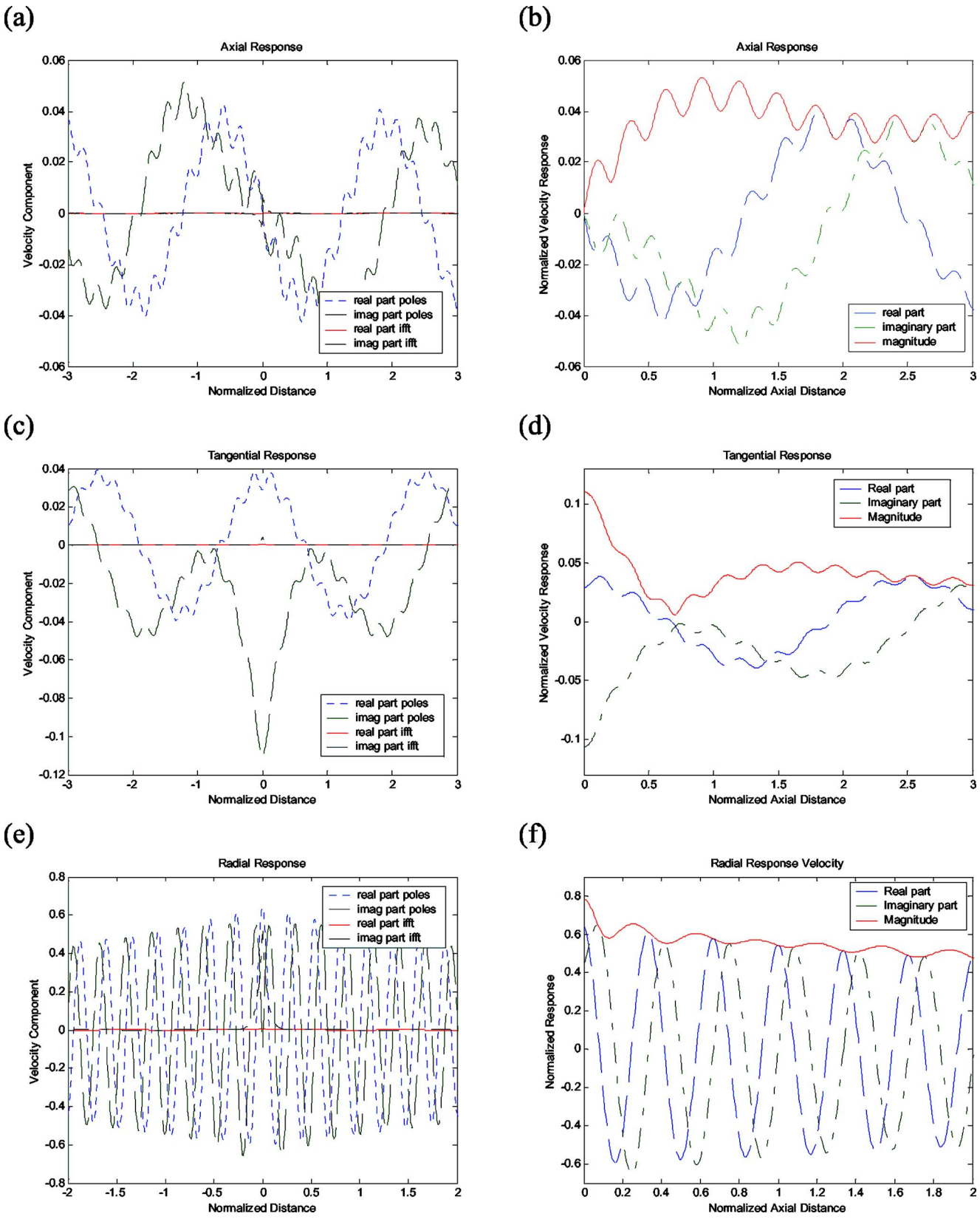


FIG. 16. (Color online) Axial, tangential and radial shell surface response for radial excitation as a function of the axial distance normalized with respect to the shell radius, for $n=2$, and $\Omega=2.0$. The response is normalized with respect to (γw_r). Axial response: (a) Contribution from the IDFT and the poles; (b) magnitude, real, and imaginary parts. Tangential response: (c) Contribution from IDFT and the poles, (d) magnitude, real, and imaginary parts. Radial response: (e) Contribution from IDFT and poles; (f) magnitude, real, and imaginary parts.

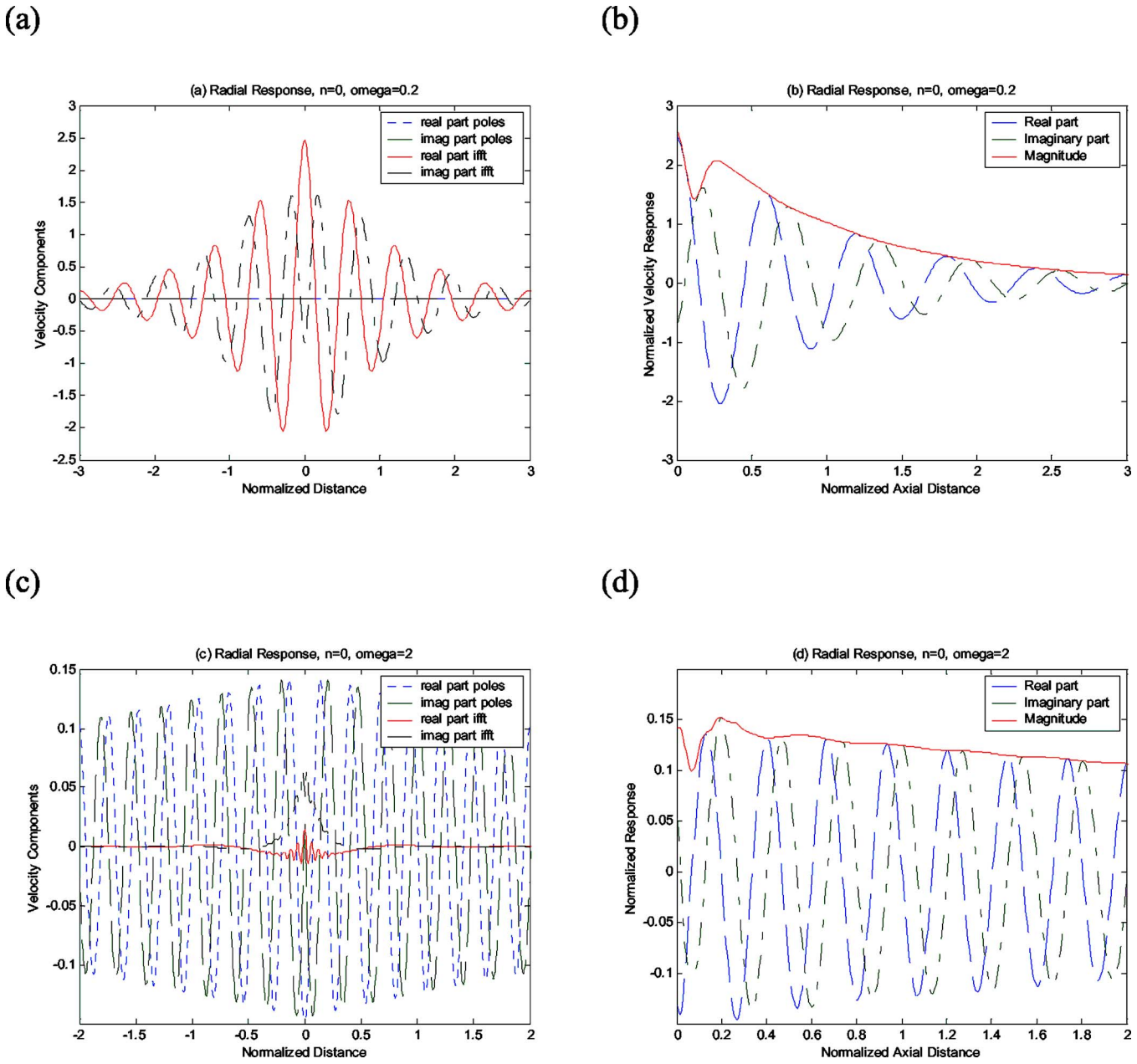


FIG. 17. (Color online) Normalized with respect to (γw_r). Radial response at the surface of the compliant layer for radial excitation for circumferential mode $n=0$. $\Omega=0.2$: (a) Contribution from IDFT and poles response, (b) magnitude, real, and imaginary parts. $\Omega=2.0$: (c) Contribution from IDFT and poles; (d) magnitude, real, and imaginary parts.

each model. The shell is excited by an incident plane wave with an angle of incidence $\theta=90°$ to the shell axial direction. Both shells are fluid-loaded externally with an *in vacuo* condition assumed for the internal space. Monostatic scattering is considered.

Figure 1 shows the geometric characteristics of the multilayer shell and the material properties (steel and compliant layer) are given in Tables I and II and are the same for both models. The radius-to-thickness ratio of the base shell is 100 (a/h).

A. Bare shell

The scattered pressure is computed for selective circumferential values of n and as a sum of the contributions from the first 30 circumferential modes. The results are presented using the far-field form function, f_∞, defined to give a nondimensional representation of the scattered pressure[1–3]

$$|f_\infty| = \left(\frac{2r}{c}\right)^{1/2} \left|\frac{p_s}{p_0}\right|. \tag{26}$$

In the case of the normally reacting compliant layer model, the results for a bare shell are obtained by setting the impedance of the compliant layer to be very high by making the stiffness of the compliant layer infinite. That is the shell model with an external compliant layer simplifies to a bare shell. In the case of the multilayer model the two layers of the shell are set to be of the same material (steel) and each of half the thickness. The results are shown in Figs. 9 and 10,

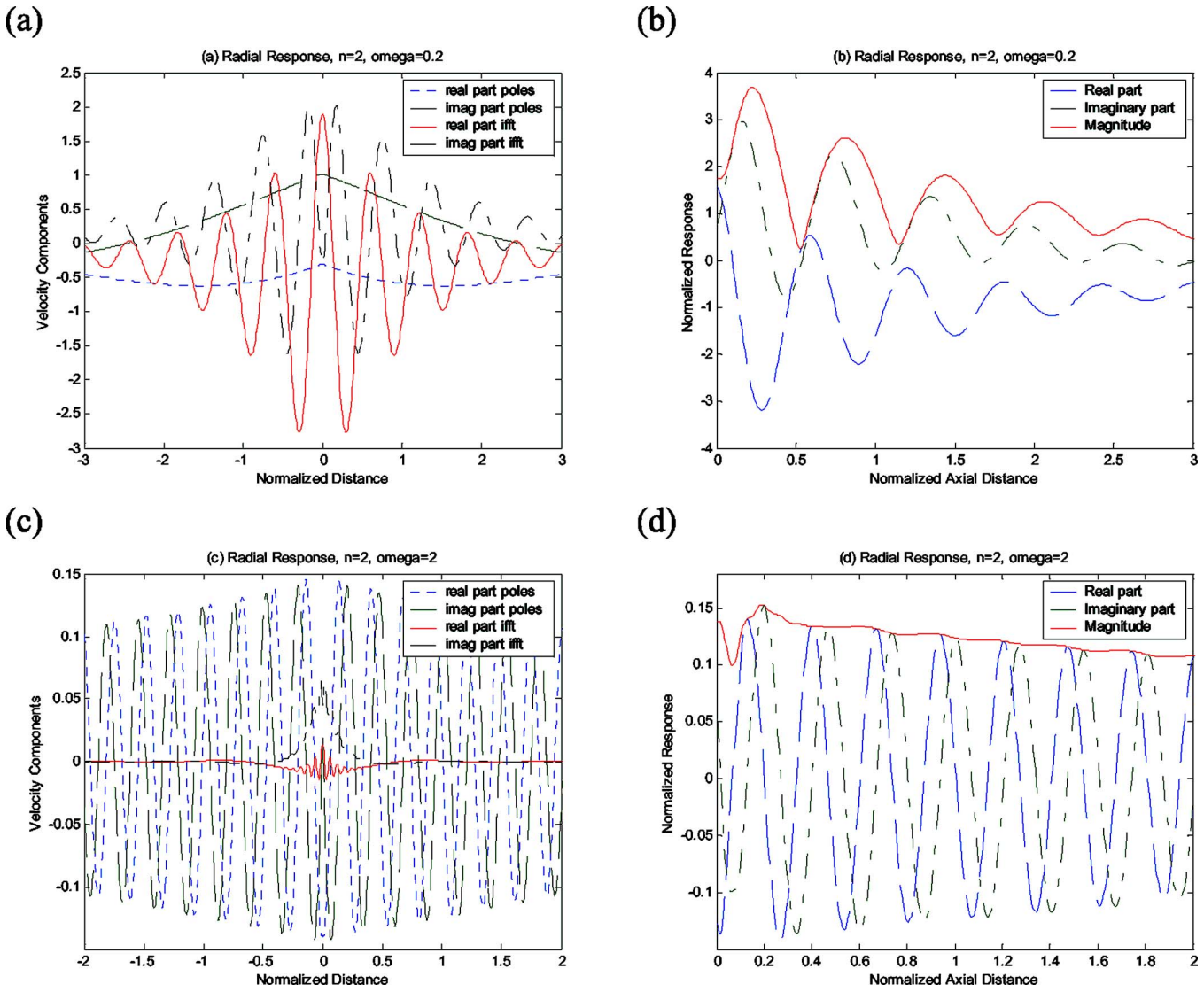


FIG. 18. (Color online)Normalized with respect to (γw_r). Radial response at the surface of the compliant layer for radial excitation for circumferential mode $n=2$. $\Omega=0.2$: (a) Contribution from IDFT and poles response, (b) magnitude, real, and imaginary parts. $\Omega=2.0$: (c) Contribution from IDFT and poles; (d) magnitude, real, and imaginary parts.

which show very good agreement between the two approaches.

B. Shell with external compliant layer

Using the parameters defined in Table II, results for the scattered pressure from the normally reacting layer model are obtained from the solution of Eq. (8)–(11). The multilayer model, with two dissimilar layers, one-representing the compliant layer requires some parameters for the compliant layer which are not needed in the normally reacting layer. One such parameter is the thickness of the compliant layer. Since this thickness can be varied, different values for the thickness of the compliant layer are considered. In this case as well results for selective individual circumferential mode numbers and for the total scattered pressure are presented using the far-field form function, f_∞. Figures 11 and 12 show the scattering results for the shells with the external compliant layer.

As can be observed from Figs. 9–12, the scattered pressure results for the individual circumferential modes match very well between the shell with the normally reacting layer and the multilayer shell. For the total scattered pressure, the results generally match except that thickness resonances associated with the compliant layer are absent in the normally reacting impedance layer model. This can possibly be modified by assuming the normally reacting compliant layer to have both a stiffness and mass impedance.

V. RESPONSE GREEN'S FUNCTIONS—NORMALLY REACTING LAYER APPROACH

Having validated using a normally reacting layer approach to model the behavior of a shell with an external compliant layer, results for the shell and compliant layer response Green's function and the far-field radiated pressure are computed. The compliant layer extends over the whole surface of the shell. For these results the shell is excited by

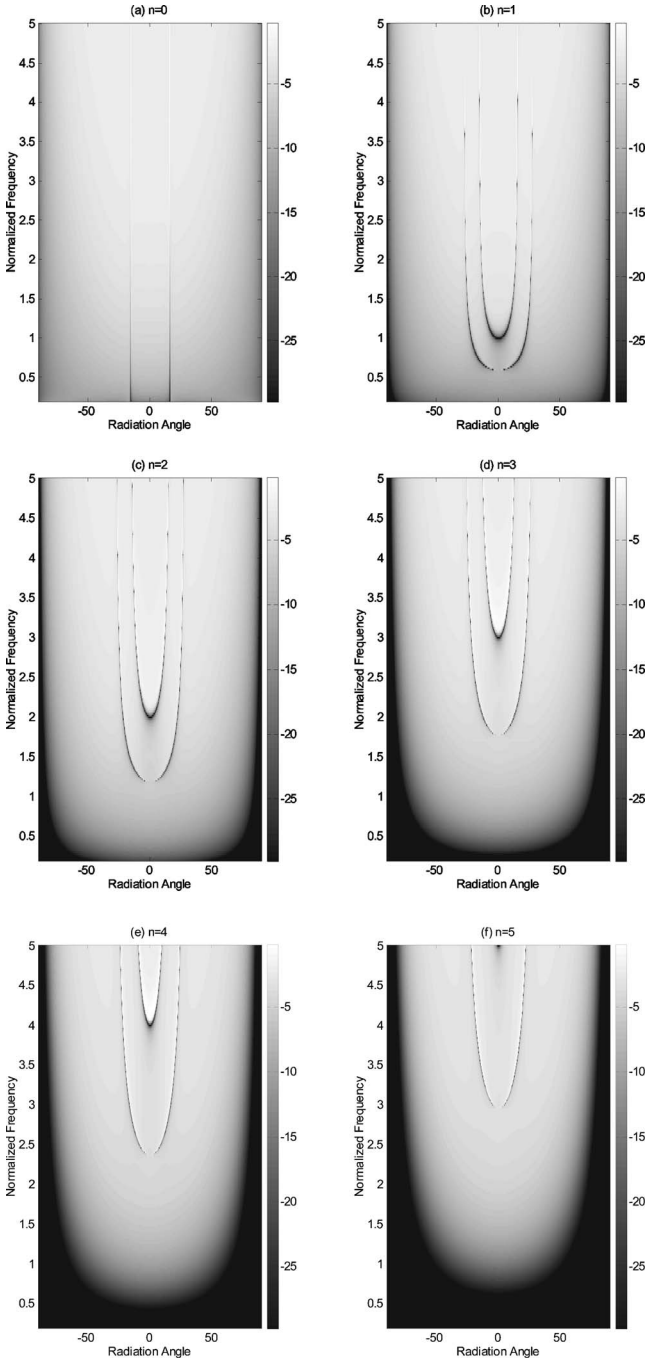


FIG. 19. Far-field radiated pressure for bare shell, (a) $n=0$, (b) $n=1$, (c) $n=2$, (d) $n=3$, (e) $n=4$, (f) $n=5$.

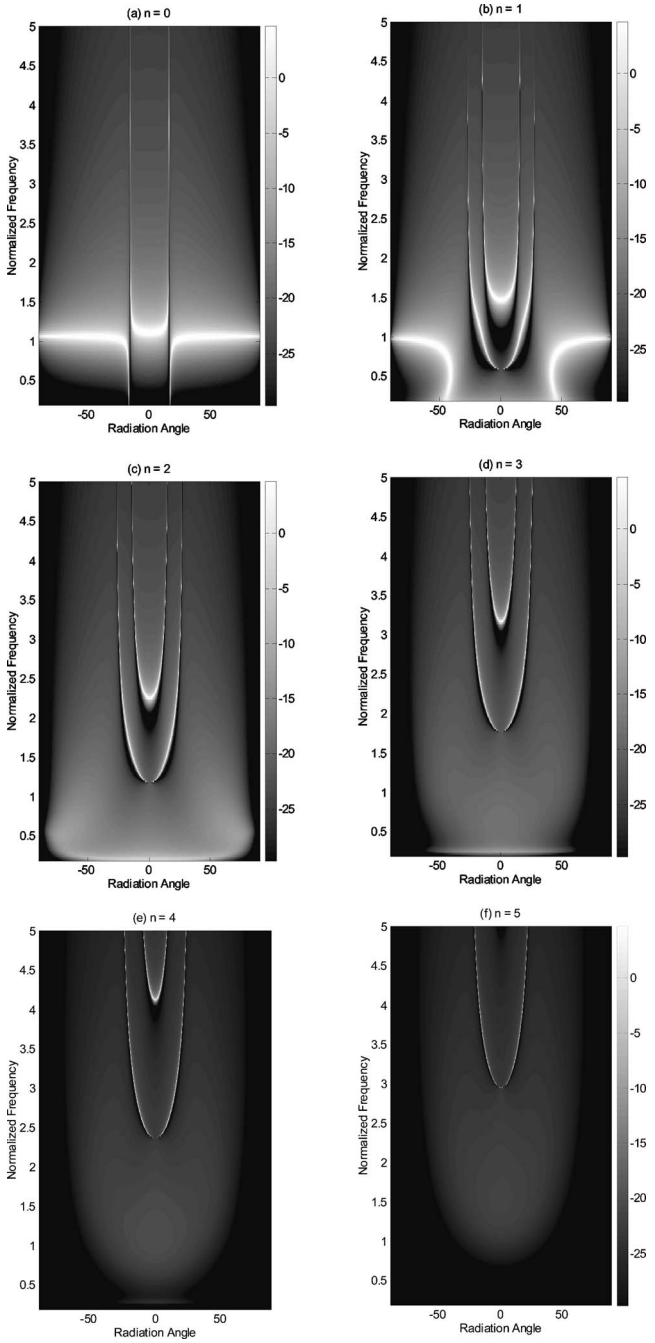


FIG. 20. Far-field radiated pressure for shell with an external compliant layer: (a) $n=0$; (b) $n=1$, (c) $n=2$, (d) $n=3$, (e) $n=4$, (f) $n=5$.

an internal radial ring force at $X=0$. The properties of the shell and the compliant layer are the same as in the previous section (Table II).

Using Eq. (19) and (16) with $\tilde{P}_i^n=0$, $\tilde{F}_{U_n}=\tilde{F}_{V_n}=0$ and $\tilde{F}_{W_n}\neq 0$, results for the response Green's functions due to an internal ring radial load are obtained and presented for two circumferential modes $n=0$ (breathing mode) and $n=2$. The mode $n=2$ is selected since it is representative of other circumferential modes greater than 0 and 1. For each mode the results are presented for two normalized frequencies, $\Omega=0.2$ and $\Omega=2$.

A. Shell surface response

With an external radial force excitation, the real part of the radial velocity component at the location of the excitation is positive. This is necessary for all frequencies and all circumferential mode numbers, since the real part of the velocity response at the point of excitation is a measure of the input power into the shell. The axial component of the response at the location where the internal force is applied is zero due the asymmetric behavior of the shell. For all conditions, the resonant fast-decaying waves (contribution seen in the IDFT) are significant only close to $X=0$. The oscillations in the response close to $X=0$, are an indication of energy exchange between the shell and the fluid.

The response of the shell for circumferential node 0 and $\Omega=0.2$ is shown in Fig. 13. It can be observed that the radial component of the surface velocity of the shell is dominated by the contribution from the IDFT, whereas the axial component of the response is dominated by both the contribution from the IDFT and the axial longitudinal waves. The magnitude of the axial response asymptotes to almost a constant value away from the excitation. The axial component of the response is equal to zero at the location where the force is applied, whereas the radial component is the highest at the excitation location.

Figure 14 shows the shell response for circumferential mode $n=0$ but for $\Omega=2$. In this case the radial response is dominated by the quasi-flexural waves, which are characterized by the short wavelength. One can clearly see the contribution from both the quasi-flexural (short wavelength) and the axial longitudinal (longer wavelength) waves in the axial response. The axial longitudinal waves have short wavelength and are attenuated with the distance, due to energy loss in acoustic radiation and/or damping of the compliant layer.

The shell response for circumferential mode $n=2$ and frequencies $\Omega=0.2$ and $\Omega=2$ are, respectively, shown in Figs. 15 and 16. For $\Omega=0.2$, the frequency is below the cut-on frequency of the axial longitudinal and torsional waves, which means that only the quasi-flexural wave can propagate. The response of the surface of the shell is dominated by the quasi-flexural waves. When $\Omega=2$, the frequency is above the cut-on frequency of the tangential waves, but below the cut-on frequency of the axial longitudinal waves. The contribution from the quasi-flexural waves (long wavelength), as well as from the tangential shear waves (short wavelength) can be clearly observed in the response. The axial and tangential responses are influenced by both the quasi-flexural and the tangential shear waves, whereas the axial response is influenced only by the quasi-flexural waves.

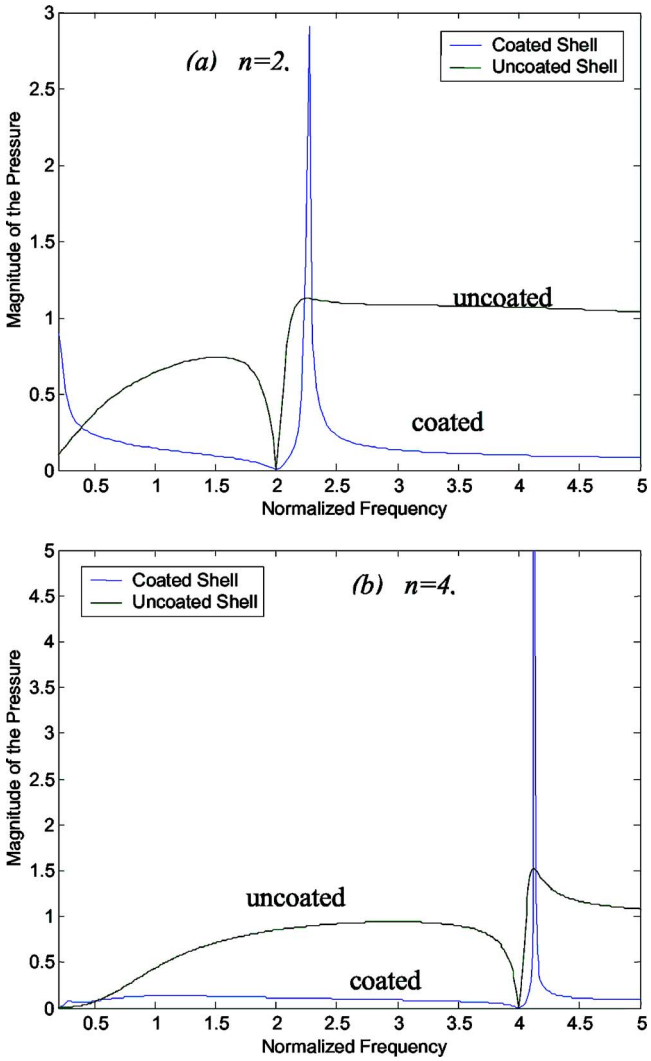


FIG. 21. (Color online) Comparison of the radiated far-field pressure between a bare shell and a shell with an external compliant layer, for $\psi=0$, and (a) $n=2$, (b) $n=4$.

B. Compliant layer surface response

The radial velocity component on the surface of the compliant layer is responsible for the radiated pressure by the cylinder. The radial velocity response at the surface of the compliant layer for circumferential mode $n=0$ and for frequencies $\Omega=0.2$ and $\Omega=2$ are presented in Fig. 17. The compliant layer response for circumferential mode $n=2$ and for frequencies $\Omega=0.2$ and $\Omega=2$ are presented in Fig. 18.

The first observation that can be made is that the magnitude of response at the surface of the compliant layer is generally lower than that on the surface of the shell. This is expected as the compliance of the layer acts as a soft decoupling layer. There is one exception and this occurs for circumferential mode $n=0$ and for $\Omega=0.2$. For this circumferential mode and frequency combination, the compliant layer exhibits a form of resonant behavior, effectively coupling the shell to the outside medium. Furthermore, upon comparing the velocity at the surface of the shell and at the surface of the compliant layer, the following observations can be made: (a) The contribution from the IDFT component of the solution is more significant at the surface of the compliant layer, (b) the oscillations in the IDFT solution (oscillations with a short wavelength close to $X=0$) are less attenuated with distance, and (c) the overall response extends over a larger section of the shell.

VI. RADIATED PRESSURE—NORMALLY REACTING LAYER APPROACH

Presented here are results for the radiation condition, that is with $\tilde{P}_i^n$ set to zero, however, scattering results can be generated by simply changing the input excitation. The far-field radiated pressure due to the excitation by an internal radial force is given by Eq. (24). The radiated pressure is normalized with respect to the radial distance Z. Results for the radiated pressure are evaluated for radiation angles $(-\pi/2<\psi<\pi/2)$, and for different frequencies, $\Omega=0-5$. The radiated pressure results are shown as contour plots with the x-axis representing the radiation angle, and the y-axis is representing the normalized frequency Ω. The radiated pressure results for the bare shell are presented in Fig. 19, while those for the coated shell are presented in Fig. 20. The pres-

sure field in the case of the bare shell is computed by setting the impedance of the compliant layer to be very high $\bar{Z}^c \rightarrow \infty$.

The radiated pressure for circumferential modes $n=0-5$ are shown. Mode $n=0$ is characterized by the absence of the supersonic tangential shear waves which are not excited. The radiation angle $\psi_c = \pm 16°$ corresponds to the critical angle of the longitudinal in-plane waves, it is constant and present for all frequencies. This angle is defined as the angle for which the velocity of the axial longitudinal wave in the shell match the acoustic wave velocity. Because of this coincidence condition, at $\psi_c = \pm 16°$, the axial longitudinal waves present a strong radiation. A lower pressure level can also be observed at an angle very close to ψ_c. This is attributed to destructive interference between the pressure field generated by the resonant fast-decaying near-field waves and the axial longitudinal waves on the shell.

The high-pressure band at $\Omega = 1.1$, correspond to the ring frequency of the fluid-loaded shell with the external compliant layer (at this frequency radial hooplike resonance occurs in the shell).

The far-field pressure pattern for a bare cylinder for circumferential mode 0 is similar to that of a cylinder with an external compliant layer. However, the ring frequency of the system is not as distinct and the radiation along the beaming angles associated with the axial longitudinal waves have a different character from those of the shell with the external compliant layer. This could be due to the much higher contribution from the radial component which has the dipolelike pattern.

For higher order circumferential modes ($n>0$), the supersonic tangential shear waves are present and contribute to the far-field pressure, and the axial longitudinal and the tangential shear waves both have a cut-on frequency. Below this frequency these waves are evanescent waves and they do not contribute to the far-field pressure.

Each supersonic component radiates along a preferred beaming angle. Two coincidence angles can be observed, one associated with the axial longitudinal waves and one associated with the tangential shear waves. These beaming coincidence angles start at the cut-on frequency of the respective waves and vary as a function of frequency. The cut-on frequency can be obtained from the dispersion curves (Figs. 3–8).

The beaming directions are symmetric about the shell axis normal and form a "wine glass" shape. The angle for each wave as a function of frequency is defined by:

$$\psi_{AL}(\Omega,n) = \sin^{-1}(k_{AL}(\Omega,n)), \tag{27}$$

$$\psi_{TS}(\Omega,n) = \sin^{-1}(k_{TS}(\Omega,n)), \tag{28}$$

where ψ_{AL} and ψ_{TS} are the beaming angle for the axial longitudinal and tangential shear waves respectively, $\bar{k}_{AL}$ and $\bar{k}_{TS}$ are the real part of the shell axial longitudinal and tangential shear wavenumbers.

While beaming of acoustic energy is strong for both wave types, it can be observed that the radiation from the axial longitudinal waves is stronger than the radiation from the tangential shear waves. This is due to the fact that the amplitude of the out of plane motion produced by the axial longitudinal wave is larger that the one produced by the tangential shear waves.

Away from the coincidence angles the radiated pressure field main contributions come from the fast-decaying waves. In between the coincidence angles of the tangential shear waves and the axial longitudinal waves, above their respective cut-on frequencies, there are two low pressure lines. These are attributed to destructive interaction of the acoustic pressure field produced by the tangential shear waves, the axial longitudinal waves and the resonant fast-decaying near-field wave on the shell. At low frequencies the radiated pattern is dipole-like due to the radial applied force.

Figure 21 shows a comparison of the radiated pressure by a bare shell and a shell with an external compliant layer, for a radiation angle $\psi = 0°$ and for circumferential modes 2 and 4. It can be seen from this figure that for the bare shell case, the beaming angles are not as distinct compared to the rest of the radiated pressure level. This is due to the fluid-loading effect. The fluid in contact with the shell attenuates the radiated power of the axial longitudinal and the tangential shear waves.

The far-field acoustic pressure generated by the radial component of the velocity field at the surface of the compliant layer does not show contributions from all the different waves present in the radial velocity field. While the radial velocity response is dominated by the quasi-flexural waves, these waves are subsonic and, therefore, do not contribute to the acoustic far-field pressure. Furthermore, the two waves introduced by the compliant layer are also subsonic and thus these waves do not have a significant influence on the far-field radiated pressure.

The resonant, fast-decaying near-field waves, present on the shell in the vicinity of the radial applied force produce an acoustic radiation in the far-field with a pattern similar to that of a dipole. This acoustic radiation is mostly noticeable at frequencies below the ring frequency of the fluid-loaded shell with the external compliant layer. This dipole radiation pattern is associated with the radial force as if the shell is not present.

As expected the far-field acoustic pressure is higher for the bare shell as compared to the shell with an external compliant layer. Depending on the circumferential mode number, the frequency, and the radiation angle, the general level can be between 15 and 20 dB higher than the case of a bare shell for the type of compliant layer considered here. The difference is greater at high frequencies. Figure 22 shows difference in dB of the radiated pressure between the bare shell and the shell with an external compliant layer for some select frequencies. As indicated, the compliant layer reduces the radiated power level considerably over the entirely frequency range except at the coincidence beaming angles.

VII. CONCLUSION

This paper presents results for the radiated pressure by an infinite cylindrical shell entirely covered with a compliant layer and immersed in a fluid. The compliant layer is mod-

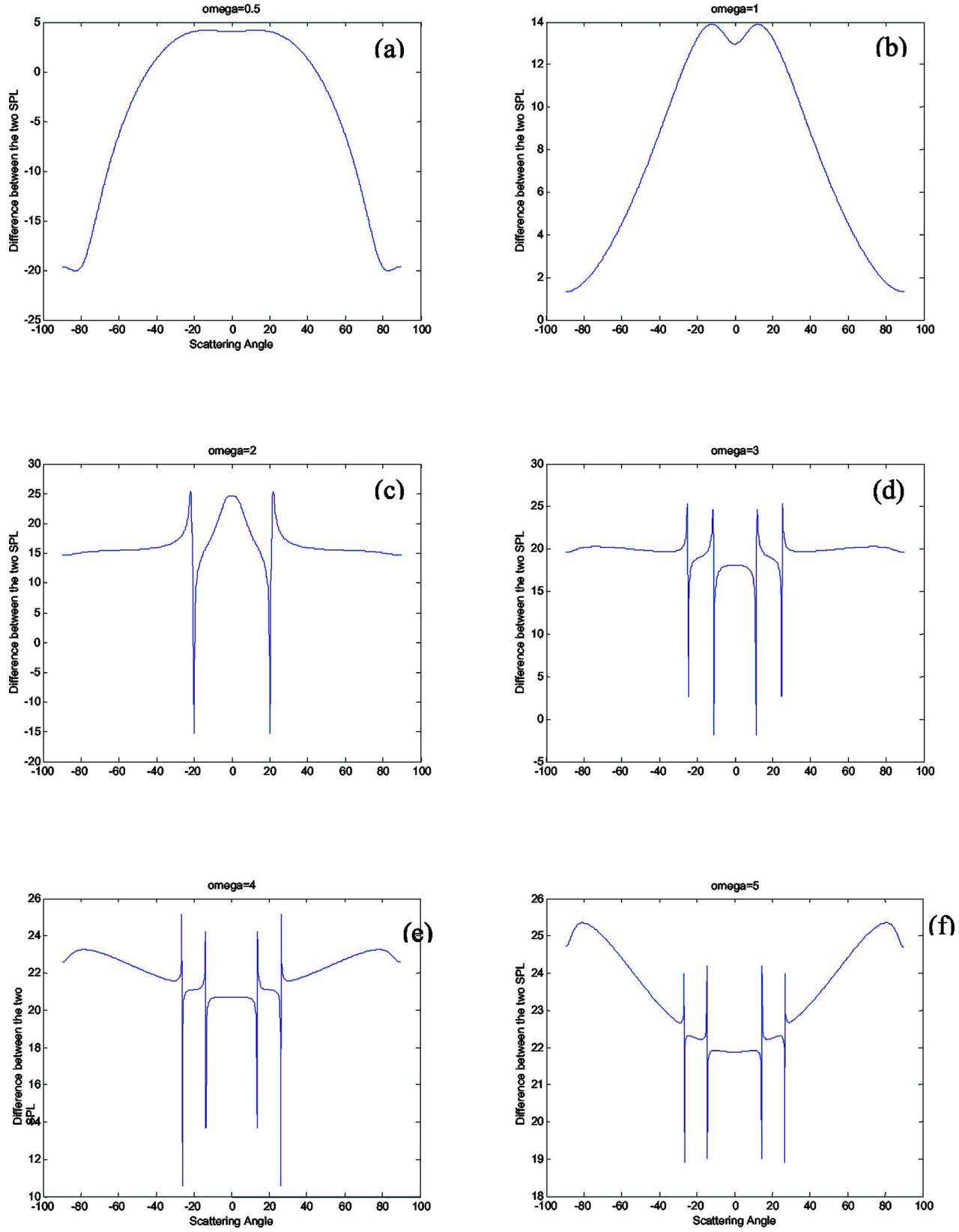


FIG. 22. (Color online) Difference in the radiated pressure levels in dB between a bare shell and a shell with an external compliant layer for $n=2$ at select frequencies: (a) $\Omega=0.5$; (b) $\Omega=1$; (c) $\Omega=2$; (d) $\Omega=3$; (e) $\Omega=4$; (f) $\Omega=5$.

eled by a normally reacting impedance layer of zero thickness. The first part of the paper demonstrates that the use of the normally reacting layer to model a compliant layer is justified. The scattering results from a shell with an external compliant layer modeled by a normally reacting impedance layer are compared to scattering results from a multilayer shell with the external layer representing the compliant layer. The results of the comparison analysis show that for both a

bare shell and a shell with an external compliant layer, the scattering results obtained from a thin shell model with a normally reacting compliant layer and a multilayer thick shell model show very good agreement, for a shell wall thickness to radius ratio of 0.01. For the shell with the external compliant layer, while the agreement in the results for the scattered pressure is still good, resonance peaks in the scattering associated with thickness resonances of the compliant layer are not obtained when the compliant layer is modeled by a normally reacting layer of zero thickness. The advantage of modeling the compliant layer by a normally reacting impedance layer is that the formulation can be extended to complex layer configurations. This is the subject of work in progress.

From the results for the shell and compliant layer surface velocity response Green's function and for the radiated pressure for the case when the shell is excited by a radial ring force, the following conclusions can be derived.

- At low frequencies the shell surface velocity response Green's function is dominated by the contributions from the resonant fast-decaying waves, the compliant layer waves, and by the quasi-flexural waves. For high frequencies the shell radial velocity response Green's function is dominated by the quasi-flexural waves. The shell axial and tangential velocity response Green's functions are dominated by the axial longitudinal and the tangential shear waves, respectively. The velocity response Green's function at the surface of the compliant layer is as expected, generally lower in magnitude than the shell surface velocity.
- At low frequencies, the radiated far field pressure is dominated by the subsonic resonant fast-decaying waves and the direct contribution of the radial internal force creating a dipole like radiation pattern. At higher frequencies, the radiated pressure is influenced by the supersonic axial longitudinal and torsional shear waves. For circumferential modes greater than 0 the far-field pressure has two predominant directions of radiation, at angles corresponding to the coincidence angle of the axial longitudinal and the tangential shear waves. These coincidence conditions create acoustic beaming for frequencies above the cut-on which is the cause of the "wine glass" pattern in the radiation pressure as a function of radiation angle and frequency.
- The radiated pressure results show that when a compliant layer is present on the shell, the radiated pressure level is considerably reduced. The reduction in the radiated pressure depends on the properties of the compliant layer (stiffness and damping), as it compares to the acoustic characteristics of the surrounding fluid.

ACKNOWLEDGMENTS

The author would like to acknowledge the financial support of ONR Code 33 without which this work would not have been possible. Also the author would like to acknowledge the contributions of Stephen Schreppler ONR Program Manager and David Feit of NSWCCD. Finally the author would like to acknowledge the preliminary contributions of Carrol Treffot who did her master's thesis research at Florida Atlantic University on the topic of this paper under the supervision of the author.

[1]R. D. Doolittle and H. Uberall, "Sound scattering by elastic cylindrical shells," J. Acoust. Soc. Am. **39**, 272–275 (1966).
[2]G. C. Gaunaurd, "Sonar cross section of a coated hollow cylinder in water," J. Acoust. Soc. Am. **61**, 360–368 (1977).
[3]L. Flax and W. G. Neubauer, "Acoustic reflection from layered elastic absorptive cylinders," J. Acoust. Soc. Am. **61**, 307–312 (1977).
[4]J. M. Cuschieri and D. Feit, "Influence of circumferential partial coating on the acoustic radiation from a fluid-loaded shell," J. Acoust. Soc. Am. **107**(6), 3196–3207 (2000).
[5]M. C. Junger and D. Feit, *Sound, Structures and their Interaction* (MIT Press, Cambridge, MA, 1995).
[6]J. M. Cuschieri, and D. Feit, "A hybrid solution for the response green's function of a fluid-loaded cylindrical shell," J. Acoust. Soc. Am. **96**, 2776–2784 (1994).
[7]B. Laulagnet, and J. L. Guyader, "Sound radiation from finite cylindrical shells, partially covered with longitudinal strips of compliant layer," J. Sound Vib. **186**, 723–742 (1995).
[8]J. M. Cuschieri, and D. Feit, "A hybrid numerical and analytical solution for the green's function of a fluid-loaded elastic plate," J. Acoust. Soc. Am. **95**, 1998–2005 (1994).

Vibro-acoustic control with a distributed sensor network

Kenneth D. Frampton[a)]
Department of Mechanical Engineering, Vanderbilt University, VU Station 351592, Nashville, Tennessee 37235-1592

(Received 12 May 2005; revised 10 January 2006; accepted 28 January 2006)

The purpose of this work is to demonstrate the ability of a distributed control system, based on a smart sensor network, to reduce acoustic radiation from a vibrating structure. The platform from which control is effected consists of a network of smart sensors, each referred to as a node. Each node possesses its own computational capability, sensor, actuator and the ability to communicate with other nodes via a wired or wireless network. The primary focus of this work is to employ existing group management middleware concepts to enable vibro-acoustic control with such a distributed network. Group management middleware is distributed software that provides for the establishment and maintenance of groups of distributed nodes and that provides for the network communication among such groups. The control objective is met by designing distributed feedback compensators that take advantage of node groups in order to effect their control. The node groups are formed based on physical proximity. The global control objective is to minimize the radiated sound power from a rectangular plate. Results of this investigation demonstrate that such a distributed control system can achieve attenuations comparable to those achieved by a centralized controller. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2178704]

PACS number(s): 43.40.Vn, 43.50.Ki [KAC] Pages: 2170–2177

I. INTRODUCTION

Active control of structural acoustic radiation has been investigated for many years and a great deal of success has been demonstrated. However, the extension of these concepts to large-scale systems has not been very successful. This lack of success is, in part, due to the overwhelming need for computational effort, data transmission, and electrical power. This need has been overwhelming in the sense that the potential applications are unable to bear the power, weight, and complex communications requirement of large-scale, centralized control systems. Recent developments in MEMS transducers, combined with networked embedded systems technologies,[1,2] have changed the focus of such applications from centralized control architectures to distributed architectures: i.e., control systems based on numerous embedded processors communicating over a network.

The technologies of active sound and vibration control have grown quickly in the past decade. The segment of this work relevant to the current investigation is the special case of active structural acoustic control (ASAC). ASAC differs from vibration control in that the goal is to use structural actuators to reduce the sound radiation instead of the structural response. This difference is particularly important below the plate coincidence frequency due to significant interaction between the modes in the radiated field.[3]

One of the earliest works in the field of active structural acoustic control was published by Jones and Fuller.[3] This research considered the control of narrow-band acoustic radiation through the use of structural actuators. Considerable attenuations in the transmitted noise were achieved through this approach. This work was expanded upon by Clark and Fuller[4–6] among others with similar results. However, each of these investigations employed feed-forward control systems. Feed-forward control is very effective when a correlated reference signal is available. However, for stochastic inputs, such as turbulent boundary layer noise, feedback control is the most effective.

The primary contribution of this work is to investigate the use of distributed control for active structural acoustic control. A distributed control system is one that consists of many localized controllers, called nodes, working cooperatively to achieve some desired performance. A distributed control system is defined here as one that actuates the system locally, but with a control law that depends on the sensor data of other nodes in the system. This is in contrast to decentralized control in which many local compensators act independently of each other. The field of distributed control has been investigated for many years.[7] The majority of these published works dealt with "weakly connected" systems in which the dynamics at each node depends on only a few system degrees of freedom while being weakly coupled to other system subcomponents. A good example of such systems is robotic swarms.[8–10] One of the novel aspects of the work reported here is that it considers distributed control of a system that is strongly coupled. This is because the dynamics experienced by any individual sensor is a function of all systems states. Recent work dealing with distributed control of strongly connected systems has been published by De' Andrea, although the configuration of the subsystem interdependencies is limited.[11] Distributed control has also been considered for control of space structure vibrations[12–14] and buckling in beams.[15]

More recently, several investigations have focused on decentralized control.[16–18] Decentralized control differs from distributed control in that with decentralized control there is

[a)]Electronic mail: ken.frampton@vanderbilt.edu

 0001-4966/2006/119(4)/2170/8/$22.50

no cooperation between system nodes. They act entirely independently. The recent work of Elliot *et al.*, which dealt with decentralized active structural acoustic control, is particularly relevant to this manuscript.[19] In this work, numerous independent rate feedback controllers were employed to minimize acoustic radiation from a vibrating plate. Both point transduction and smart-material transduction were considered. The results demonstrated the ability of decentralized controllers in this context. However, it seems plausible that, by employing networked embedded systems technology, some degree of internode cooperation could easily be incorporated into such a control system, resulting in improved performance while retaining the advantages of a decentralized architecture. The use of such a cooperative, distributed platform for active structural acoustic control is the focus of this work.

The distributed hardware platform that is assumed in this investigation is one where each node possesses the following: a microprocessor of limited computational capability and memory; a sensor and actuator along with the necessary signal conditioning hardware; and a network communications link with other nodes that is of limited bandwidth. The objective of this distributed controller is the same as a centralized control system: to minimize the sound power radiated from a vibrating plate in the presence of disturbances. However, for the purposes of this investigation, the distributed controllers must achieve their performance while operating within the constraints of a networked embedded system. Working within these memory, computation, and communications constraints means that the control law must be relatively simple and the information that is communicated between nodes must be limited. This will ensure that the final system not only fits within the constraints of the current investigation, but that it is scalable and therefore may be applied to even larger systems. To meet the first constraint the control law employed is constant gain output feedback. Such controllers have been found to be quite effective at attenuating broadband disturbances, and are relatively robust to system uncertainty, particularly when the transducers are collocated and dual.[20] The second constraint is met by limiting the sensor information available to each node to a prescribed subset of the total sensor information in the system.

The work presented here consists of the computational investigation of distributed active structural acoustic control in a simply supported plate. Two cases are considered in which the plate is equipped with numerous point-force/point-velocity transducer pairs as well as piezoelectric transducers. Distributed feedback compensators are designed which interact with each other by sharing sensor information and whose goal is to minimize the radiated sound power. Sensor information is shared among groups of nodes which are defined by their physical proximity to each other. The performance of these group-based distributed control approaches is evaluated by comparing their performance with a centralized control system that uses the same sensors and actuators and that expends an equal amount of control energy. The discussion begins with a description of the physical system model including plate dynamics, transduction, and radiated acoustic energy. This is followed by a brief description of centralized control design and the motivation for distributed control approaches. Finally, distributed feedback control designs are described and results are presented which demonstrate the effectiveness of group-based ASAC with a distributed sensor network.

II. SYSTEM MODELING

The effectiveness of such a distributed approach to ASAC is demonstrated through computational results. The system considered is a simply supported plate embedded in an infinite baffle and exposed to an infinite acoustic medium on one side. The objective of the distributed controller is to minimize the radiated sound power when the plate is subjected to a broadband disturbance input. Modeling of the plate and acoustic radiation is described in this section along with a discussion on distributed feedback control design.

A. The plate model

The plant under consideration is a simply supported plate subject to a random disturbance. The plate is equipped with either piezoelectric transducers that act simultaneously as sensors and actuators[21] or point-force/point-velocity transduction. The plate modeling technique follows that presented by Hagood *et al.*,[22] which uses a Rayleigh-Ritz technique to create the equations of motion for the full electromechanical system. This representation includes the mass and stiffness contributions of the piezoelectric transducers as well as the dynamics of the transducers themselves. The solution assumes that the structural displacement can be expressed as a summation of orthogonal functions. Using the modes of a simply supported plate results in an expansion of the following form:

$$w(x,y,t) = \sum_{n=1}^{N} \psi_n(x,y) q_n(t) = \sum_{n=1}^{N} \sin\left(\frac{n\pi x}{a}\right)\sin\left(\frac{n\pi y}{b}\right) q_n(t), \tag{1}$$

where $w(x,t)$ is plate displacement, $\psi_n(x,y)$ is the nth mode shape, and $q_n(t)$ is the generalized coordinate of the nth mode. The result is a set of coupled ordinary differential equations of the form

$$[\mathbf{M}_p + \mathbf{M}_t]\{\ddot{\mathbf{q}}(t)\} + [\mathbf{K}_p + \mathbf{K}_t]\{\mathbf{q}(t)\} = \{\mathbf{Q}^d(t)\} + [\Theta]\{\mathbf{v}(t)\} + \{\mathbf{Q}^p(t)\}, \tag{2}$$

where $\mathbf{M}_p$ and $\mathbf{M}_t$ are the mass matrices for the plate and piezoceramic transducers, respectively, $\mathbf{K}_p$ and $\mathbf{K}_t$ are the stiffness matrices for are the plate and piezoceramic transducers, respectively, $\mathbf{q}_n$ is a vector of generalized coordinates, $\mathbf{Q}^d(t)$ is the vector of generalized disturbance forces. Control is affected either by point forces, represented by the generalized force vector $\mathbf{Q}^p(t)$, or by piezoelectric transducers through the electromechanical coupling matrix, Θ, which relates the applied control voltages, $\mathbf{v}$, to the model equations. The disturbance generalized force vector, $\mathbf{Q}^d(t)$, is related to the disturbance input, $F^d(t)$. The disturbance used in the simulations consisted of an independent random input to

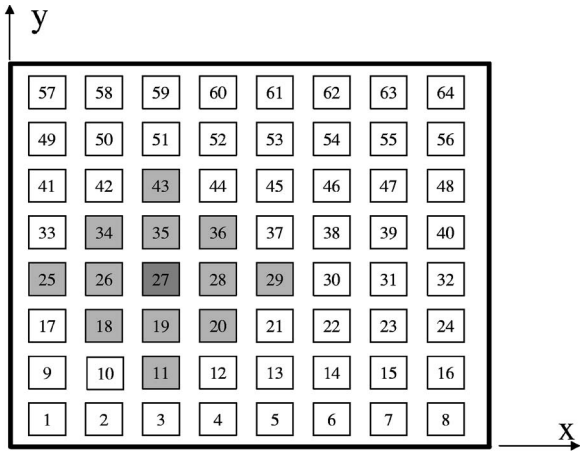


FIG. 1. Schematic of the plate and transducers.

each mode. Clearly, this disturbance is not representative of a realistic one such as turbulent boundary layer inputs (which would consist of independent "colored" random inputs to each mode). The disturbance represents a kind of "worst case" scenario.

The plate model can be cast in state variable form as follows:[23]

$$\dot{\mathbf{x}}_s = \mathbf{A}_s\mathbf{x}_s + \mathbf{B}_s\mathbf{u}_s,$$

$$\mathbf{y}_s = \mathbf{C}_s\mathbf{x}_s, \tag{3}$$

where $\mathbf{x}_s$ is the state vector containing the generalized coordinates, $q_n(t)$ and their derivatives, $\mathbf{u}_s$ is a vector of control and disturbance inputs, and $\mathbf{y}_s$ is the output from each sensor (either piezoceramic[22] or point velocity) such that

$$\mathbf{x}_s = \{\mathbf{q}\ \dot{\mathbf{q}}\}^T, \tag{4}$$

$$\mathbf{u}_s = \begin{Bmatrix} \mathbf{u}_d \\ \mathbf{u}_c \end{Bmatrix}^T = \begin{cases} \{\mathbf{F}^d(t)\ \mathbf{v}(t)\}^T & \text{for piezoelectric transduction} \\ \{\mathbf{F}^d(t)\ \mathbf{Q}^p(t)\}^T & \text{for point transduction} \end{cases} \tag{5}$$

$$\mathbf{y}_s = [\mathbf{0}\ \mathbf{T}^T]\mathbf{x}_s \text{ or } [\mathbf{0}\ ?], \tag{6}$$

where $\mathbf{u}_d$ corresponds to the disturbance input, $\mathbf{u}_c$ corresponds to the control input. In both piezoelectric and point transduction $\mathbf{F}^d(t)$ is a vector of disturbance generalized forces. In the case of piezoelectric transduction $\mathbf{v}(t)$ is the vector of control voltages and $\boldsymbol{\Theta}$ is the vector of electromechanical coupling terms. In the case of point-force/point-velocity transduction $\boldsymbol{\Psi}$ is the vector of mode shape functions evaluated at the points of transduction.

The specific arrangement utilized in this investigation consisted of 64 rectangular transducers uniformly arrayed on the plate as shown in Fig. 1. The plate is assumed to have x,y dimensions of 1.2 and 1 m, respectively, and to be 4 mm thick. The material properties used to develop the plate model were those for 2024-T4 aluminum. Each transducer was 40 by 50 mm and therefore all transducers covered about 10% of the total plate surface area. The physical parameters used to model the piezoceramic those for PZT type G-1195 and the transducers were assumed to be 0.2 mm thick. To ensure accuracy in the frequency range of interest, 40 modes [i.e., $N=40$ in Eq. (1)] were used in the plate model; however, results are shown only for the first 20 modes.

The author recognizes that constructing a plate with 64 transducers, processors, and the necessary power and network communications is not a very realistic approach to controlling acoustic radiation from a single plate. Clearly, such a networked control system would be more suitable for a large-scale system such as an aircraft fuselage. However, the work presented here is a first step in demonstrating the capabilities of distributed control and the promise that it has for large-scale systems.

B. The acoustic model

Modeling of the radiated sound power follows the approach taken by Bauman *et al.*,[24] as well as Currey and Cunefare.[25] These results are combined with those of Clark and Cox[26,23] in order to obtain a state variable representation suitable for control system design. This approach begins by expressing the pressure radiated to the far field with the Rayleigh integral such that

$$p(t) = \frac{j\omega\rho}{2\pi}\int_S \frac{\dot{w}(S,t)e^{jk(R-r_s)}}{(R-r_s)}dS, \tag{7}$$

where ω is the excitation frequency, ρ is the fluid density, $\dot{w}(S,t)$ is the plate velocity over the surface S, and R is the distance to the far-field point of interest. The far-field intensity can be calculated by integrating the power over a hemisphere in the far field and of radius R. Performing this integration, and adopting the notation of Bauman *et al.* results in[24]

$$\Pi = \frac{1}{\pi}\int_0^\infty \mathbf{V}^H(j\omega)\mathbf{M}(j\omega)\mathbf{V}(j\omega)d\omega, \tag{8}$$

where Π is the total radiated power, $\mathbf{V}(j\omega)$ is the matrix of transformed modal velocities (the transform of the time rate of change of the generalized coordinate vector $\{\mathbf{q}\}$) and $\mathbf{M}(j\omega)$ is the frequency-dependent matrix that defines the acoustic interaction among modes.[27] The diagonal entries of M are proportional to the modal radiation efficiencies and the off-diagonal terms pertain to the mutual radiation efficiencies of the structural modes. Bauman *et al.* described a factorization technique for partitioning this matrix such that[24]

$$\mathbf{M}(s) = \mathbf{G}^T(-s)\mathbf{G}(s), \tag{9}$$

where $G(s)$ is referred to as the radiation filter. These filters, when provided with the plate modal velocity amplitudes as inputs, generate a signal that is proportional to the far-field pressure. The square of this output is equal to the radiated power. There are numerous methods for generating a filter given its Laplace domain description. The approach used by Clark and Cox is used here to generate a state variable

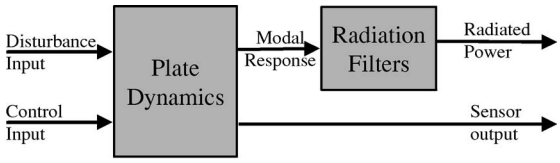


FIG. 2. Schematic of the coupled system model.

representation.[26] Such a state variable representation of the acoustic system has the form

$$\dot{\mathbf{x}}_a = \mathbf{A}_a\mathbf{x}_a + \mathbf{B}_a\mathbf{u}_a,$$

$$\mathbf{y}_a = \mathbf{C}_a\mathbf{x}_a, \tag{10}$$

where the input vector, $\mathbf{u}_a$, contains the modal velocities, the output, $\mathbf{y}_a$, is proportional to the far-field pressure, and the acoustic state vector, $\mathbf{x}_a$, is populated by variables that are of mathematical construct and have no specific physical relationship. The final full acoustic model had 200 states and was coupled with the first 20 plate modes.

C. The coupled model

Given the state variable models of the structure [Eq. (3)] and the acoustic radiation [Eq. (10)], a coupled structural acoustic model can be constructed as depicted in Fig. 2. As shown, the plate is subjected to both disturbance inputs and control inputs. The response of the plate drives the acoustic system model. The coupled system outputs are the structural sensor outputs as well as the response of the acoustic model. The sensor outputs are used by the feedback compensator to construct the control inputs. The acoustic system response is used as a design variable for the control system as described in the following section. The coupled system is subsequently referred to as $(\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D})$ with states $\mathbf{x}$, inputs $\mathbf{u}$, and outputs $\mathbf{y}$. In the simulations discussed later the full coupled system was comprised of 280 states. This model size demonstrated sufficient convergence of the results for the kinetic energy and radiated power from the first 20 plate modes. A similar convergence conclusion was reached by Clark and Cox.[26]

III. CENTRALIZED VIBRO-ACOUSTIC CONTROL

As described previously, a considerable amount of work has been done on centralized control of structural acoustic radiation. One such centralized control method is explored here for three reasons. First, the distributed control design used in this investigation builds on this centralized control approach. Second, the performance of the centralized controller serves as a basis from which to judge the performance of distributed control. And third, the form of the resulting centralized compensator provides very strong support for the idea that distributed control is a viable approach.

A. Design and performance of a centralized compensator

Since the envisioned distributed control system is comprised of numerous nodes of very limited resources, a control scheme requiring very little computational effort was selected: constant gain output feedback. For the centralized control case the control law implemented is

$$\mathbf{u}_c = -\mathbf{K}\mathbf{y}, \tag{11}$$

where $\mathbf{K}$ is the feedback gain matrix, $\mathbf{u}_c$ is the control input to the structure, and $\mathbf{y}$ is the sensor output vector. The feedback gain matrix can be found by minimizing the cost functional,[28]

$$J = \int_{t_0}^{t_\infty} [\mathbf{x}^T\mathbf{C}_r^T\mathbf{W}\mathbf{C}_r\mathbf{x} + \mathbf{u}_c^T\mathbf{R}\mathbf{u}_c]dt, \tag{12}$$

where J is the cost to be minimized, $\mathbf{x}$ is the coupled system state vector, $\mathbf{C}_r$ is that portion of the coupled system $\mathbf{C}$ matrix that provides the radiated sound power output, $\mathbf{W}$ is the penalty weight associated with the radiated power, and $\mathbf{R}$ is the control effort penalty. Details concerning the calculation of the feedback gain matrix that minimizes Eq. (12) can be found in Refs. 26 and 28.

The penalty matrices, $\mathbf{R}$ and $\mathbf{W}$, were both set equal to a constant (subsequently referred to as R and W), times the identity matrix. The objective of minimizing radiated sound power was accomplished by selecting a performance penalty weight, W, that resulted in a control system performance that one could reasonably expect to obtain in a laboratory setting.[21] While it is theoretically possible to increase the performance penalty, resulting in significantly larger control signals and attenuations, practical limitations would prevent such improvements. This value of W was then used for all subsequent centralized and distributed control cases. However, the control effort penalty, R, was changed in each case so that the system H_2-norm between the disturbance input and all control outputs was the same. This was done in order to ensure a fair comparison basis for different control architectures; i.e., all controllers expend an equal amount of control energy to achieve their performance.

The effectiveness of such centralized controllers has been well documented in the past[23,26] and is shown here in Fig. 3. Figure 3 depicts the sound power ratio as a function of frequency with and without feedback control. This is the ratio, in decibels, between the radiated sound power and the

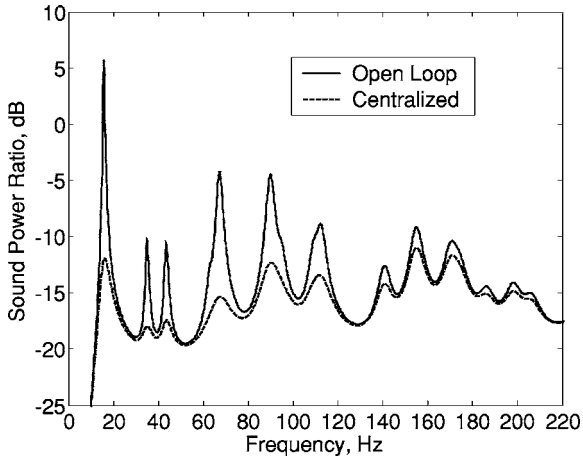


FIG. 3. Sound-power ratio for the uncontrolled system and the centralized compensator using point transduction.

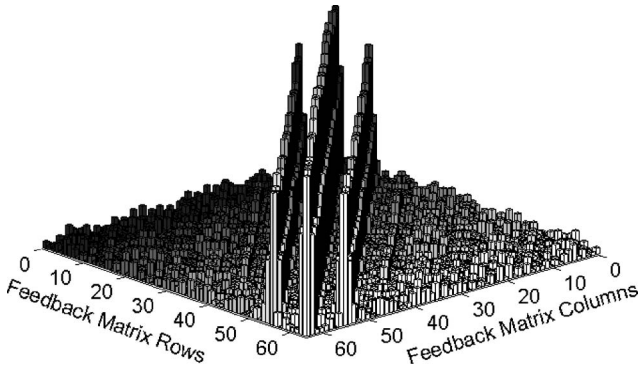


FIG. 4. Magnitude of the centralized feedback matrix elements.

disturbance input power. In this case point-force actuation and point-velocity sensing were used to achieve control. As shown in Fig. 3, the uncontrolled system response clearly shows the modal influences of the radiated sound power. The dominant radiator, as expected, is the (1,1) mode located at 17 Hz, while the next biggest contributors to the sound power are the (1,3) and (3,1) modes located at 67 and 91 Hz, respectively. Also demonstrated in Fig. 3 is the effectiveness of the centralized compensator attenuating the radiated sound power. These results compare very well with those reported by Clark and Cox[26] which used a nearly identical control approach.

The results of Fig. 3 were not unexpected, but the structure of the feedback gain matrix was surprising. The magnitudes of all elements of the centralized feedback gain matrix, **K**, are plotted in Fig. 4. Note that the structure of this matrix is block diagonal. In other words, the feedback gains away from the main diagonal are very small compared to those near the main diagonal. What this means is that each individual control signal is approximately only a function of a small subset of the total sensor signals available. Even more surprising is the fact that each individual control signal is dominated by the sensor signals in its immediate vicinity. This is shown more clearly in Fig. 5, which shows the feedback gains used to create the control signals for actuators 11 and 36 mapped onto the actuator grid of Fig. 1. So, Fig. 5(a) shows the gain applied to each of the 64 sensors and summed to form the control signal sent to actuator 36. Note that the control signal for actuator 36 is dominated by its own sensor (the largest amplitude gain comes from collocated sensor 36). It also depends strongly on the signals from those sensors immediately adjacent to it. However, the gains applied to sensor signals further away are very small in comparison. This "clustering" of gains was also noted for various effort penalties and for piezoelectric transduction. Therefore, one might infer from this that a good approximation to this optimal centralized controller would be to construct the control signal for each actuator only from those sensors that are "nearby." This result is the motivation for the distributed control architecture used here and described in the next section.

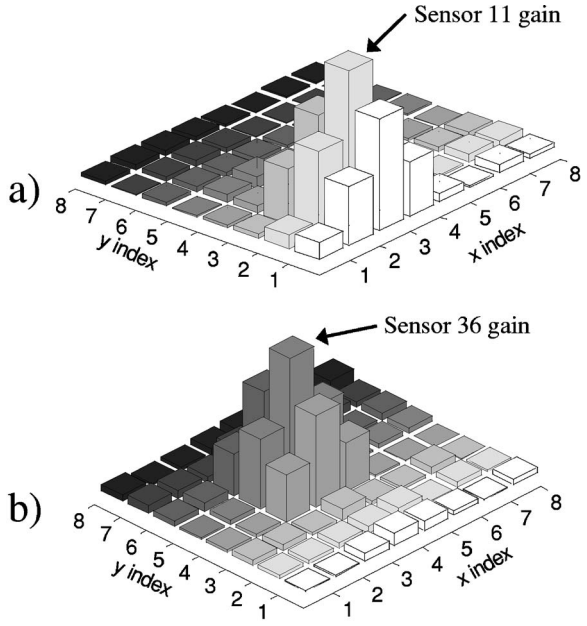


FIG. 5. Magnitude of the centralized feedback gain matrix mapped onto the plate transducer array: (a) control actuator 11; (b) control actuator 36.

IV. DISTRIBUTED CONTROL SYSTEM DESIGN AND RESULTS

Motivated by the structure of the centralized controller discussed previously, a distributed control architecture is considered based on groups of nodes in close physical proximity. In this manner a control architecture is developed that performs nearly as well as the centralized approach, while also being scalable. That is, the distributed control architecture based on groups could conceivably be deployed in very large-scale systems while staying within communications constraint of such a system. This distributed control concept is described in the following sections along with simulation results which demonstrate its capabilities.

A. Designing the distributed control system

As stated previously, the control architecture utilized in this investigation is based on groups. A group is defined as a collection of nodes that are associated with each other by the exchange of sensor signals. All nodes within a group receive the instantaneous sensor signals of all other members in the group by network communication. Then, each node calculates its own local control signal as a function of its own sensor signal and all sensor signals received from fellow group members. Therefore, each local control algorithm is a multi-input, single-output (MISO) control system. This is shown schematically in Fig. 6. The performance of various size groups will be compared; however, the size of each group is set such that it is a fraction of the total number of nodes in the system. Since the number of members in each group is a fraction of the total number of nodes, the size and complexity of each node's control law will not grow as the size of the system increases.

The grouping used here, motivated by the structure of the centralized compensator, is referred to as geographic grouping and is based on the physical proximity of nodes.

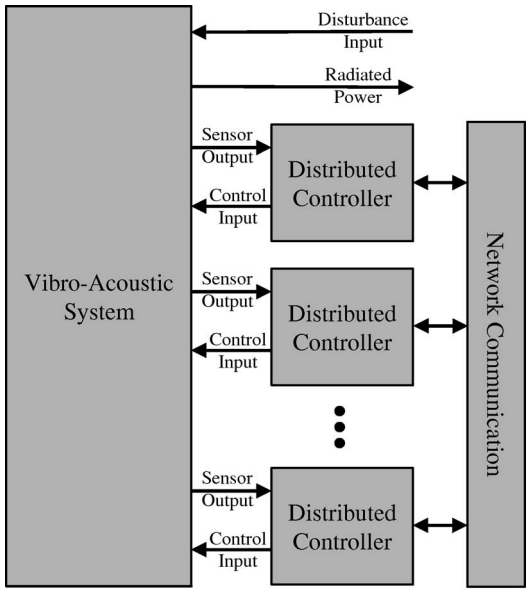


FIG. 6. Schematic of the distributed control system.

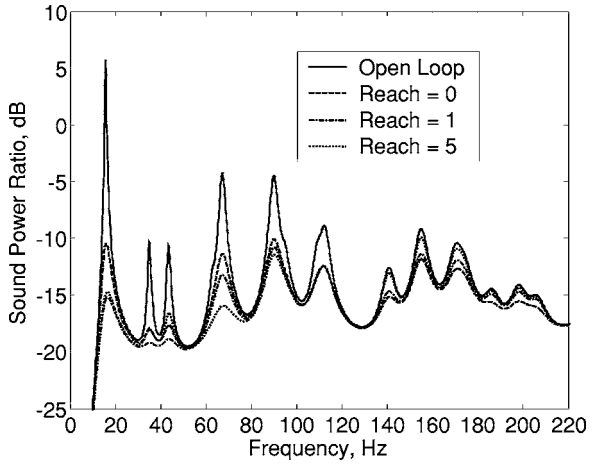


FIG. 7. Sound-power ratio for the uncontrolled system and distributed compensators of reach 0 (decentralized control), 1, and 5 using point transduction.

These groups are created by considering a specific node and including in its group all nodes within a certain number of "hops" from the node being considered. A hop is one step either in the x or y direction. The maximum number of hops is called the "reach" of the group, denoted by R. With this control architecture, a group is formed around every node in the system. Every node then has a controller that receives sensor signal inputs from every other node within its reach. (Note that this group is also the same set of nodes to which the specific node sends its sensor signal.) Figure 1 depicts an example of a geographic group centered on node 27 (shown in dark gray) and with $R=2$. The center node receives sensor signals from all other nodes within 2 hops (shown in light gray). Note, furthermore, that a group having a reach of zero would have only one node in it and the resulting control would be completely localized, with no interdependence on other nodes in the system. This is frequently referred to as decentralized control and it is the type of control that was used in Refs. 16–19. In this investigation, control architectures having reaches of 0 through 5 are considered.

If a distributed control system, such as that described here, were employed with wired communications it is very likely that the communication network would be a point-to-point or a peer-type network. In other words, it is likely that each node would have direct, wired communications with each of its neighboring nodes. This type of configuration would minimize the wires needed in comparison to a switched router network. Messages would be passed along from one node to the next in order to transmit across the network. Therefore, the concept of reach would be very closely related to the time necessary to convey messages and it would also be closely related to the total network traffic volume. On the other hand, if a wireless network were employed, then each node would have a particular radius in which it could transmit and receive. This radius would also be closely related to the reach used to define geographic architectures.

The distributed controllers are designed in the same manner as the centralized controller described previously. However, each distributed compensator is a MISO system with a control law of the form

$$u_i = -\mathbf{K}\mathbf{y}_i, \tag{13}$$

where $\mathbf{K}$ is the feedback gain vector, u_i is the (scalar) local control signal, and $\mathbf{y}_i$ is the subset of the sensor output vector associated with the group in question. The feedback gain matrix is found by minimizing the cost functional,

$$J_i = \int_{t_0}^{t_\infty} [\mathbf{x}^T \mathbf{C}_r^T \mathbf{W} \mathbf{C}_r \mathbf{x} + u_i^T R u_i] dt, \tag{14}$$

where J_i is the cost associated with the ith node compensator. Note that each distributed compensator is "locally" optimal. However, the overall system is not globally optimal.

The objective of each node's compensator is to minimize the radiated sound power. As described in the centralized case, the value of $\mathbf{W}$ was held constant. The compensator for each node was designed independently based on the open-loop plant and employing the method outlined previously. Once each node was designed, all nodes were appropriately connected to the open loop plant. Then, the H_2-norm between disturbance and all control signals was calculated to provide a measure of the total system control effort. The value of R was then adjusted and all compensators were redesigned until the control effort norm equaled that of the centralized controller.

B. Performance of the distributed control system

The performance of distributed controllers using point-force/point-velocity transduction is demonstrated in Fig. 7, which shows the sound-power ratio for an uncontrolled plate, and a plate controlled by distributed compensators of reach 0, 1, and 5. One will note in Fig. 7 that all of the distributed control systems are able to achieve significant reductions in the overall radiated sound power. However, significant differences are apparent among the distributed compensators.

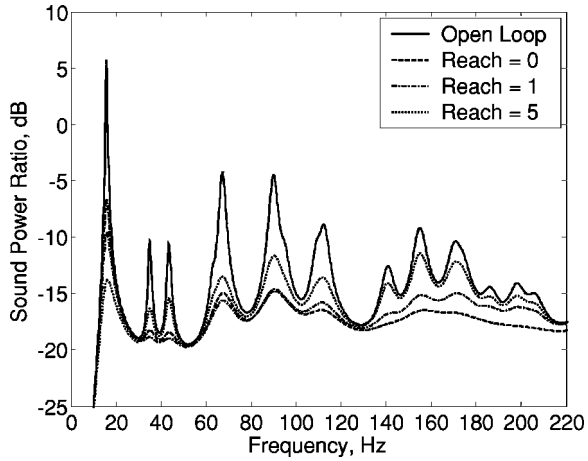


FIG. 8. Sound-power ratio for the uncontrolled system and distributed compensators of reach 0 (decentralized control), 1, and 5 using piezoelectric transduction.

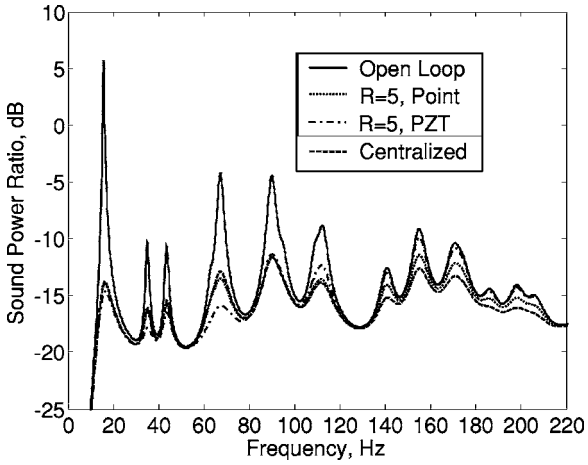


FIG. 9. Sound-power ratio for the uncontrolled system; distributed compensators of reach 5 with point and piezoelectric transduction; and a centralized controller using piezoelectric transduction.

The performance of the distributed compensators clearly improves as the reach of the system increases. However, there are significant differences in performance as a function of frequency. Most notably, the reach 5 compensator achieves significantly better reduction near the fundamental peak near 20 Hz, and (to a lesser extent) the modal peak near 70 Hz. On the other hand, the reach 0 outperforms the reach 5 system at all peaks above 140 Hz.

Although it provides interesting results, the use of point-force/point-velocity transduction is not very practical. A more practical approach is the use of piezoelectric transducers simultaneously as sensors and actuators. While this approach is not trivial, it has been demonstrated to be a feasible method.[21] The performance of distributed controllers employing such piezoelectric transducers is demonstrated in Fig. 8, which shows the sound-power ratio for an uncontrolled plate, and a plate controlled by distributed compensators of reach 0, 1, and 5.

As in the point-transduction case, the distributed control systems are able to achieve significant reductions in the overall radiated sound power. However, while the total radiated power decreases as the reach increases, the improvement is slight. In fact, one will note that the attenuations decrease with increasing reach for most modes with the exception of the fundamental. It is this ability to attenuate the fundamental (1,1) mode which makes the reach 5 system the best total power attenuator. But, for any other specific mode, the reach 5 system is not the best performer. Furthermore, for any mode above 100 Hz the decentralized (reach 0) system achieves the best reductions. This behavior is due to the fact that piezoelectric transducers are sensitive to strain (or strain rate) which results in low sensitivity to low-frequency modes. However, as the reach of the distributed compensators is increased, the combination of piezoceramic transducers that span the plate (or a large fraction of the plate) will have an improved sensitivity to these long-wavelength modes resulting in an improved ability to control them. A similar comparison between point and piezoelectric transduction in decentralized control was noted by Elliot *et al.*[19]

Finally, the sound-power ratio for the uncontrolled case, the case where a distributed controller of reach 5 is used and the case of a traditional, centralized controller (using piezoelectric transduction) is shown in Fig. 9. Note that the performance of the two control approaches is indistinguishable except at higher frequencies. Clearly, a distributed controller of reach 5 is able to perform nearly as well as a traditional centralized controller under these conditions. This is encouraging for future work in applying such scalable distributed control systems for use in larger, more complex systems.

V. CONCLUSIONS

The feedback control of acoustic radiation from a vibrating plate with a distributed sensor network has been presented. A distributed control architecture has been described that is based on the grouping of sensor nodes according to their physical proximity (called geographic grouping). The design of distributed control compensators was described and the performances of several such systems were compared. It was noted that this type of distributed control is capable of significantly reducing acoustic radiation. It was also noted that the distributed control performance approaches that of a traditional centralized control system if the size of the groups is sufficiently large. However, the only performance improvement of distributed over decentralized control was in attenuating the fundamental mode.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation CAREER Program under Grant No. 0134224 and the DARPA/IXO NEST program (F33615-01-C-1903).

[1]R. S. Raji, "Smart networks for control," IEEE Spectrum **31**, Issue 6, 49 (1994).
[2]I. D. McCammon and S. C. Jacobsen, "Communication and control for distributed microsystems," IEEE Control Syst. Mag. **10**, Issue 2, 48–50 (1990).
[3]C. R. Fuller, "Experiments on reduction of aircraft interior noise using active control of fuselage vibration," J. Acoust. Soc. Am. **78**, No. S1, 79 (1985).
[4]R. L. Clark and C. R. Fuller, "Control of sound radiation with adaptive

structures," J. Intell. Mater. Syst. Struct. **2**, No. 3, 431–452 (1991).
[5]R. L. Clark and C. R. Fuller, "Active structural acoustic control with adaptive structures and wavenumber considerations," J. Intell. Mater. Syst. Struct. **3**, No. 2, 296–315 (1992).
[6]R. L. Clark and C. R. Fuller, "Experiments on active control of structurally radiated sound using multiple piezoceramic actuators," J. Acoust. Soc. Am. **91**, No. 6, 3313–3320 (1992).
[7]N. R. Sandell, P. Varaiya, M. Athans, and M. G. Safonov, "Survey of decentralized control methods for large scale systems," IEEE Trans. Autom. Control **AC-23**, No. 2, 108–128 (1978).
[8]F. Giulietti, L. Pollini, and M. Innocenti, "Autonomous formation flight," IEEE Control Syst. Mag. **20**(6), 34–44 (2000).
[9]D. Stilwell and B. Bishop, "Platoons of underwater vehicles: Communication, feedback and decentralized control," IEEE Control Syst. Mag. **20**(6), 45–52 (2000).
[10]M. J. Krieger, J. B. Billeter, and L. Keller, "Ant-like task allocation and recruitment in cooperative robots," Nature (London) **406**(6799), 992–995 (2000).
[11]R. S. Chandra, J. Fowler, and R. D'Andrea, "Control of interconnected systems of finite spatial extent," *IEEE Conference on Decision and Control*, pp. 238–239 (2002).
[12]S. R. Hall, E. F. Crawley, and J. P. How, "Hierarchic control architecture for intelligent structures," J. Guid. Control Dyn. **14**, No. 3, 503–512 (1991).
[13]J. P. How and S. R. Hall, "Local control design methodologies for a hierarchic control architecture," J. Guid. Control Dyn. **15**, No. 3, 654–663 (1992).
[14]G. S. West-Vukovich, E. J. Davison, and P. C. Hughes, "The decentralized control of flexible space structures," IEEE Trans. Autom. Control **AC-29**, No. 10, 866–879 (1984).
[15]T. Hogg and B. A. Huberman, "Controlling smart matter," Journal of Smart Material Systems and Structures **7**, No. 1, R1–R14 (1998).
[16]P. Gardonio, E. Bianchi, and S. J. Elliott, "Smart panel with multiple decentralised units for the control of sound transmission. Part I Theoretical predictions, Part II Design of the decentralised control units, Part III Control system implementation," J. Sound Vib. **274**, 163–232 (2004).
[17]O. N. Baumann, W. P. Engels, and S. J. Elliott, "A Comparison of Centralized and Decentralized Control for the Reduction of Kinetic Energy and Radiated Sound Power," Proceedings of Active04, Williamsburg, VA, 2004.
[18]M. Baudry, A. Berry, and P. Micheau, "Decentralized Active Vibration Control of a Flexible Plate using PZT Actuator-PVDF Sensor Pairs," Proceedings of Active04, Williamsburg, VA, 2004.
[19]S. J. Elliot, P. Gardonio, T. C. Sors, and M. J. Brennan, "Active vibroacoustic control with multiple local feedback loops," J. Acoust. Soc. Am. **111**, No. 2, 908–915 (2002).
[20]J. Q. Sun, "Some observations on physical duality and collocation of structural control sensors and actuators," J. Sound Vib. **194**, 765–770 (1996).
[21]J. S. Vipperman, "Adaptive Piezoelectric Sensoriactuators for Active Structural Acoustic Control," Ph.D. thesis, Duke University, 1996.
[22]N. W. Hagood, W. H. Chung, and A. von Flotow, "Modeling of piezoelectric actuator dynamics for active structural control," J. Intell. Mater. Syst. Struct. **1**, No. 3, 327–354 (1990).
[23]R. Clark, W. Saunders, and G. Gibbs, *Adaptive Structures: Dynamics and Controls* (Wiley, New York, 1998).
[24]W. T. Baumann, W. R. Saunders, and H. H. Robertshaw, "Active suppression of acoustic radiation from impulsively excited structures," J. Acoust. Soc. Am. **90**, No. 6, 3202–3208 (1991).
[25]M. N. Currey and K. A. Cunefare, "The radiation modes of baffled finite plates," J. Acoust. Soc. Am. **98**, No. 3, 1570–1580 (1995).
[26]R. L. Clark and D. E. Cox, "Multi-variable structural acoustic control with static compensation," J. Acoust. Soc. Am. **102**, No. 5, 2747–2756 (1997).
[27]K. A. Cunnefare, "The minimum multimodal radiation efficiency of baffled finite beams," J. Acoust. Soc. Am. **90**, No. 5, 2521–2529 (1991).
[28]W. Levine and M. Athans, "On the determination of the optimal constant output feedback gains for linear multivariable systems," IEEE Trans. Autom. Control **AC-15**, No. 1, 44–48 (1970).

Noise annoyance and activity disturbance before and after the erection of a roadside noise barrier[a)]

Mats E. Nilsson[b)] and Birgitta Berglund
Institute of Environmental Medicine, Karolinska Institutet and Department of Psychology, Stockholm University, Stockholm, SE-106 91 Sweden

(Received 4 September 2005; revised 5 January 2006; accepted 6 January 2006)

Questionnaire studies were conducted in a residential area before and after the erection of a 2.25 m high noise barrier of conventional type along a heavily traveled road (19 600 vehicles/24 h). The interval between studies was two years. Houses closest to the barrier received a sound-level reduction from ~70.0 to 62.5 dB L_{den} at the most exposed facade. The sound-level reduction decreased with distance to the road, and was negligible for houses at more than 100 m distance. Up to this distance, the noise barrier reduced residents' noise annoyance outdoors and indoors as well as improved speech communication outdoors. Indoors, speech communication and sleep disturbance were slightly but nonsignificantly improved. Predictions of the number of annoyed persons from published exposure-response curves (in L_{den}) agreed with the percentage of residents being annoyed when indoors, before and after the barrier. Conversely, the percentage of residents being annoyed when outdoors clearly exceeded the predictions. These results suggest that these exposure-response curves may be used in predicting indoor situations, but they should not be applied in situations where outdoor annoyance is at focus. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2169906]

PACS number(s): 43.50.Qp, 43.50.Gf, 43.50.Sr [DKW] Pages: 2178–2188

I. INTRODUCTION

A large proportion of a country's population is typically exposed to unhealthy levels of road-traffic noise (WHO, 2000). For example, approximately 20% of the population of the European Union is exposed during the daytime to continuous equivalent sound levels of road-traffic noise greater than 65 dBA (CEC, 1996). To create acceptable sound environments in residential areas along heavily traveled main roads and motorways, the two most common procedures are to build roadside barriers and to improve window insulation. Thus far, the efficiency of roadside barriers has typically been predicted on the basis of sound-level reduction. The goal of the present research was to evaluate a barrier's efficiency to reduce residents' annoyance and activity disturbances. As pointed out by Fidell, Silvati, and Haboly (2002), opportunities and resources for this type of study are rare because it requires a longitudinal, noise-barrier intervention design. The present research contributes to the noise-barrier literature by presenting data on (1) the efficiency of a roadside barrier to reduce annoyance and activity disturbances at various distances to the road, and highlighting, (2) the ability of published synthesized curves of noise annoyance (Miedema and Oudshoorn, 2001) to predict the change in the proportion of annoyed persons accomplished by the noise barrier.

The relationship between road-traffic noise exposure and annoyance has earlier been studied extensively for steady-state situations (for reviews, see Fidell, Barber, and Schultz, 1991; Miedema and Vos, 1998; WHO, 2000). The proportions of annoyed residents were then compared among groups of residents who live in differently exposed residential areas. The variation in noise exposure was, thus, accomplished by selecting residential areas at different distance to a road and/or with different traffic-flow composition and volume. Synthesized exposure-response curves based on several such cross-sectional studies constitute a tool for predicting the extent of annoyance geographically under "steady-state circumstances." However, if a substantial change in road-traffic noise exposure occurs in an area, the corresponding *change* in proportion of annoyed residents in the same area is typically greater than would be predicted from the synthesized (cross-sectional) exposure-response curves (Brown, 1987; Brown, Hall, and Kyle-Little, 1985; Griffiths and Raw, 1986, 1989; Öhrström, 2004; but see Fields, 1993). The same may be true for changes in aircraft noise exposure (Raw and Griffiths, 1985; Fidell, Silvati, and Haboly, 2002; Horonjeff and Robert, 1997).

To our knowledge, Lambert (1978) and Kastka *et al.* (1995) are the only published longitudinal annoyance studies on roadside noise barriers. In Lambert (1978), the interval between the before and after study was one year; the after study took place only six months after a freeway noise barrier had been built. The relationship between annoyance and noise exposure was well described by a single curve, suggesting that in principle the reduction in annoyance could be predicted from the corresponding sound-level reduction. However, no nonbarrier control condition was used and no

[a)]Portions of this work were presented in "Evaluation of a road-traffic noise barrier: An intervention study," Proceedings of the Tenth International Congress on Sound and Vibration, Stockholm, July 2003.

[b)]Author to whom correspondence should be addressed. Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden. Electronic mail: mats.nilsson@psychology.su.se

 0001-4966/2006/119(4)/2178/11/$22.50

FIG. 1. Aerial photograph of studied residential areas taken before the noise barrier was erected. The white lines show location of the future barrier. The apartment buildings of the reference area are marked by arrows.

comparison was yet possible to make with synthesized exposure-effect curves (i.e., Schultz, 1978). In Kastka *et al.* (1995), the study-interval was over 12 years and the after studies for four barriers were between 8 and 10 years. In spite of potential confounders, they argue for a greater than expected reduction in annoyance as compared to the change predicted from cross-sectional studies. However, Kastka *et al.* also refer to a study by Kastka and Paulsen (1979; in German) in which "short-term" effects of a noise barrier was found to be smaller than expected.

The present longitudinal study assesses the efficiency of the noise-barrier intervention as such. It involves a more recent traffic situation and a substantially shorter time span than used by Kastka *et al.* (1995), (2 years apart in 1999 and 2001 versus 12 years apart in 1974 and 1986), but longer time span, between barrier erection and the after study, than used by Lambert (1978), (15 versus 6 months). Importantly, a reference area was included in the present study as a no-barrier control of potential confounders that may affect residents' annoyance and activity disturbance (e.g., changes in traffic volume or other kinds of noise exposure). Furthermore, the noise-barrier's efficiency in reducing sound level, annoyance, and activity disturbances was determined separately for indoor and outdoor living conditions.

II. METHOD

A. The noise barrier

The barrier was an 800 m long conventional reflective noise barrier of wood, each 50 m interspersed with 1.5 m long sections of glass. It was 2.25 m high and 65 mm thick. The barrier was located between the main road and a small local road to the residential areas. Its distance to the center of the main road was 10 m, and its distance to facades of houses closest to the road was 12 m. Figure 1 shows an aerial photograph with the location of the barrier (white lines) indicated relative to the detached houses and the main road.

B. Design, study population, and participation rate

A longitudinal study was designed with the noise barrier acting as an intervention between two waves of questionnaire data collection. The questionnaire waves were conducted on two occasions, two years apart, in the same residential areas: (1) In May-June 1999, three months before the building start of the noise barrier ("before-study"), and (2) in May-June 2001, 15 months after the barrier was erected ("after-study"). Furthermore, questionnaire data was collected in (a) an "experimental area" along a part of the road where the noise barrier was built (see detached houses behind the barrier in Fig. 1), and in (b) a "reference area," with no noise barrier, along another part of the road adjacent to where the noise barrier was built (see arrows pointing at apartment buildings in Fig. 1).

In order to assure very similar road-traffic noise exposure, the experimental and reference areas should be exposed to noise from the same section of the road. Since all detached houses along the road were to be noise protected by the barrier, an apartment building area was selected as the reference area. This area would serve as a good reference for the change in traffic-noise exposure.

The experimental and reference areas involved a total of 329 households in detached houses and 236 households in apartment buildings, respectively. The corresponding two study populations were all the 644 and 363 adult residents (18 years or older) living in the respective houses and apartments. In the before-study, the residents' participation rate was 77% ($n=495$) and 66% ($n=241$) in the experimental and reference area, respectively, and correspondingly in the after study 72% ($n=462$) and 69% ($n=250$). Out of these two original study populations of residents, 59% ($n=379$) and 46% ($n=166$) in the experimental and reference area, respectively, participated in *both* waves, that is, both before and after the noise-barrier intervention.

Thus, the questionnaire data consists of three sets: Groups of respondents (a) who only participated in the before-study, (b) who only participated in the after-study, or (c) who participated in both the study waves. Since the present research evaluates the consequences for residents' health and wellbeing of *changes* in the traffic-noise exposure caused by the noise barrier, the data analyses to be presented were restricted to data set (c), that is, residents in the "experimental" and "reference" groups, who participated in both study waves. A great advantage of this type of comparison is that personal variables can be assumed to be invariant between study waves.

C. Traffic noise assessment

According to a traffic count in 1994, the number of vehicles on the main road was 19 000 vehicles/day. During two weeks, a corresponding traffic count was conducted automatically in the after-barrier study. Traffic counts were also secured in connection with acoustic measurements in both the before-study and after-study (see below). Thus, apart from the acoustic measurements, sufficient data were collected for estimating reliably the change in traffic flows between the two study waves. The noise exposure from the other roads was negligible because of infrequent traffic.

The Nordic Calculation Model (Swedish EPA, 1996) was used for computing L_{den} contours of road-traffic noise as free-field values at facade, 1.5 m above ground, for the two

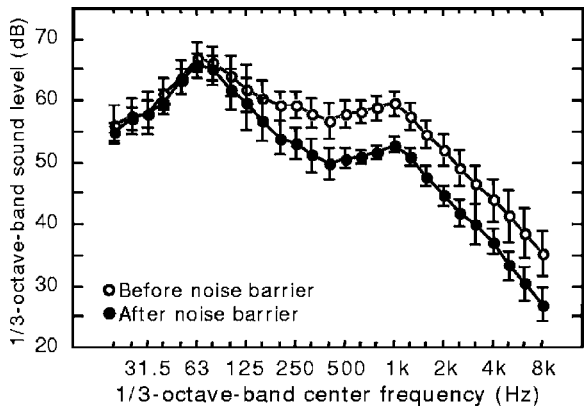


FIG. 2. Spectra for 45 min measurements conducted in eight gardens closest to the main road (23 m). Points show arithmetic averages of 1/3-octave-band sound levels ($L_{eq,45min}$) from eight measurements conducted before (open circles) and after the barrier (filled circles) was erected (error bars: ±1 standard deviation).

occasions with and without the noise barrier. These calculations were based on traffic flow data from the automatic traffic count. Corresponding contours were also computed for $L_{Aeq,24h}$, because this is currently the main traffic-noise index used in Sweden. Computed L_{den}-values were on average 2.5 dB higher than corresponding $L_{Aeq,24h}$-values. The L_{den}-contours will be used in the following data presentation, because (a) L_{den} is the noise index recommended by the European Community (EC, 2002) and (b) updated synthesis curves of noise annoyance are available for L_{den}, but not for $L_{Aeq,24h}$ (Miedema and Oudshoorn, 2001).

During both study waves, acoustic outdoor and indoor measurements were simultaneously conducted in the "experimental area" at each of 26 addresses at various distances to the main road (23–236 m). Of these, eight were used for characterizing the shielding efficiency of the barrier (see Fig. 2). These recordings were located close to the road (23 m) in houses and gardens that were to be completely shielded by the noise barrier. In the after study, the acoustic measurements were conducted in the same location, the same month, the same weekday and the same hour of the day as were the measurements in the before-study. Outdoor measurement microphones were located ~3 m away from the short side of the house, at 1 m distance behind the facade facing the main road. For building sites closest to the road, the perpendicular distance between the center of the road and the outdoor microphone location was ~23 m. Indoor recordings were made in living rooms on the ground floor (floor level ~1.5 m above the ground), 1 m behind a closed window facing the main road (approx. parallel with the outdoor microphone relative to the road). Microphone height was always 1.6 m for both outdoor and indoor recordings. The duration of the recordings was 45 min, during which period the passing vehicles were always counted manually.

All acoustic measurements were conducted during the month of May in 1999 and 2001. According to data from a stationary weather station nearby, average wind speed of the month was 3 m/s in both years. Average temperature were 10 and 11 °C, and average relative humidities were 63% and 59%, in 1999 and 2001, respectively. Thus, meteorological conditions during the measurements were similar for both years.

D. Questionnaire

Questionnaires were delivered and collected at respondents' homes by personnel officially representing Stockholm University. The data collection procedure was the same in the before-study and after-study. On both occasions, all adult members of all households of the residential area were asked to participate.

The same questionnaire was used in the before-study as in the after-study. It contained 64 questions, which were answered in approximately 15 min. The entire questionnaire is displayed in Swedish in Nilsson and Berglund (2002). The questionnaire was self-administered in both the experimental and the reference areas. There were a few slight differences in how the questions were phrased in the questionnaires distributed in the two areas; for example, reference to "the garden" in the experimental area (detached houses) was replaced by "outdoors close to the apartment" in the reference area (apartment buildings with cultivated green areas).

The main part of the questionnaire consisted of questions on noise annoyance, interference with speech communication and sleep disturbance. Questions on noise annoyance referred to road-traffic noise from the main road either as perceived indoor or as perceived outdoors. Annoyance was measured using a visual analog scale, with endpoints marked "not at all annoyed" (coded 1) and "extremely annoyed" (coded 7). Each annoyance question was preceded by a question on how often the road-traffic noise from the main road was heard (response alternatives: "Never," "sometimes," "often"). The respondents were instructed to always answer both questions. The annoyance response was always taken as valid independent of the answer to the preceding question, on how often the source was heard. However, a *missing* annoyance response combined with a "never-heard" answer was scored as "not at all annoyed." This coding of data reduced the internal response loss on annoyance questions from 14% to 4.7%.

How traffic noise interfered with speech communication was assessed for specific activities outdoors or indoors (relaxed conversation, talking/listening in telephone/cell phone, listening to television/radio). Category scale response alternatives were "never," "seldom," "sometimes," "often," and "very often" disturbed. The same kinds of questions on specific activities and the same response alternatives were used for sleep disturbance, though with reference to situations indoors only (difficulties to fall asleep, awakened too early in the morning, poor sleep quality).

E. Method of analysis

For reasons of traffic safety, the noise barrier was built in such a way that it did not fully protect houses close to a few local-road entrances from the main road (Fig. 1). In order to produce fair results on the efficiency of the noise barrier, respondents from these particular houses and houses behind them along the local roads perpendicular to the main

TABLE I. Number of respondents and houses at various distances (m) to the center of the main road.

	Experimental area						Reference area
	<25 m	51–75 m	76–100 m	101–150 m	151–225 m	>225 m	55–220 m
No. of respondents	52	47	31	35	62	77	166
No. of houses	29	26	18	23	45	48	126[a]

[a]Apartments.

road were excluded from the data analyses. Data from 75 out of 379 respondents were excluded. Thus, the analyses were based on data from 304 respondents in the experimental areas, who all lived in houses to be fully noise protected by the barrier.

In the experimental area, data was analyzed separately for six groups of residents classified according to their houses' distance to the main road. Residents in the reference areas were treated as a single group. Compared to a classification based on noise exposure, the obvious advantage of a classification by distance is that residents will be members of the same group in the data analysis of the before-study and the after-study. (The noise barrier affects the noise exposure at the facade, not the distance to the road.) Table I presents number of respondents and houses (or apartments in the reference area) in the distance-to-the-road groups. The two first distance-to-the-road groups (<25 and 51–75 m) correspond to the first and second row of houses in the experimental area (cf. Fig. 1).

Moreover, the percentage of annoyed residents in the experimental area was calculated for each group of residents whose houses were found within computed noise contours (L_{den}). To obtain the percentages, the 1–7 analogue annoyance scale was linearly transformed to a 0–100 scale; values >50 and >72 were classified as "annoyed" (%A) and "highly annoyed" (%HA), respectively (cf. Miedema and Oudshoorn, 2001). The number of residents within each of the six 2.5 dB step contours ranged from 14 to 80 (Mdn =43.5) in the before-study and from 25 to 107 (Mdn =43.5) in the after-study.

In the present comparison of results on change due to the noise barrier, individual residents were the *unit* of the longitudinal data analysis (neither houses nor apartments). The average number of participants from each household was 1.5 (range: 1–3). The residents were carefully instructed to fill in the questionnaire independently of other members of the household (or apartment). In the statistical analysis it was, therefore, assumed that residents' responses were independent of each other. By including two or more respondents from the same household, the variances in the annoyance responses may have been underestimated. As a consequence, the confidence intervals would also have been larger. However, parallel statistical analyses were also made with only one randomly selected resident per household (Nilsson and Berglund, 2002, in Swedish). The conclusions drawn from the results of these control analyses agreed well with those presented below.

III. RESULTS

A. Traffic volume

According to the automatic traffic count during the after study, the traffic volume was on average 19 600 vehicles/24 h in working days (6% heavy vehicles). The traffic volume during rush hours, based on 26 45 min manual traffic counts, was on average 912 vehicles/45 min in the before-study and increased to on average 1003 vehicles/45 min in the after-study. This slight increase (<10%) in traffic volume corresponds to an increase in sound level of <0.5 dB and may, therefore, be considered negligible.

B. Noise exposure outdoors

For the experimental area, Table II gives two sets of cross-tabulations of the number of *respondents* with regard to computed noise contours by distances of homes to the main road; one set for the before-barrier study and another set for the after-barrier study (marked BS and AS, respectively).

All the 52 respondents, who lived closest to the road (<25 m), were located in the 70.0–72.5 dB contour before the noise barrier and moved to the 62.5–65.0 dB contour after the barrier was built. At this short distance, the barrier caused an L_{den} decrease of approximately 7.5 dB. The 47 respondents, who lived next closest to the road (51–75 m) were located in one of the two contours between 55.0 and 60.0 dB and moved to one of the two contours between 50.0 and 55.0 dB after the barrier. This corresponds to an L_{den} decrease of approximately 5.0 dB. The barrier caused the 31 respondents living at 76–100 m distance to the road to be moved from one of the four contours between 47.5 and 57.5 dB to one of the four contours between 45.0 and 55.0 dB. The corresponding L_{den} decrease was approximately 2.5 dB L_{den}. Only a few of the residents residing more than 100 m from the main road were moved to lower contours because of the noise barrier. At best, they received a small decrease in dB L_{den}.

Figure 2 shows average 1/3-octave band spectra of eight 45 min measurements, before (open circles) and after (filled circles) the noise barrier was built. These measurements were from the eight gardens closest to the road (<25 m), which all would become completely shielded by the noise barrier. As expected, the noise barrier did not reduce the sound levels in the low-frequency bands, that is, below ~ the

TABLE II. Number of respondents in the experimental area according to noise contour (dB L_{den}), distance to center of main road (m), and study occasion (before and after the barrier was erected).

Contour dB L_{den}	<25 m		51–75 m		76–100 m		101–150 m		151–225 m		>225 m		*Total*	
	BS[a]	AS[b]	BS[a]	AS[b]	BS[a]	AS[b]	BS[a]	AS[b]	BS[a]	AS[b]	BS[a]	AS[b]	*BS*[a]	*AS*[b]
≤45.0	0	0	0	0	0	0	0	2	20	34	60	71	*80*	*107*
>45.0–47.5	0	0	0	0	0	3	4	15	34	28	17	6	*55*	*52*
>47.5–50.0	0	0	0	0	3	17	23	18	6	0	0	0	*32*	*35*
>50.0–52.5	0	0	0	24	11	9	6	0	2	0	0	0	*19*	*33*
>52.5–55.0	0	0	0	23	15	2	2	0	0	0	0	0	*17*	*25*
>55.0–57.5	0	0	33	0	2	0	0	0	0	0	0	0	*35*	*0*
>57.5–60.0	0	0	14	0	0	0	0	0	0	0	0	0	*14*	*0*
>60.0–62.5	0	0	0	0	0	0	0	0	0	0	0	0	*0*	*0*
>62.5–65.0	0	52	0	0	0	0	0	0	0	0	0	0	*0*	*52*
>65.0–67.5	0	0	0	0	0	0	0	0	0	0	0	0	*0*	*0*
>67.5–70.0	0	0	0	0	0	0	0	0	0	0	0	0	*0*	*0*
>70.0–72.5	52	0	0	0	0	0	0	0	0	0	0	0	*52*	*0*

[a]BS="Before-study," conducted before the barrier was erected.
[b]AS="After-study," conducted after the barrier was erected.

200 Hz band. Outdoors, the median sound levels, in dB $L_{Aeq,45min}$, of the eight gardens closest to the main road (<25 m) were 66.0 dB (range: 64.0–70.0 dB) in the before-study and 59.5 dB (range: 58.0–63.0 dB) in the after-study. The difference value at the same garden sites before and after the barrier was built was 6.5 dB (range: 5.5–7.5 dB $L_{Aeq,45min}$). Based on computed L_{den} contours and measured $L_{Aeq,45min}$ the accomplished reduction of the barrier was estimated to approximately 7 dB for houses closest to the main road (<25 m). This is within the interval of 5–12 dB expected from a conventional reflective noise barrier (I-INCE, 1999; Watts, 2002).

Sound-levels were also measured at distances further away from the main road (63–236 m). In most cases, these levels agreed well with the predicted values. However, occasionally higher levels were measured because of the influence of additional sound sources, such as bird song and children playing.

C. Noise exposure indoors

In the eight homes closest to the road (<25 m), which subsequently were all fully shielded by the barrier, the measured median *indoor* sound level ($L_{Aeq,45min}$) was in the before-study 43 dB (range: 42–48 dB) and in the after-study 41 dB (range: 36–46 dB). Thus, the noise barrier did not change the overall indoor sound levels much. This is explained by additional sound sources (e.g., ventilation systems and refrigerators), which contributed substantially to the overall indoor sound levels. It is most likely that these between-study indoor sound levels were invariant in the houses because sound insulation and stationary indoor sources would not normally change much.

D. Annoyance as a function of distance to the road

In the text to follow, the barrier's effect on road-traffic noise annoyance will be presented in two forms, either (a) as the average annoyance scale values (arithmetic mean) of the 1–7 analog scale or (b) as the percentage of respondents who reported that they were annoyed (%A) and highly annoyed (%HA) according to the two internationally adopted cutoffs on the analog scale (Miedema and Oudshoorn, 2001).

1. Average annoyance (scale 1–7)

In the experimental area, the *average* annoyance with road-traffic noise exhibited an overall decrease with increasing distance to the main road (Fig. 3, two left panels). This trend was found in the before-study (open circles) as well as the after-study (filled circles). For residents living in houses up to ~100 m from the main road, noise annoyance outdoors (upper panel) and noise annoyance indoor (lower panel) were significantly lower with the noise barrier than without ($p < 0.05$). Significant differences in noise annoyance between the two study occasions was neither found for residents in houses >100 m from the main road, nor for residents in the reference apartment-building area. The right panels of Fig. 3 show these significant and insignificant annoyance differences with distance (bars indicate 95% confidence intervals). In the reference area, slight between-study change in traffic flow affected insignificantly the outdoor and indoor annoyance reports (Fig. 3, two right panels).

However, the smooth trend of decreasing outdoor annoyance with increasing distance to the road was interrupted at 50–75 m distance in both the before and the after study (Fig. 3, upper left panel). At this distance the residents reported slightly more annoyance outdoors than the residents closer to the main road. This reverse rank order of effects was not found for indoor annoyance (lower left panel). The reason for the reverse effect is probably the location of the houses and their garden relative to the road (Fig. 1): Houses closest to the main road (<25 m) are located in front of their (noise-shielded) gardens, whereas houses 51–75 m from the road are located behind their gardens (shielded by first row of houses but not second row). Thus, the second row of houses' gardens were not only exposed to the (direct) road-

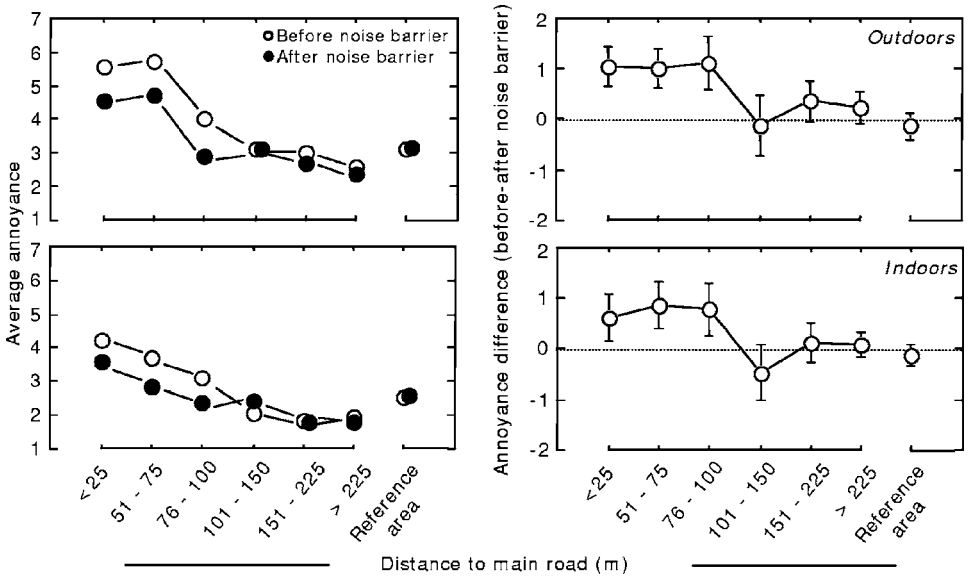


FIG. 3. Road-traffic noise annoyance (left panels) and annoyance difference between study waves (right panels), each plotted as a function of distance to the main road. Upper panels: outdoor annoyance. Lower panels: indoor annoyance. Left panels: Open symbols=before the barrier; filled symbols=after the barrier. Bars: 95% confidence intervals.

traffic noise from the main road (partly shielded by first-row houses), but also to the acoustic reflexes caused by second-row houses' facades.

2. Percentage of annoyed residents

Figure 4 shows the results for *percentage* of annoyed (left panel) or highly annoyed (right panel) residents. An overall decrease was found in both %A and %HA in groups of residents living at the six categories of increasing distances to the main road. This annoyance decrease was larger without the barrier than with the barrier (open and filled circles, respectively) and larger outdoors than indoor (upper and lower panels, respectively). In the reference area no difference was found in %A and %HA between the before and after study. As expected, in the experimental area, a greater number of residents were annoyed by road-traffic noise outdoors (upper panels) than indoor (lower panels). Significant reductions in %A and %HA were only observed for the three groups of residents whose houses were within 100 m from the main road ($p<0.05$, sign-test). Thus, the same conclusions were drawn from data on *percentage* of annoyed residents (%A and %HA, cf. Fig. 4) as were drawn from data on *average* annoyance (Fig. 3).

The questionnaire included questions on annoyance with many other sound sources than road-traffic noise from the main road, for example, road-traffic noise from local roads, aircraft noise, railway noise, noise from lawn movers, etc. Since road traffic from the main road was the predominant source of noise annoyance, annoyance results for the other noise sources are not presented here. Residents' heard these other noises infrequently and in most cases they were given very low annoyance values (Nilsson and Berglund, 2002, in Swedish).

E. Activity disturbance as a function of distance to the road

A comparison of results on activity disturbance showed that the rank orders of the percentages of disturbed residents were the same in the residential and reference areas. Road-traffic noise interfered most with speech communication outdoors, somewhat less with sleep indoor and least with speech communication indoor. At most, 12% of the residents in the

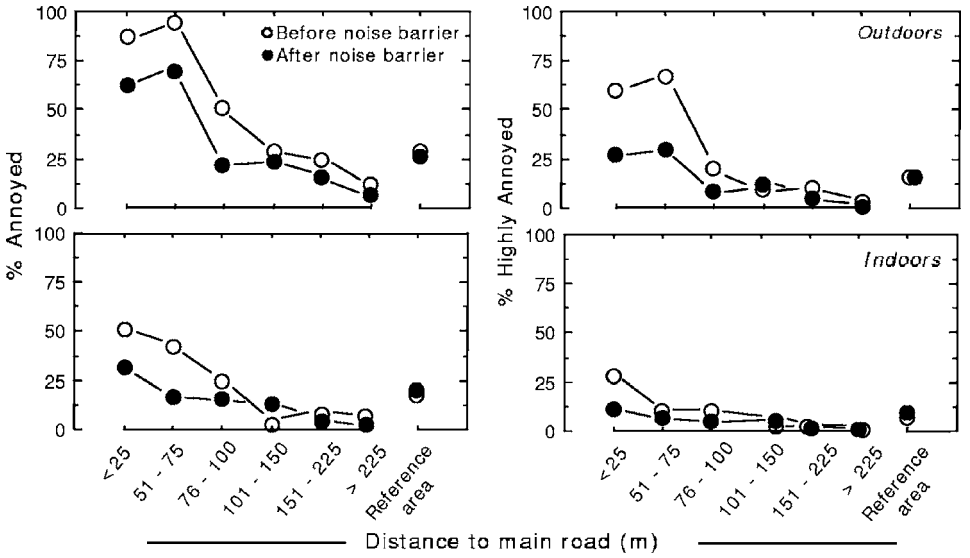


FIG. 4. Percent annoyed (%A, left panels) and percent highly annoyed (%HA, right panels), each plotted as a function of distances to the main road. Upper panels: percent annoyed outdoors. Lower panels: Percent annoyed indoors. Open symbols: Before the barrier. Filled symbols: After the barrier.

TABLE III. Percent respondents reporting activity disturbance *often* or *very often* due to road-traffic noise at various distances (m) to the center of the main road.

Item[c]	Study	Experimental area <25 m	51–75 m	76–100 m	101–150 m	151–225 m	>225 m	Reference area 55–220 m
Difficulties to hear…				*Interference with speech communication outdoors*				
…speech outdoors[d]	Before	71	60	6	3	0	0	11
	After	46[a]	40[a]	10	3	0	1	12
…television/radio	Before	33	50	7	3	2	0	9
outdoors[d]	After	16	34	7	3	0	3	8
…telephone outdoors[d]	Before	46	48	7	3	2	0	8
	After	15[b]	30[a]	7	3	3	0	11
				Interference with speech communication indoors				
…speech indoors	Before	10	9	0	0	0	0	2
	After	4	2	0	0	0	1	4
…television/radio	Before	13	4	0	0	0	0	2
indoors	After	6	0	0	0	0	0	4
…telephone indoors	Before	2	0	0	0	0	0	1
	After	4	0	0	0	0	0	1
				Sleep disturbance (indoors)				
Difficulties falling	Before	4	4	0	0	2	1	9
asleep	After	6	6	3	0	3	0	5
Awakened too early	Before	21	24	23	6	2	3	7
	After	20	18	7	9	3	1	8
Poor sleep quality	Before	8	11	10	0	2	1	4
	After	6	6	7	0	3	3	7

[a]$p<0.05$.

[b]$p<0.01$; Probabilities refer to sign-test for repeated measurements, testing if the number of residents that changed classification ("often-or-very-often-disturbed" vs "not-often-or-very-often-disturbed) between the two study occasions exceeds the number expected by chance. Non-significant differences are not indicated in the Table.

[c]"How often does road-traffic noise from the main road cause the following problems?" Response alternatives: "Never," "seldom," "sometimes," "often," "very often."

[d]In the experimental area (houses), "outdoors" was defined as "outdoors in your garden," in the reference area (apartments) "outdoors" was defined as "outdoors close to your apartment."

reference area reported any of the activity disturbances (Table III). The results for the experimental area are presented in more detail below.

1. Activity disturbances in gardens

The upper part of Table III shows percentages of residents who reported that road-traffic noise interfered with speech communication outdoors in their gardens (often + very often). These interferences were substantial in the first two rows of gardens relative to the main road, between 33% and 71% of the residents in the before-study. After the barrier had been built, speech intelligibility and telephone conversation was found to have improved significantly for residents living up to 75 m from the main road. However, these problems prevailed, since 46% of the residents in gardens closest to the road, and 40% of those in the next row of gardens, reported that they still often or very often found it difficult to hear speech outdoors. Although the barrier caused a significant improvement also for telephone listening outdoors, the road-traffic noise still interfered for 15% of the residents in the first row and 30% in the second row of gardens.

2. Activity disturbances indoor

The middle or lower parts of Table III show the percentages of residents who reported that road-traffic noise interfered (often + very often) indoors with speech communication or sleep disturbance, respectively. Difficulties to hear conversation in rooms, television/radio or telephone were not reported to be a frequent indoor problem; at most 9%–13% of the residents in the first two rows of houses experienced such disturbances. Although fewer residents reported communication problems after the erection of the barrier, the differences were not statistically significant ($p>0.2$, sign-test). Sleep disturbances due to road-traffic noise was not much improved by the barrier (Table III). Within 100 m distance to the road, 21%–24% of the residents reported early awakenings in the morning and 8%–10% reported poor sleep quality (often + very often) due to the road-traffic noise. The noise barrier did not significantly reduce these disturbances (for early awakenings, $p>0.1$, sign-test). Notably, 18%–20% of the residents in the first two rows of houses were awakened too early also after the barrier had been built. An obvious reason is that master bedrooms were typically located at the first floor, which was not fully protected by the noise barrier.

F. Annoyance as a function of day-evening-night sound level (L_{den})

Figure 5 shows percentages of annoyed (%A, left panels) and highly annoyed (%HA, right panels) residents as a function of computed L_{den} contours (center values are given,

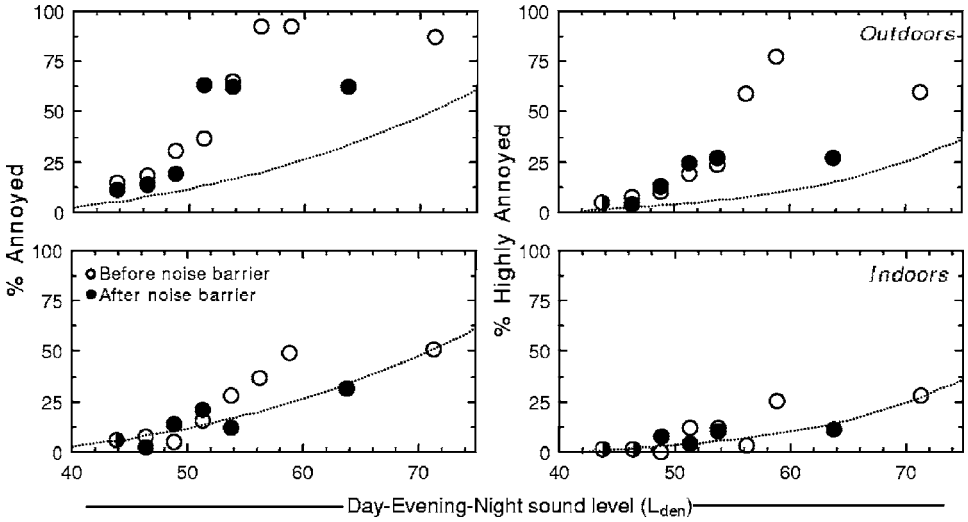


FIG. 5. Percent annoyed (%A, left panels) and percent highly annoyed (%HA, right panels), each plotted as a function of computed L_{den} contours in dB. Upper panels: Outdoor %A or %HA. Lower panels: Indoor %A or %HA. Open symbols: before the barrier. Filled symbols: after the barrier. L_{den} is given as center values of the computed 2.5 dB contours (43.75 dB refer to <45 dB, 51.25 dB refer to 50.0–52.5 dB, etc.) Thin line: Miedema and Oudshoorn's (2001) synthesized annoyance curve.

43.75 dB for <45 dB contour, 51.25 dB for 50.0–52.5 dB contour, etc). The two upper panels give the results for outdoor annoyance and the two lower panels for indoor annoyance. To facilitate comparison, corresponding synthesized exposure-response curves from Miedema and Oudshoorn (2001) have been drawn as thin lines in the four panels of Fig. 5. Please observe that these synthesized exposure-response curves (thin lines) refer to road-traffic annoyance in general, with no reference to or request of the respondents to distinguish between indoor annoyance and outdoor annoyance. The last two columns of Table II present the number of residents within each contour, the range of which is from 14 to 107 residents.

For *outdoor* annoyance (upper panels), the %A and %HA residents found to be exposed above 50 dB L_{den} were much greater than predicted from Miedema and Oudshoorn's (2001) synthesized curves, both before and after the noise barrier was built. This is notably true for residents in the *before*-study with computed L_{den} exposures of 55–60 dB. In fact, a higher percentage of these residents were classified as annoyed than in the groups of residents closest to the road (exposed to >70 dB L_{den}). As pointed out above (in connection with Fig. 3), the residents living 51–75 m from the main road violate a potentially smooth monotonic annoyance function, possibly because their annoyance in the gardens was caused also by acoustic facade reflections. Notably, the curves' annoyance predictions already underestimate the reported annoyance of residents living in gardens closest to the road (<25 m).

For *indoor* annoyance (lower panels), the %A and %HA agreed better with Miedema and Oudshoorn's (2001) synthesized exposure-response curves. Distinct deviations were only found for residents within the range of 55.0–60.0 dB L_{den}, especially for %A. The residents who deviate from the predictions, notably, lived in the second row of houses, 51–75 m from the main road.

IV. DISCUSSION

The noise barrier was found to reduce residents' noise annoyance (outdoors and indoors) and to improve their speech communication (outdoors). These effects were significant only for residents living up to 100 m from the noisy road, not for residents living further away. It may be concluded that the noise barrier caused the observed changes in annoyance and activity disturbances, because corresponding changes were not found in the reference area along the same road where no barrier was built.

A. Distance-to-road effects of roadside barrier

Lambert (1978) and Kastka *et al.* (1995) found reductions in residents' annoyance up to 150 m distance to the road; at larger distances, no or very small effects were found. We found corresponding effects up to 100 m from the road, but not at greater distances. The traffic-volumes and the barrier heights were greater in the Lambert (100 000 vehicles/24 h; 5.8 m) and the Kastka *et al.* (20 000–50 000 vehicles/24 h; 3.2–6 m) studies than in the present study (20 000 vehicles/24 h, 2.5 m). These two factors may explain why Lambert and Kastka *et al.* still found effects at 100–150 m distances from the main road.

The before-study and after-study of Kastka *et al.* (1995) were conducted 12 years apart, involved four roadside barriers, and a mix of the same and different participating residents. Apart from securing comparisons for the same group of residents, the present shorter between-study period of two years better guards against other time-related confounders, such as residents' age, time of residency, traffic volume, and vehicle models. In this respect, the present study is more similar to the study of Lambert (1978), which was also conducted in one residential area before and after the noise barrier was built, but with a post-barrier period to the after-study of only six months and no without-barrier reference area. The effects of the roadside barrier on outdoor and indoor noise annoyance (or activity disturbances) were not separated by Lambert or Kastka *et al.* The present noise barrier was found to reduce both outdoor *and* indoor annoyance up to a distance of 100 m from the main road. Moreover, as will be discussed below, the annoyance reductions indoors, but *not* outdoors, were well predicted by Miedema and Oudschoorn's (2001) synthesized exposure-effect curves. These differences in annoyance reduction support the view that indoor and outdoor annoyance should be measured separately.

Residents living 51–75 m from the road ($n=52$) reported slightly higher annoyance and activity disturbance than residents closer to the road (<25 m; $n=47$), although the latter group of residents had a higher predicted noise exposure. Compared to residents living closer to the road, the gardens of residents living 51–75 m from the road (2nd row) were less well shielded from the road traffic noise. These gardens were located behind the first-row houses and in front of the second-row houses. However, the size of the annoyance *reduction* caused by the roadside barrier was the same for this group of residents as for the other two groups living within 100 m from the road (Fig. 3, right panels). This supports the validity of the annoyance assessments as a measure of noise barrier efficiency.

B. Outdoor and indoor effects of roadside barrier

As expected, residents' annoyance and interference with speech communication were greater in the gardens than inside in the houses, simply because the noise exposure was higher outdoors than indoors. The consistently lower annoyance indoors than outdoors supports that the residents were able to report the two types of annoyances separately. The reduction accomplished by the barrier indoors was particularly distinct for annoyance. For instance, up to 100 m distance to the road, the road barrier significantly affected the indoor annoyance. At further distances, the indoor annoyance reached a slightly lower level than outdoor annoyance (cf. Fig. 3).

The difference in measured indoor sound levels (all sources inclusive) before and after the barrier was erected was small (~2 dB as measured with closed window in houses closest to the road). Still, the indoor annoyance was significantly reduced (e.g., Fig. 3). It is therefore likely that the noise barrier actually improved residents' *perceived* sound environment indoors. Road-traffic may be heard and thereby be annoying at levels below background indoor sound levels. Therefore, a reduction in road-traffic noise exposure may have an effect on annoyance although the effect on the overall measured sound level, including all sources, is small (e.g., Fidell and Teffeteller, 1981). Moreover, it is probable that residents' reports of indoor annoyance would be based also on their experience of road-traffic noise with open windows. In such circumstances, the sound levels of the road traffic noise would be considerably higher than with windows closed, perhaps 15 dB (Berglund and Nilsson, 2001; Berglund and Nilsson, 2002).

C. Comparison with recent synthesis curves

The present results were compared with Miedema and Oudshoorn's (2001) synthesized exposure-response curves. Two aspects of this comparison are discussed below.

First, the studies in Miedema and Oudshoorn's (2001) synthesis refer to road-traffic annoyance in general, whereas we separated annoyance as experienced in the garden or inside the house. Percentage annoyed (%A or %HA) *indoors*, obtained before and after the barrier was built (Fig. 5, lower panels), agrees fairly well with predictions from the synthesized curves for road-traffic annoyance. However, the percentages annoyed *outdoors* (Fig. 5, upper panels), obtained before and after the barrier was built, were much greater than predicted by the synthesized curves, for levels >50 dB L_{den}. These comparisons indicate that the "general annoyance" assessed in previous studies would correspond to indoor annoyance. This conclusion is indirectly supported by Australian studies in which aircraft annoyance inside dwellings was found to have a stronger relationship with exposures outdoors than indoors (Job, Bullen, and Burgess, 1991). Our findings suggest that the synthesis curves are not appropriate for predicting annoyance in situations where the outdoor environment is of particular interest, such as residential areas with cultivated gardens. Of course, definite conclusions cannot be drawn on the basis of a single study. However, our annoyance comparison further indicates that the noise-barrier reductions in sound level actually reduced the indoor annoyance to an extent that it approximately adhere to the Miedema and Oudshoorn's (2001) synthesis curve.

Second, the Miedema and Oudshoorn's (2001) synthesis was based on cross-sectional studies in which different geographical sites with steady-state exposures were compared. In contrast, the present longitudinal study compared the same site before and after a change in road-traffic noise exposure. Previous longitudinal research on traffic volume suggests that when a decrease (or increase) in noise exposure occurs, the *change* in percentage annoyed residents is greater than would be predicted from steady-state circumstances. The present study does not give full support to this hypothesis. For annoyance indoors, the percentages annoyed residents agreed with the *absolute values* derived from the synthesized curves (Fig. 5, lower panels). Consequently, the observed *changes* in percentage annoyed residents indoor was also consistent with the synthesized curves. For annoyance outdoors, the percentages annoyed were clearly higher than the *absolute values* derived from the synthesized curves, both in the before-study and the after-study. However, the difference between observed and predicted *change* was not very dramatic. In the group of residents closest to the road, %A and %HA was reduced with 25 and 35 percentage points, respectively (cf. Fig. 5, upper panels, data points >60 dB L_{den}). The synthesized curve predicts only slightly lower reductions, 15 and 20 percentage points, respectively. Thus, unlike previous studies on traffic-volume-change (e.g., Brown, 1987; Brown, Hall, and Kyle-Little, 1985; Griffiths and Raw, 1986, 1989), the present noise-barrier study did not show an excessive effect on annoyance. This agrees with Griffiths and Raw's (1986) suggestion that reduction of noise at the ear (e.g., noise barrier) may not have the same effectiveness as changes of noise at the source (e.g., traffic-volume reduction).

Other factors than sound-level reduction have also been shown to affect a barrier's abatement efficiency. For example, there is evidence that a visually occluded sound source would be perceived as slightly louder (Aylor and Marks, 1976) and noisier (Watts, Chinn, and Godfrey, 1999) than when it is fully visible. This would suggest that a barrier's visual shielding would partly counteract the perceived loudness reduction accomplished by a sound-level reduction. Attitudes are another potential factor. For example, traffic-

volume reductions may favorably affect residents' attitudes toward the noise source whereas erection of a barrier does not. Obvious effects of reduced traffic volume are reduced risks for traffic accidents and pollution from car exhausts. Attitudes toward the noise source are recognized as an important factor in noise annoyance (McKennel, 1963; Job, 1988a; Fields, 1993; Miedema and Vos, 1999), and it has specifically been suggested as a primary cause of excessive reactions to changes in traffic exposure (cf. Job, 1988b).

D. Roadside barrier efficiency to reduce annoyance and activity disturbances

It should be observed that, although the barrier did improve the sound environment, severe noise problems remained in the residential area. For example, 46% of the residents living closest to the main road (<25 m) often or very often experienced interference with speech communication outdoors in their gardens, and 20% were often or very often awoken too early because of road-traffic noise (see Table III). These disturbances are not surprising, because the noise exposure continued to be high, >60 dB L_{den}. Such values surpass the current guideline value of The World Health Organization (WHO, 2000) based on *daytime* and *evening* annoyance in residential areas, that is, 50 dB or 55 dB $L_{Aeq,16h}$, for moderate or serious annoyance, respectively.

V. CONCLUSIONS

The following conclusions were drawn from the longitudinal noise study of a conventional reflective noise barrier:

(1) The noise barrier reduced residents' noise annoyance *outdoors* and *indoors* up to distances of ~100 m to the main road.
(2) The noise barrier reduced residents' disturbances of speech communication *outdoors* up to distances of ~75 m to the main road. Notably, disturbances of speech communications *indoors* were only reported by a small number of residents before as well as after the noise barrier.
(3) For road-traffic noise exposures >50 dB L_{den}, the percentage of residents who were annoyed or highly annoyed *outdoors* was much larger than predicted from Miedema and Oudshoorn's (2001) synthesized exposure-response curves. The extent of annoyance *indoors, after* the barrier had been erected, agreed better with the synthesized curve. Thus, the synthesized curve would not be applicable for situations where outdoor annoyance is at focus (e.g., residential areas with cultivated gardens).
(4) The noise barrier's shielding efficiency was ~7 dB. Because of high initial exposures (70 dB L_{den} at facades closest to the road), the road-traffic noise exposures of residents, remaining after the barrier was built, still surpassed current WHO (2000) guideline values for annoyance in residential areas. It is, therefore, not surprising that serious noise problems prevailed, despite the noise barrier.

ACKNOWLEDGMENTS

This research was sponsored by grants from the Swedish Foundation for Strategic Environmental Research (MISTRA), the Swedish National Road Administration, and the Swedish Agency for Innovation Systems (Vinnova). We are grateful to Margareta Grandin, Swedish National Road Administration, and Magnus Lindqvist, the Stockholm City, for their valuable contribution to this noise-barrier intervention study. We would also like to thank the following persons for their assistance: Östen Axelsson, Elisabet Borg, Claire Anne Eriksen, Owe Gustavsson, Åsa Lindgren Dawidsson, Jonas Rosenström, Britth Sandin, and Carola Åberg.

Aylor, D. E., and Marks, L. E. (**1976**). "Perception of noise transmitted through barriers," J. Acoust. Soc. Am. **59**, 397–400.

Berglund, B., and Nilsson, M. E. (**2001**). "Variation in perceived soundscape due to shielding building and facade," in *Inter-Noise 2001*, edited by R. Boone (Maastricht, The Netherlands, Nederlands Akoestisch Genootschap), Vol. **3**, pp. 1253–1256.

Berglund, B., and Nilsson, M. E. (**2002**). "Soundscapes perceived indoors and outdoors at quiet and noisy sides of apartment buildings," in *Inter Noise 2002*, edited by A. Selamet, R. Singh, and G. C. Maling, Paper N488 (Marston Hall Ames, IA, Institute of Noise Control Engineering of the USA (INCE)). [Available on CD-ROM.]).

Brown, A. L. (**1987**). "Responses to an increase in road traffic noise," J. Sound Vib. **117**, 69–79.

Brown, A. L., Hall, A., and Kyle-Little, J. (**1985**). "Responses to a reduction in road traffic noise exposure," J. Sound Vib. **117**, 69–79.

CEC. (1996). *Green Paper of the European Commission: Future Noise Policy*, COM(96) 540 final (Brussels, European Commission).

EC. (2002). "Directive 2002/49/EC of the European Parliament and the Council of 25 June 2002 relating to The Assessment and Management of Environmental Noise," *Official Journal of the European Communities*, L 189/12, 18.7.2002. [http://www.europa.eu.int/comm/environment/noise]

Fidell, S., Barber, D. S., and Schultz, T. J. (**1991**). "Updating a dosage-effect relationship for prevalence of annoyance due to general transportation noise," J. Acoust. Soc. Am. **89**, 221–233.

Fidell, S., Silvati, L., and Haboly, E. (**2002**). "Social survey of community response to a step change in aircraft noise exposure," J. Acoust. Soc. Am. **111**, 200–209.

Fidell, S., and Teffeteller, S. (**1981**). "Scaling the annoyance of intrusive sounds," J. Sound Vib. **78**, 291–298.

Fields, J. M. (**1993**). "Effect of personal and situational variables on noise annoyance in residential areas," J. Acoust. Soc. Am. **93**, 2753–2763.

Griffiths, I. D., and Raw, G. J. (**1986**). "Community and individual response to changes in traffic noise exposure," J. Sound Vib. **111**, 209–217.

Griffiths, I. D., and Raw, G. J. (**1989**). "Adaptation to changes in traffic noise exposure," J. Sound Vib. **132**, 331–336.

Horonjeff, R. D., and Robert, W. E. (1997). *Attitudinal Responses to Changes in Noise Exposure in Residential Communities*. (NASA/CR-97-205813, Hampton, Virginia: Langely Research Center).

I-INCE. (1999). "Technical assessment of the effectiveness of noise walls (I-INCE Publication 99-1, Final report)," Noise/News International, **7**, 137–161.

Job, R. F. S. (**1988a**). "Community response to noise: A review of factors influencing the relationship between noise exposure and reaction," J. Acoust. Soc. Am. **83**, 991–1001.

Job, R. F. S. (**1988b**). "Over-reaction to changes in noise exposure: The possible effect of attitude," J. Sound Vib. **126**, 550–552.

Job, R. F. S., Bullen, R. B., and Burgess, D. H. (**1991**). "Noise-induced reaction in a work community adjacent to aircraft runways: The Royal Australian Airforce." in *Inter-Noise 1991*, edited by A. Lawrence (Poughkeepsie, NY, Noise Control Foundation), pp. 895–898.

Kastka, J., Buchta, E., Ritterstaedt, U., Paulsen, R., and Mau, U. (**1995**). "The long term effect of noise protection barriers on the annoyance response of residents," J. Sound Vib. **184**, 823–852.

Kastka, J., and Paulsen, R. (1979). *Undersuchung über die Subjective und Objective Wirksamkeit von Schallschutz-einrichtungen und ihre Nebenwirkung auf die Anlieger* (Forschungsbereicht für den Bundesminister für Verkehr).

Lambert, R. F. (**1978**). "Experimental evaluation of a freeway noise barrier," Noise Control Eng., **Sep-Oct**, 86–94.
McKennel, A. C. (1963). *Aircraft Noise Annoyance around London (Heathrow)*, U.K. Government Survey Report SS. 337 (London, British Central Office of Information).
Miedema, H. M. E., and Oudshoorn, C. G. M. (**2001**). "Annoyance from transportation noise: Relationships with exposure metrics DNL and DENL and their confidence intervals," Environ. Health Perspect. **109**, 409–416.
Miedema, H. M. E., and Vos, H. (**1998**). "Exposure-response relationships for transportation noise," J. Acoust. Soc. Am. **104**, 3432–3445.
Miedema, H. M. E., and Vos, H. (**1999**). "Demographic and attitudinal factors that modify annoyance from transportation noise," J. Acoust. Soc. Am. **105**, 3336–3344.
Nilsson, M. E., and Berglund, B. (2002). "Upplevelse av bullerskyddsåtgärder" [Perception of noise abatement efficiency]. Final Report to the Swedish National Road Administration. Archives of the Center for Sensory Research, **7**, 1–40. (In Swedish).
Öhrström, E. (**2004**). "Longitudinal surveys on effects of changes in road traffic noise-annoyance, activity disturbance, and psycho-social well-being," J. Acoust. Soc. Am. **115**, 719–729.
Raw, G. J., and Griffiths, I. D. (**1985**). "Adaptation to changes in traffic noise exposure," J. Sound Vib. **101**, 273–275.
Schultz, T. J. (**1978**). "Synthesis of social surveys on noise annoyance," J. Acoust. Soc. Am. **64**, 377–405.
Swedish EPA. (1996). *Vägtrafikbuller. Nordisk beräkningsmodell, reviderad 1996 [Road-traffic Noise. The Nordic Calculation Method, revised 1996]*, Report 4653 (Stockholm, Swedish Environmental Protection Agency).
Watts, G. (**2002**). "Barrier designs to reduce road traffic noise," Proceedings of the Institution of Civil Engineers-Transport, **153**, 79–86.
Watts, G., Chinn, L., and Godfrey, N. (**1999**). "The effects of vegetation on the perception of traffic noise," Appl. Acoust. **56**, 39–56.
WHO. (**2000**). *Guidelines for Community Noise* (World Health Organization, Geneva).

Relaxation of sound fields in rooms of diffusely reflecting boundaries and its application in acoustical radiosity simulation

Honghu Zhang[a)]
Architectural Department, South China University of Technology, Wushan Guangzhou, Guangdong 510641, China

(Received 28 May 2005; revised 3 January 2006; accepted 7 January 2006)

The acoustical radiosity method is a computationally expensive acoustical simulation algorithm that assumes an enclosure with ideal diffuse reflecting boundaries. Miles observed that for such an enclosure, the sound energy decay of every point on the boundaries will gradually converge to exponential manner with a uniform decay rate. Therefore, the ratio of radiosity between every pair of points on the boundaries will converge to a constant, and the radiosity across the boundaries will approach a fixed distribution during the sound decay process, where radiosity is defined as the acoustic power per unit area leaving (or being received by) a point on a boundary. We call this phenomenon the "relaxation" of the sound field. In this paper, we study the relaxation in rooms of different shapes with different boundary absorptions. Criteria based on the relaxation of the sound field are proposed to terminate the costly and unnecessary radiosity computation in the later phase, which can then be replaced by a fast regression step to speed up the acoustical radiosity simulation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2169920]

PACS number(s): 43.55.Br, 43.55.Ka [JBS] Pages: 2189–2200

I. INTRODUCTION

Originally developed for illumination and thermal engineering applications and later adopted for computer graphics, the *radiosity method* models the propagation of radiant flux (light) in environments with ideal diffuse reflecting boundaries.[1] With a few minor changes, it can be adapted to model the propagation of acoustic power in similar environments.[2]

As shown by Bot and Bocquillet,[3] the acoustical radiosity method is similar to ray-tracing methods that ignore the phase of the acoustic waves (which is a valid assumption when the wavelength is much less than the environment dimensions). As a historical note, the method's use in acoustics also predates its use in computer graphics.[2,4–6]

For some environments with simple geometries, acoustical radiosity calculations can be carried out analytically.[7–11] For example, Carroll *et al.* gave an analytical solution in a spherical geometry,[7] and Kuttruff showed a solution for an environment composed of two infinite parallel planes.[11] For more complex environments, however, the equations need to be solved numerically by meshing boundaries into surface elements or patches of a finite area.[2,12–15] Depending on the number of surface elements, this can be computationally expensive.

The primary difference between the radiosity method for radiant flux transfer and the acoustical radiosity method is the introduction of a time factor for the finite speed of sound, which makes acoustical radiosity time-dependent[16,17] and even more computationally expensive. There is, therefore, a need to reduce the computational load and improve the calculation time wherever possible.

One approach is to reduce the computations required to simulate the "reverberation tail." For example, Rougeron *et al.* gave an average method to estimate the radiosity in the later phase of energy decay instead of calculating it explicitly.[17] There is, however, a critical question behind such methods, which, to our knowledge has not been previously discussed in the literature: When does the energy decay enter into the later phase? We present in this paper a solution that is based on Miles' theory.[2]

Miles observed that if the boundaries of an enclosure are ideal diffuse reflectors, the sound energy decay at all points on the boundaries will gradually approach, in the later phase, an exponential manner under a uniform decay rate, regardless of the initial energy distribution supplied by the source(s). In other words, the exponential manner and the uniform rate of the decay are intrinsic properties of the geometry and absorption of the enclosure, which cause the ratio of radiosity between every pair of points on the boundaries to be a constant and the relative distribution of radiosity across the environment to become fixed. Therefore, an enclosure can be viewed as a transformer or an operator that transforms the initial arbitrary energy distribution within it into the final relative radiosity distribution with exponential decay at a uniform rate. We say that the sound field system is "relaxed" when the radiosity distribution sufficiently approaches this final state. The numerical radiosity calculations can then be terminated because they provide no further useful information.

Several authors have presented iterative methods to calculate the decay rate.[18,19] Among those, Kuttruff proposed a simple scheme by which the normalized radiosity distribution on the boundaries and uniform later decay rate can be obtained simultaneously.[19] He applied this scheme to calculate reverberation time for several geometrically simple enclosures.

[a)]Electronic mail: zhanghonghu@yeah.net

The problem is, therefore, to decide the time when the system has been sufficiently relaxed. In this paper, we study the relaxation properties of several room geometries and propose strategies to divide the simulation of sound decay into two phases based on the relaxation of the sound field. The radiosity method is only applied in the early phase. In the later phase, a simple and fast regression scheme is adopted to replace the expensive and unnecessary radiosity calculations.

This paper is organized as follows. Section II briefly outlines the acoustical radiosity method, with more detailed presentations to be found in literature.[20,21] Section III establishes parameters to measure the degree of the sound field relaxation, while Sec. IV examines the relaxation properties of rooms with different shapes and boundary absorptions. In Sec. V, stopping criteria for the radiosity calculations are studied.

II. ACOUSTICAL RADIOSITY

Acoustical radiosity method is based on geometrical acoustics. Sound energy is assumed to propagate in space like particles or rays at sound speed c, while the wave phases are ignored. Furthermore, the boundaries of enclosure are assumed to be purely diffusely reflective, so that their reflection coefficients are independent of the incidence angle.

The rate at which the energy leaves from the unit area of the boundary is defined as *radiosity*[21] (or "radiation density" by some authors[4,19,21]). We denote the radiosity of the infinitesimal area at point x and time t by $B(x,t)$. The radiosity of an infinitesimal boundary element is the sum of the contribution of the sound source(s) and the contribution of other wall elements of the enclosure, which is represented by the *radiosity equation*:

$$B(x,t) = \int_S K(x',x)B[x',t-R(x,x')/c]\rho(x)ds' + \rho(x)E(x,t), \tag{1}$$

where $R(x,x')$ is the distance between boundary points x and x', c is the sound speed, $\rho(x)$ is the reflection coefficient of the surface at point x, $E(x,t)$ is contributed by sound source(s), and $\rho(x)E(x,t)$ is called the *emittance*.

In room acoustics simulations, the object of interest is generally the impulse response, and so the emittance comes from the contribution of an impulsive point source. ds' denotes the infinitesimal area of the surface element at point x', while $K(x,x')$ is the ratio of the energy radiated from the boundary element at point x' that reaches the boundary element at x to the energy radiated from the boundary element at x'. Under the assumption of the Lambertian reflection at boundaries, we have:

$$K(x,x') = \frac{\cos\theta\cos\theta'}{\pi R(x,x')^2}v(x,x'), \tag{2}$$

where θ and θ' denotes the angles between the line which links x and x' and the surface normal vectors at x and x'. $v(x,x')$ equals 1 if the points x and x' are visible to each other; otherwise it equals 0.

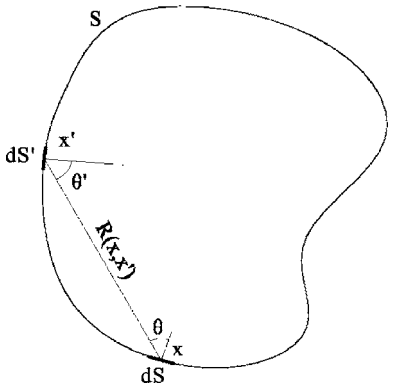


FIG. 1. A simple illustration to help understand the radiosity equation and the related quantities.

The integral is determined for the domain S composed of all boundaries. Figure 1 illustrates some quantities in the radiosity equation. The influence of air absorption is not explicitly shown in Eq. (1). However, if an impulsive source is adopted, the effect of air absorption could be easily compensated.[16,21]

Some other authors also define radiosity (or "irradiation density"[19]) as the rate at which the energy reaches the unit area on the boundaries. Therefore, the radiosity equation assumes another equivalent form as suggested by Kuttruff:[19]

$$B(x,t) = \int_S K(x',x)\rho(x')B[x',t-R(x,x')/c]ds' + E(x,t). \tag{3}$$

Once $B(x,t)$ of all surface elements is known, their contributions to the intensity of a receiver in the room can be easily obtained.[21] Therefore, the primary task of the radiosity calculations is to solve the radiosity equation numerically.

Suppose the boundaries of the enclosure are discretized into N small finite plane elements and time is discretized into sequential equal time intervals tagged by consecutive numbers, and where the radiosity distribution of every element is assumed to be constant. Equation (1) can then be revised as:

$$B_i(n) = \sum_{j=1}^{N} F_{ij}\rho_i B_j(n - R_{ij}/c) + \rho_i E_i(n) \tag{4}$$

where F_{ij} is called *form factor* from element i to element j. It represents the fraction of the energy radiated by element i that reaches element j to all energy radiated by element i. A more detailed discussion on form factor can be found in literature.[22–28]

n represents the nth time interval, or the nth iterative calculation step. The interval is typically several milliseconds, depending on the desired sampling resolution for the acoustical simulation. The energy incident on an element i in the same time interval is summed to determine $B_i(n)$.

Equation (4) shows that $B_i(n)$ involves the calculation of the contributions from the radiosity of all the other elements before the step n, which is already known. Therefore, generally speaking, the computation cost at each step n in the radiosity process has time complexity $O(N^2)$. Step by step,

the radiosity of all N elements can be obtained for each time step n until the radiosity calculations are terminated.

III. SOUND FIELD RELAXATION MEASURES

Miles has observed that after the sound sources stop emitting, the radiosity $B(x,t)$ of all points x on the boundaries will gradually approach an exponential decay $b(x)e^{-\lambda t}$, with λ being constant across all boundaries. Now, we consider the $B(x,t)$ of all surface elements at time t as an infinite dimensional vector and denote it by $B(t)$ and define vector b based on $b(x)$. According to Miles, the vector b, if normalized, will be unique for a given enclosure and independent of the initial sound source(s). In other words, the normalized vector b is independent of the initial energy distribution in the enclosure or the state of the early phase of the sound decay. Miles' argument can be rewritten as follows: In the later phase of the sound decay, the vector $B(t)$ will become more and more parallel to the vector b.

We call angle Ψ_t between the vector $B(t)$ and b or angle Ψ_π between two vectors $B(\tau)$ and $B(t)$ the *relaxation angle*. Several parameters based on the relaxation angle can be observed during the sound decay process and established as measures of the relaxation of sound fields.

We first consider the cosine of the relaxation angle:

$$\mu(t) = \cos \Psi_t = \frac{\langle B(t), b\rangle}{\|B(t)\| \cdot \|b\|} \to 1, \quad t \to \infty, \tag{5}$$

$$\mu(\tau,t) = \cos \Psi_\pi = \frac{\langle B(\tau), B(t)\rangle}{\|B(\tau)\| \cdot \|B(t)\|} \to 1, \quad \tau,t \to \infty, \tag{6}$$

where $\langle\cdot,\cdot\rangle$ is the inner product of two vectors and $\|\ \|$ is the corresponding norm as usually defined in inner product space:

$$\langle B(\tau), B(t)\rangle = \int_S B(x,\tau)B(x,t)ds, \tag{7}$$

$$\|B(t)\|^2 = \int_S B(x,t)^2 ds. \tag{8}$$

If the vectors $B(\tau)$, $B(t)$, and b become coplanar—in inner product space, coplanar means the three vectors are linearly dependent, or one of them could be linearly expressed by the others—then there is a relationship between the parameters $\mu(t)$, $\mu(\tau)$, and $\mu(\tau,t)$:

$$\Psi_\pi = \arccos[\mu(\tau,t)] = |\Psi_\tau - \Psi_t| = |\arccos[\mu(\tau)] - \arccos[\mu(t)]|. \tag{9}$$

The corresponding discrete forms are:

$$\mu(n) = \cos \Psi_n = \frac{\langle\ B(n), b)\rangle}{\|B(n)\| \cdot \|b\|} = \frac{\sum_{i=1}^N B_i(n)b_i}{\sqrt{(\sum_{i=1}^N B_i(n)^2)(\sum_{i=1}^N b_i^2)}} \to 1, \quad n \to \infty, \tag{10}$$

TABLE I. Configuration of rooms.

Rooms	Size (m)	Volume (1000 m^3)	Surface meshing	Elements number
Long	$60\times4\times4$	0.96	$30\times3\times3$	258
Flat	$60\times60\times4$	14.4	$30\times30\times3$	2160
Normally shaped	$60\times30\times8$	14.4	$30\times15\times6$	1440
Cubic 1	24.33^3	14.4	10^3	1000
Cubic 2	60^3	216	20^3	2400

$$\mu(n,m) = \cos \Psi_{nm} = \frac{\langle B(n), B(m)\rangle}{\|B(n)\| \cdot \|B(m)\|} = \frac{\sum_{i=1}^N B_i(n)B_i(m)}{\sqrt{(\sum_{i=1}^N B_i(n)^2)(\sum_{i=1}^N B_i(m)^2)}} \to 1, \quad n,m \to \infty. \tag{11}$$

In addition, Eq. (9) has another discrete form:

$$\Psi_{nm} = \arccos[\mu(n,m)] = |\Psi_n - \Psi_m| = |\arccos[\mu(n)] - \arccos[\mu(m)]| \tag{12}$$

We will use the properties of relaxation angle, or parameters related to it, to measure the degree of the relaxation of the sound field. For example, when the sound field gets more and more relaxed in the later phase of sound decay, the cosine of the relaxation angle, μ, will approach unity.

It should be noted that the parameter μ [in the paper, $\mu(t)$, $\mu(n)$, or their approximations are sometimes denoted by μ for simplicity] itself is not a good measure of the relaxation, as it does not approach unity monotonically. One reason for this is that the distances between pairs of boundary elements are different. In other words, the energy exchanges between these pairs are not synchronized. Some part of the energy may be frequently transformed or operated by the enclosure, while other parts may not. The convergence of the system will, therefore, be perturbed by the less-frequently transformed part of the energy. For example, in a long room, the energy traveling along the length dimension will tend to perturb the system's convergence. Therefore, drops and oscillations will tend to occur in the μ curve during the sound decay process.

Strictly speaking, if the μ parameter from time t to ∞ is always greater than a constant μ_0, we say that "*at time t the sound field is relaxed above the level* μ_0." Such a constant μ_0, being smaller than μ value at any time after t, definitely has a least upper bound $\bar{\mu}$, which is equal to the greatest lower bound of the μ value from time t to ∞, increases monotonically with time, and so is a good indication of the relaxation. In practice, however, it may be difficult to declare at what μ_0 level (except for $\mu_0=0$) the sound field is relaxed above at the time t, so we need to estimate the μ_0 level.

IV. RELAXATION FEATURES IN DIFFERENT ROOMS

The relaxation properties of the sound fields in four types of rectangular rooms are studied. Table I shows the room configuration details, including the size, volume, sur-

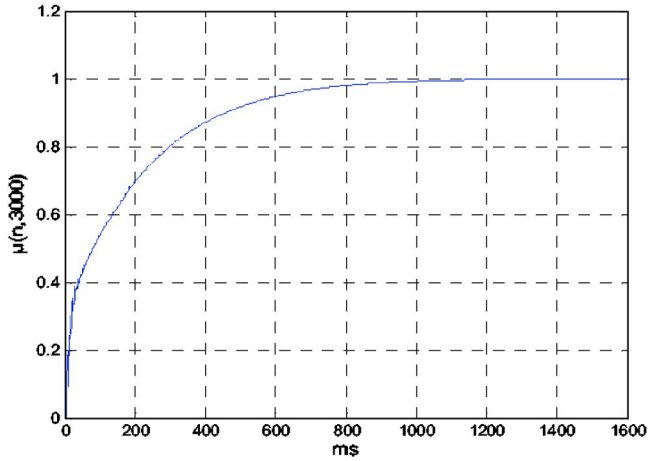


FIG. 2. (Color online) The parameter $\mu(n,3000)$ for the long room, $\alpha=0.15$.

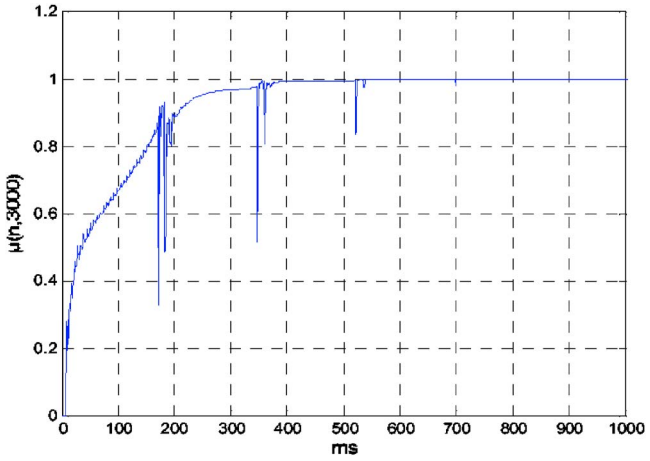


FIG. 3. (Color online) The parameter $\mu(n,3000)$ for the long room, $\alpha=0.35$.

face meshing, and total number of boundary elements. The size and surface meshing of the rooms are expressed with the format length × width × height. For example, the long room has the size $60\times4\times4$ meter and thus its volume is 960 m^3. The surface meshing of the long room is $30\times3\times3$, which means the two surfaces along the length and width direction are equally meshed into $30\times3=90$ elements, the two surfaces along the length and height direction are also meshed into $30\times3=90$ elements, and the two surfaces along the width and height directions are equally meshed into $3\times3=9$ elements. Since the boundaries are meshed into rectangular surface elements that are either parallel or perpendicular to each other, the form factor between each pair of elements is calculated analytically.[2] For convenience, the time interval for each step of plane response calculation is set to one millisecond. An omni-directional point source is located at (2, 2, 2) in all rooms.

Cases which have uniformly and nonuniformly distributed boundary absorption coefficients are studied. For conciseness, only the cases having uniformly distributed boundary absorption coefficients are presented in this paper. Three absorption coefficients, 0.15, 0.35, and 0.55 are considered.

A. Long room

Case 1: Figure 2 shows the value of $\mu(n,3000)$, as an approximation of the value $\mu(n)=\mu(n,\infty)$, in the long room when the boundaries are uniformly assigned absorption coefficient 0.15. In this case, μ increases monotonically and smoothly from 0 to 1.

Case 2: Figure 3 shows the μ curve of the case in the long room when absorption coefficient rises to 0.35. The μ value presents a very similar trend to that of Case 1 (Fig. 2), except for three drops that occur at about $t=175$, 360, and 523 ms. These drops are due to the sound energy which is reflected from the walls at the far end of the room and travels along the room length. Note that $t=175$ ms corresponds to the distance between the sound source to the furthest back wall, or it is the time when the emittance, which means the external disturbance of the system, ceases. For comparison, we changed the absorption coefficient of the back wall to 1.0 while keeping the absorption coefficient of the remaining walls unchanged at 0.35. The result is shown in Fig. 4. Although the back wall has a very small area ($4\times4=16$ m^2) compared with the total room area of 992 m^2, the effect is obvious: The time to relaxation decreased (or conversely, the *relaxation speed* increased). Also, the drops in the μ curve in Fig. 3 disappeared, leaving only a small oscillation at about $t=175$ ms.

Case 3: Figure 5 shows the case of the long room with absorption coefficient $\alpha=0.55$. For precision, $\mu(n,6000)$ is calculated to approximate $\mu(n)$. In this irregularly shaped and highly absorptive room, the sound field relaxes very slowly. In addition, there are periodic drops in the μ curve. The time intervals between every two sequential drops are in accordance with the length of the room. Because the sound energy traveling across the short dimensions of the room is more highly attenuated upon reflection, the disturbance due to the energy traveling along the long dimension is even greater. Regardless, the amplitude of the drops decreases and the oscillation of the μ curve becomes less with increasing time.

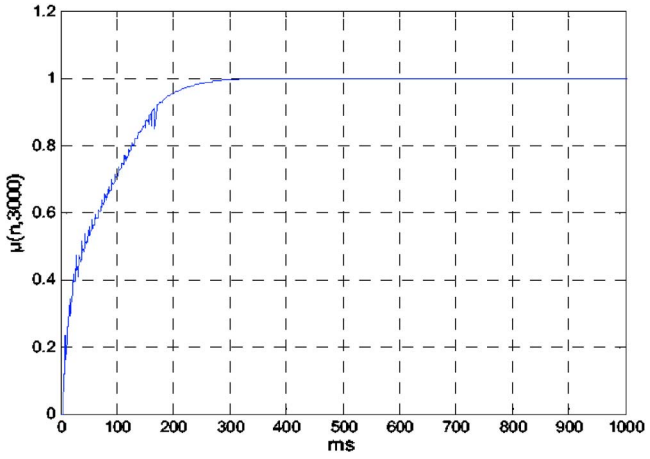


FIG. 4. (Color online) The parameter $\mu(n,3000)$ of the comparison case in the long room.

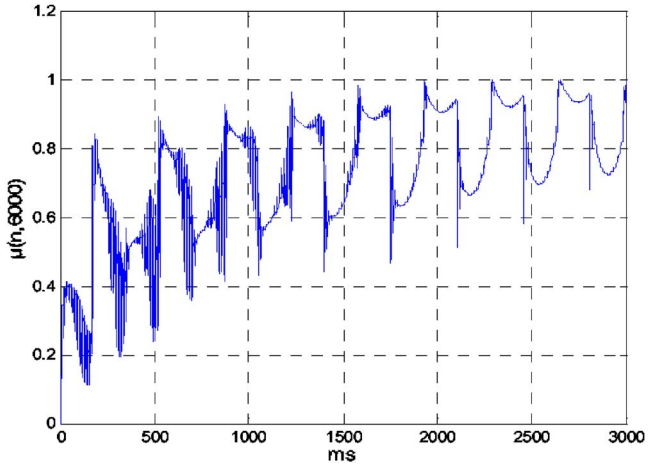


FIG. 5. (Color online) The parameter $\mu(n,6000)$ for the long room, $\alpha=0.55$.

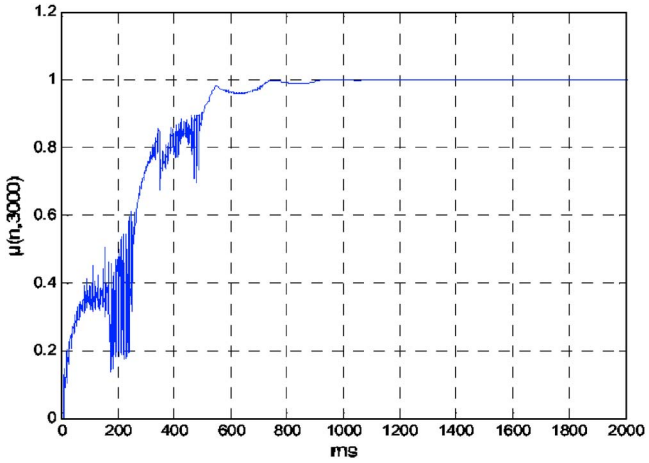


FIG. 7. (Color online) The parameter $\mu(n,3000)$ for the flat room, $\alpha=0.55$.

B. Flat room

Case 4: This case is the flat room with boundary absorption coefficient $\alpha=0.15$. Although its volume is 15 times larger than the volume of the above long room, the relation speed of the sound field in this room is greater than that of the long room. For example, after $t=321$ ms, $\mu(n,3000)$ is over 0.90, while in Case 1 (long room, $\alpha=0.15$), the $\mu(n,3000)$ value does not exceed 0.90 until $t=459$ ms. The μ curve (not shown) is also quite smooth.

Case 5: This case is the flat room, but with $\alpha=0.35$. Figure 6 shows the μ curve for this case, where we can see severe oscillation around the time when the emittance from the far walls perturbs the system. The length of time for the oscillation is longer than that of Case 2 (long room, $\alpha=0.35$) because of the room geometry. Afterwards, the curve is quite smooth.

Case 6: Figure 7 shows the μ curve of the case of the flat room with $\alpha=0.55$. The sound field relaxes more slowly, with $\mu(n,3000)$ not exceeding 0.9 until $t=500$ ms. The periodicity of the oscillations also becomes somewhat stronger. There are another two drops with severe oscillations after the emittance ceases. However, compared with the μ curve of Case 3 (long room, $\alpha=0.55$) in Fig. 5, where the periodic drops appear for a long time, the μ curve in this flat room with high boundary absorption is quite smooth. Consequently, the relaxation speed is considerably greater than that for long room.

C. Normally shaped room

Case 7: This case is the normally shaped room with $\alpha=0.15$. Even though this room has the same volume as the flat room, it relaxes more quickly. Its μ curve increases almost monotonically except for a few small drops at, for example, $t=188$ ms. The μ value in this case is larger than that in Case 4 (flat room, $\alpha=0.15$) almost everywhere. After $t=225$ ms, $\mu(n,3000)$ is over 0.9.

Case 8: This is the case where the boundary absorption coefficient $\alpha=0.35$. After about $t=188$ ms, the μ value jumps very quickly to near its maximum value. The relaxation speed is greater than Case 5 (flat room, $\alpha=0.35$) with the same boundary absorption coefficient and same volume.

Case 9: Figure 8 shows the case in the normally shaped room with $\alpha=0.55$. The μ value also jumps very quickly to a high level after the emittance runs out, and after then we

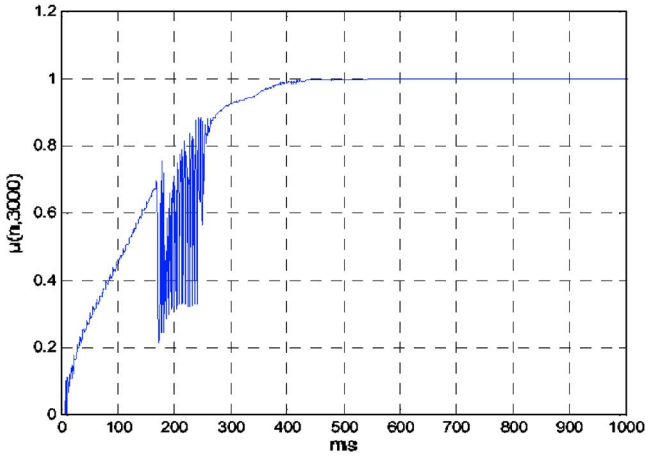


FIG. 6. (Color online) The parameter $\mu(n,3000)$ for the flat room, $\alpha=0.35$.

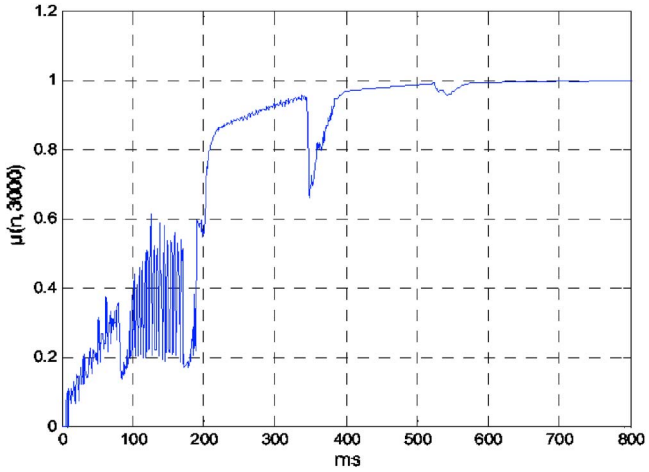


FIG. 8. (Color online) The parameter $\mu(n,3000)$ for the normally shaped room, $\alpha=0.55$.

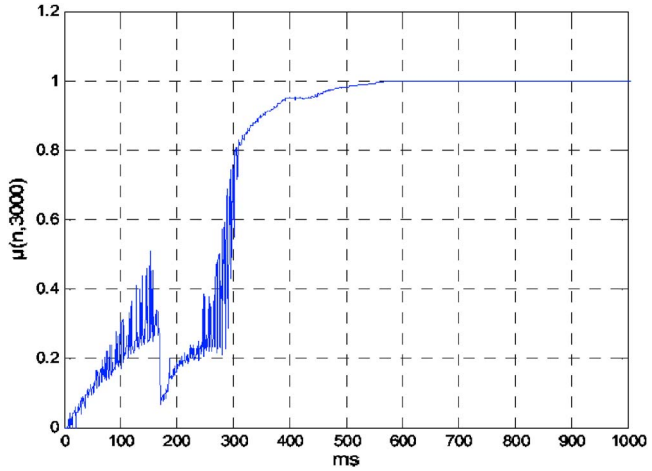


FIG. 9. (Color online) The parameter $\mu(n,3000)$ for cube 2, $\alpha=0.55$.

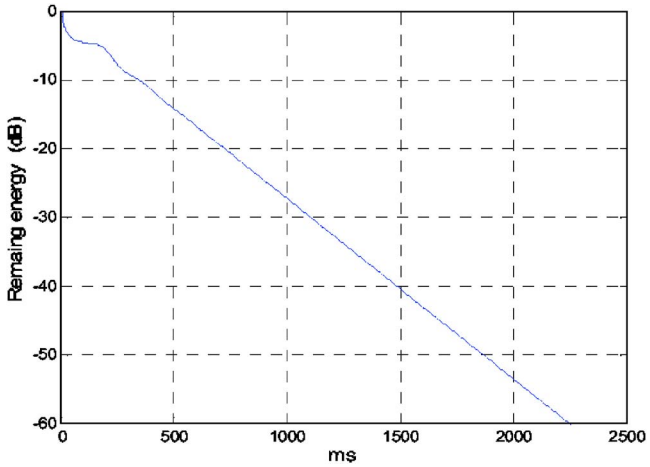


FIG. 10. (Color online) Remaining energy in the cube 2, $\alpha=0.55$.

still see two obvious drops in the μ curve. Comparing the μ curves of this case and Case 6 (flat room, $\alpha=0.55$), it is easy to see that the relaxation speed here is faster than that for the flat room. For example, in this case, the oscillation of the μ curve around the later two drops is much smaller than that for Case 6.

D. Cubes

We also studied the relaxation properties in two cubes with three absorption coefficients 0.15, 0.35, and 0.55, which are denoted as case 10–15, respectively. The results indicated that, for cube 1, the relaxation speed is faster than the irregularly shaped rooms of the same volume, the flat and the normally shaped room. And the relaxation speed for cube 1 is also faster than that in cube 2, as the latter cube has a larger volume. As for the cube 2, which has the same maximum dimension 60 m as the first three rooms (the long, flat, and the normally shaped room), the relaxation speed for this more regularly shaped room is still generally faster. For conciseness, we only show the μ curve for Case 15 (cube 2, $\alpha=0.55$) in Fig. 9.

Case 15: The furthest distance between the points on the boundaries to the sound source is about 100 meters. It takes about 300 ms for sound to travel between these points. We can see in Fig. 9 that after about 300 ms, the oscillation of the μ curve is quickly damped, and the μ value rises up to a high level. After about $t=348$ ms, the μ value is greater than 0.9. Figure 10 shows the remaining energy level in room during the sound decay. For this large room with its high boundary absorption coefficient, however, it still takes more than 2200 ms for all of the energy in the room to decay from 0 to -60 dB.

E. Interim summary

We now summarize the relaxation properties determined for the different rooms above:

(1) For all rooms, if the boundaries are quite reflective, for example, when $\alpha=0.15$, the relaxation speed of the sound field is fast, and the μ curve oscillates only slightly. With increasing boundary absorption, the relaxation speed decreases, and the oscillation of the μ curve becomes more severe.
(2) For irregularly shaped enclosures, such as very long or flat rooms, the relaxation speed of the sound field is generally slow, and periodic drops occur in the otherwise monotonically increasing μ curve. As the room shape becomes more regular, the relaxation speed increases and the severity of the periodic drops gradually decreases.
(3) If the emittance at step n on any surface element $\rho_i E_i(n)$ is nonzero, it perturbs the system and causes oscillations in the μ curve. Rapid oscillations are therefore often seen in the early part of the μ curve until emittance runs out (e.g., Fig. 6 and similar figures), and with increasing time the severity of the oscillations will be reduced. The oscillation of μ curve in the regular rooms is generally less severe than that in the irregular ones, especially when the boundary absorption is high. For all rooms, there generally seems a very severe oscillating range around the time when the emittance from the farthest walls to the sound source ceases and then the μ curve rises to a high level.

V. RADIOSITY CALCULATION STOPPING CRITERIA

In most energy-based sound simulation algorithms, such as ray-tracing methods, radiosity methods, and so forth, the remaining energy in system often acts as a criterion to stop the simulation process, which we call the *energy criterion* (EC). Having observed the relaxation properties of the rooms we studied above, we propose several criteria for stopping the radiosity calculations based on the relaxation of the sound field to speed up simulation. We refer to each of these criteria as a *relaxation criterion* (RC).

A. Energy criterion

To date, relatively few authors have published their studies on acoustical radiosity simulation of sound field. Among them, Kang systematically studied certain types of rooms with acoustical radiosity methods.[13–15] In his work, the criterion for terminating the radiosity process was that when the

remaining energy in the room during the simulation process declined to 10^{-6} of its original quantity. Here, we denote the criterion that Kang used as EC60. For reasonably sized enclosures with very absorptive boundaries, this criterion may work well. (Note Case 15 in this paper, where the boundary absorption coefficient is quite high, but it still takes a long time for the EC60 criterion to be satisfied due to its large volume.) However, for enclosures with more reflective boundaries, a comparatively long time will be needed for the energy to decay to this level. If the sound field has already become very relaxed before the EC60 criterion is satisfied, much of the radiosity computation will have yielded relatively little information. If the boundaries are finely meshed, the problem will be even more significant. In practice, this describes many practical engineering problems, including concert and opera halls.

Many acoustical indices, including EDT, T30, and SPL, depend mainly on the early part of the sound decay. If the sound field is sufficiently relaxed at about −30 dB, over half the computational burden (to −60 dB) can be ignored, with the reverberant tail of the sound decay estimated using faster techniques. Thus, using EC to control the simulation process may be costly.

B. Simple relaxation criterion

When the sound field relaxes sufficiently after the emittance ceases, for example, if the relaxation measurement μ reaches a high level and does not oscillate severely, we may consider that the sound decay has entered its later phase, and we can then terminate the radiosity computation. There are three strong arguments for the relaxation criterion:

(1) In practice, there are many rooms that are neither very irregularly shaped nor with a very high surface absorption, where a high relaxation speed is expected.
(2) For many acoustical indices, the precision of the early part of sound decay is more important than that of the later part. In enclosures with diffusely reflecting boundaries, the later part of sound decay provides little useful informative if the later decay rate is known.
(3) In many cases, the speed of the simulation is more critical than its precision. For example, during the architectural design stage, we often need just a good estimate of the sound field to improve the design. There is, therefore, a very practical need to terminate the time-consuming radiosity computation as quickly as possible.

The first problem is how to estimate the μ value at a time interval n. Generally, we cannot use the above $\mu(n,3000)$ or $\mu(n,6000)$ that are obtainable only after at least 3000 or 6000 radiosity calculation steps. We need a measure which can estimate the relaxation degree of the system as early as possible. Therefore, we use the cosine of the relaxation angle between two vectors $B(n)$ and $B(n-k)$, the parameter $\mu(n,n-k)$, to determine whether the radiosity at time interval n is relaxed enough. Although it seems better if k is larger when viewing $\mu(n,n-k)$ as an approximation of μ, in practice, a small k is desired to speed up the computation. It should be noted that while it may be hard for $\mu(n,n-k)$ to approximate the absolute value of μ when k is small, $\mu(n,n-k)$ is a good indicator that reflects the severity of oscillation of μ value.

We check whether the value $1-\mu(n,n-k)$ has, for the first time after the last emittance has been released, continuously been lower than a pre-set threshold 10^{-l}, such as 10^{-5}, for preset p steps. If so, we consider the system to be sufficiently relaxed.

Here we introduce another parameter RC(s) and its discrete form RC(k):

$$\mathrm{RC}(s) = \lg[1 - \mu(t,t-s)], \tag{13}$$

$$\mathrm{RC}(k) = \lg[1 - \mu(n,n-k)]. \tag{14}$$

The parameter RC(k), whose value ranges from 0 to $-\infty$, magnifies the change of $\mu(n,n-k)$, whose value range is from 0 to 1, especially in later phase of sound decay, where $\mu(n,n-k)$ increasingly approaches unity. Therefore, the above relaxation criterion can be rewritten as RC(k) dropping below $-l$ continuously for p steps.

Note that the time complexity of the calculation cost for $B(n)$ at every step is $O(N^2)$, while the calculation of the $\mu(n,n-k)$ value involves only a N-step loop or a time complexity of only $O(N)$. Therefore, the additional computational cost for $\mu(n,n-k)$ is comparatively trivial, especially when elements number N is large.

Since the radiosity on each surface element tends to decay exponentially under the uniform decay rate in the later phase of the sound decay, the sum of the energy on all the surface elements also decays in this way. After the radiosity computation is terminated, the logarithm of the sum of energy on all surface elements at the last p steps is used to do a linear regression to calculate the decay rate. The time varying radiosity of every surface element can then be added as a reverberation tail by extrapolating the data of the last p steps of radiosity calculation using the decay rate obtained above.

C. Comparison of EC and RC

We compared the effect and efficiency of the simple RC with EC60. Here we choose $k=5$, the threshold $1-\mu(n,n-5) \leqslant 10^{-5}$ or RC(5)=−5, and $p=30$.

The termination times of the radiosity calculations decided by the criteria EC60 and RC in the 15 cases are listed in Table II. We divide the space in each room into equal small cubes. At every center point of the small cubes, the impulse response is calculated using radiosity method under two criteria. A regression scheme will be taken if RC stopped the radiosity process early. In the long room, the size of small cubes is $1\times1\times1$ meter, so there are 960 center points in the long room. In the flat and the normally shaped room, the cube size is $2\times2\times2$ meter and there are 1800 points in the two rooms, respectively. Cubes 1 and 2 are also divided into 1000 and 1728 small cubes whose lengths of side are 2.43 and 5 meter, respectively. Acoustical indices EDT, T30, and SPL are calculated based on those impulse responses. In Table II, the maximum and average differences between

TABLE II. Comparison between EC and RC.

Case	Terminating time (ms)		Maximum difference (%)			Average difference (%)			Energy remaining (dB)
	EC60	RC	EDT	RT30	SPL	EDT	RT30	SPL	
1	986	574	0.02	0.24	0.00	0.00	0.18	0.00	−35.1
2	—	—	—	—	—	—	—	—	—
3	—	—	—	—	—	—	—	—	—
4	1880	550	1.75	5.39	0.08	0.70	1.53	0.03	−17.9
5	856	676	0.00	0.16	0.00	0.00	0.03	0.00	−47.5
6	—	—	—	—	—	—	—	—	—
7	2939	505	3.80	2.82	0.29	1.90	0.64	0.12	−10.6
8	1193	422	0.35	0.37	0.00	0.03	0.20	0.00	−21.6
9	726	610	0.28	0.04	0.00	0.00	0.01	0.00	−50.1
10	4144	297	13.9	7.58	0.67	6.93	7.57	0.58	−4.6
11	1624	299	2.69	3.96	0.09	1.45	3.85	0.06	−11.5
12	918	366	0.72	1.68	0.00	0.07	1.28	0.00	−24.2
13	10201	645	2.72	1.55	0.14	1.26	1.55	0.05	−4.3
14	3984	654	17.6	10.0	0.16	3.74	9.83	0.12	−10.7
15	2249	675	4.50	10.8	0.03	1.20	9.87	0.03	−18.8

these indices obtained under two criteria are listed, and the levels of remaining energy in the rooms at the time when the RC is satisfied are also given.

The cases marked by dashed lines in Table II are those where the time required by the RC is more than EC60. In other words, RC is less efficient than EC60 or fails in those cases. For the long room, RC fails when the absorption coefficient is 0.35 or 0.55. (But for the comparison case in the long room, where the back wall is absent or its absorption coefficient is 1, RC still works to give a shorter termination time than EC60, though it is not shown in Table II.) As the room shape becomes more regular, the working range of RC expands. For the flat room, RC works at absorption coefficient 0.35. And for the normally shaped room, which has the same volume as the flat room, RC still works with the high absorption coefficient 0.55. In the two cubes, with absorption coefficient 0.55, the termination time of RC is largely less than that of EC60.

Obviously, if the ratio between the termination times required by RC and EC60 is smaller, the efficiency of RC is higher. From Table II, we can see that as the room shape becomes more regular, the efficiency of RC improves. For example, we compare the cases with low boundary absorption coefficient 0.15. In the long room (Case 1), RC reduces the computation required by EC60 by half, since the former requires the radiosity process to calculate for 574 ms while the later requires 986 ms. In the flat room (Case 4) the ratio is 550/1880=0.29, near one-thirds. In the normally shaped room (Case 7), the ratio is 505/2939=0.17, less than one-fifth. And in the cube 2 (Case 13), the time at which EC60 is satisfied will be a very large value 10 201 ms, while the time satisfying RC is only 645 ms; the ratio amazingly reaches 645/10 201=0.06, meaning that the relaxation speed in such a room is so high that enormous radiosity calculation time savings can be achieved in comparison to using EC60.

We can see that the differences between the acoustical indices obtained with the two stopping criteria are generally small, especially when RC does not terminate the radiosity process too early. In Case 13 (cube 2, α=0.15), even when all the remaining energy in the room declines only to −4.3 dB, RC is ready to terminate the radiosity process and the maximum difference of the acoustical indices at 1728 points between both criteria is less than 3%.

We must note, however, in some cases where the differences of the indices EDT and T30 between the two criteria are fairly large. For example, the maximum difference in EDT in Case 10 (cube 1, α=0.15) is 13.9% and in Case 14 (cube 2, α=0.35), is 17.6%. This is not because the relaxation speeds are low. Rather, it is primarily because of the properties of the sound field in these cases and the definition of these acoustical indices. Figure 11 shows the sound level decay curves at point (2.5, 2.5, 17.5) in Case 14. The curves are calculated from the impulse responses under the two criteria using the Schroeder integral method,[29] where the dash dot curve is obtained with RC and the solid curve is obtained

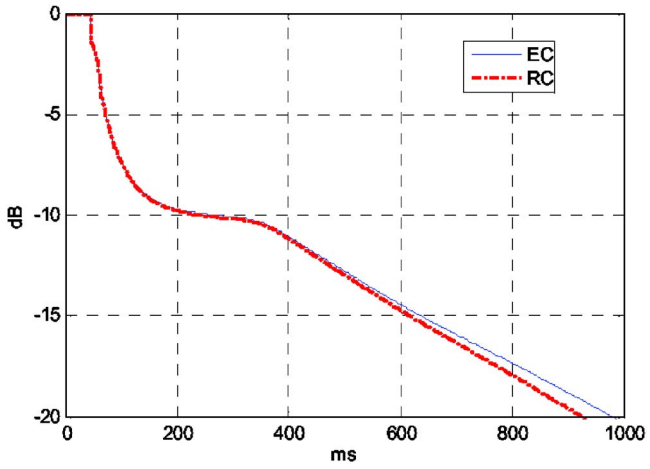


FIG. 11. (Color online) The sound level decay curves at the point (2.5, 2.5, 17.5) in cube 2, obtained under two criteria, α=0.35.

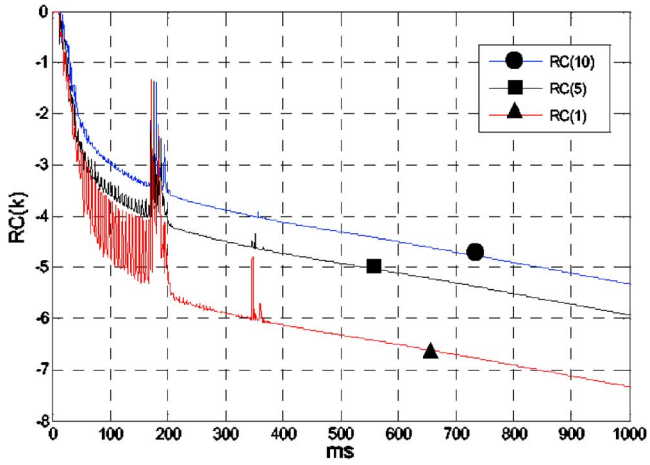


FIG. 12. (Color online) The parameter RC(k) in the long room, α=0.15, k=1, 5, and 10.

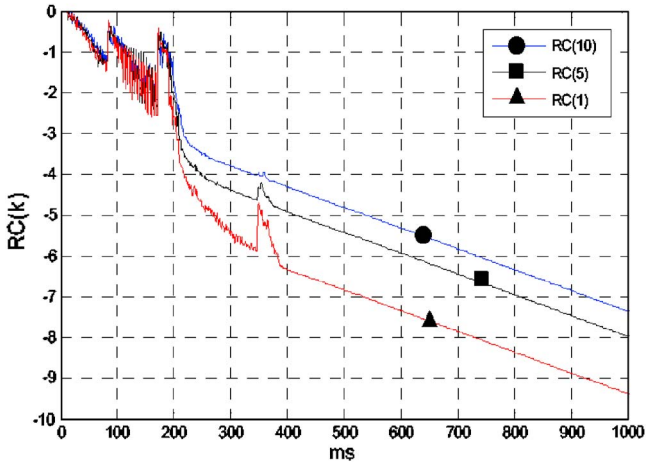


FIG. 13. (Color online) The parameter RC(k) in the normally shaped room, α=0.15, k=1, 5, and 10.

with EC. The curves show that the sound level has quite different decay rates in the early and later parts. Between the two parts, there is a transition that happens to be near the sound level −10 dB. Since EDT depends on the decay rate of sound level from 0 to −10 dB, the shape of the curves here may cause large errors and differences between two curves, even though they are very close to each other. (This means for many other acoustical indices whose calculation does not depend on linear regression of decay curve, for example, Ts, the center time, the differences may be not obvious.)

In such circumstances, a more strict relaxation criterion may, therefore, need to be considered, for example, using a higher threshold to prevent the radiosity calculations from being terminated too quickly.

Since RC will fail in some circumstances where EC is more efficient, another criterion might consist of a combination of RC and EC. That is, the radiosity calculations can be terminated when either RC or EC is satisfied.

D. Relaxation criterion based on relaxation angle decline

The above simple RC involves the selections of a step interval k, the threshold of RC(k), and the step number p. Tactful choices may depend on experience. For example, since we observed periodic drops occurring in the μ curves, it seems a good idea to choose p in such a way that the time span of p steps equals the length of a drop period. In the long room, the length of the dropping period is about 176 ms, which is in accordance with the approximate time for sound traversing the length of the room (60 m). As we have set the length of a step as 1 ms in this paper, it means that p=176 may be a good choice. In the flat room, the normally shaped room, and in Cube 2, the drop periods are even longer than 176 ms. However, a smaller number p=30 still works very well in the cases we studied, when the threshold of RC(k) is set as −5 and k=5.

It is interesting to find some less empirical criteria. Enlightened by the observations of the simulation results, another relaxation criterion based on the decline of the relaxation angles is proposed. First, two important observations are noted.

(1) The RC(k) curves tend to converge to a straight line in the later phase of sound decay, especially obvious in rooms with reflective boundaries. If the boundary absorption coefficient is high, the RC(k) curve may be more complex in the later phase of sound decay, but the linear trend is still obvious. Figure 12 shows the curves of RC(1), RC(5), and RC(10) of Case 1 (long room, α =0.15). The oscillation of curve RC(1) is the most severe. With increasing k, the RC(k) curve becomes smoother.
(2) In the early phase of the sound decay, the three curves in Fig. 12 are neighboring closely. There is a sudden jump at about t=175 ms when the emittance ceases. After this, the three curves begin to then converge to three parallel straight lines.

Similar phenomenon are seen in Fig. 13, which shows Case 7 (normally shaped room, α=0.15). Figures 14 and 15, where the results of Cases 8 and 9 (normally shaped room, α=0.35 and 0.55) are presented, show that when the boundary absorption is high, the shape of the RC(k) curves become

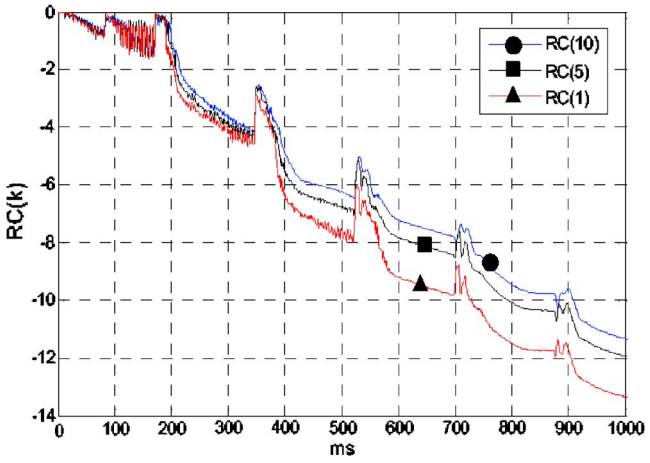


FIG. 14. (Color online) The parameter RC(k) in the normally shaped room, α=0.35, k=1, 5, and 10.

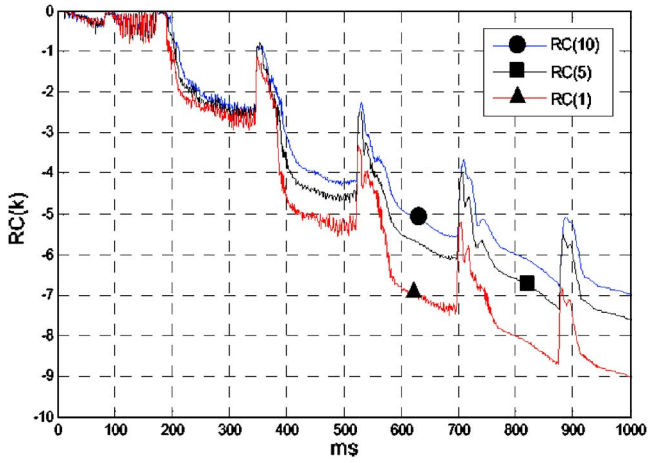


FIG. 15. (Color online) The parameter RC(k) in the normally shaped room, α=0.55, k=1, 5, and 10.

more complex. For example, periodic jumps appear in the RC(k) curves in the later phase of sound decay, while the corresponding μ curves are quite smooth. This is because RC(k) value amplifies the details of $\mu(n,n-k)$. However, the linear trend of the RC(k) curves in the later phase of sound decay is obvious. It is also clear to see the RC(k) curves tend to separate from each other with a constant distance.

If we try to explain the phenomena, two assumptions must be adopted:

(1) Theoretically, it will take an infinitely long time for the relaxation angle to converge to zero. The linear trend of RC(k) curves in the later phase implies that we could assume the decline of the relaxation angle Ψ_t between $B(x,t)$ and $b(x)$ also has an exponential manner when the time is sufficiently long or the sound field has sufficiently relaxed.
(2) Vectors $B(x,t)$, $B(x,t-s)$, and $b(x)$ (where s is a constant) will tend to become more coplanar with increasing time and the sound field has sufficiently relaxed. As we know, if $B(x,t)$, $B(x,t-s)$, and $b(x)$ are normalized, the vector $B(x,t)$, $B(x,t-s)$ will converge to $b(x)$. The assumption means that the trace of $B(x,t)$ converging to $b(x)$ will eventually be along a smooth curve which is continuous and differentiable at point $b(x)$ on the unit hypersphere in the inner product space, or the trace on the unit hypersphere will converge to a fixed great circle of the hypersphere at point $b(x)$, and $B(x,t)$ converges to $b(x)$ along a fixed direction that is the tangent direction of the great circle at point $b(x)$, as illustrated by Fig. 16.

With the above assumptions, the linear and parallel trends of the RC(k) curves in the later phase are easy to explain. We assume:

$$\Psi_t = \arccos[\mu(t)] \to e^{-ht+d}, \quad t \to \infty. \tag{15}$$

In Eq. (15), h is the decline rate of the relaxation angle Ψ_t and d is a constant. We have:

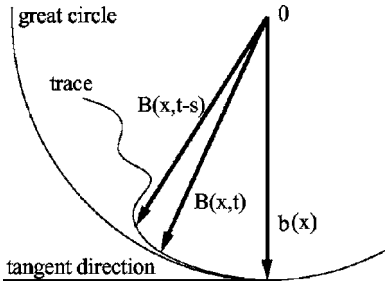


FIG. 16. The way that the normalized radiosity vector $B(x,t)$, and $B(x,t-s)$ converge to their limit $b(x)$.

$$\lg[1-\mu(t)] = \lg(1-\cos\Psi_t) = \lg\left[2\sin^2\left(\frac{\Psi_t}{2}\right)\right]$$
$$\to \lg\left[2\left(\frac{\Psi_t}{2}\right)^2\right] = 2\lg\Psi_t - \lg 2$$
$$\to Ht + D, \quad t \to \infty \tag{16}$$

where $H=-2h\lg e$ and $D=2d\lg e-\lg 2$. Equation (16) shows that $\lg[1-\mu(t)]$ will converge to a linear form when time t is sufficiently large.

Now we consider RC(s)=$\lg[1-\mu(t,t-s)]$, where s is a constant. We have

$$\mathrm{RC}(s) = \lg[1-\mu(t,t-s)] = \lg(1-\cos\Psi_{t,t-s}) = \lg[1$$
$$-\cos(\Psi_t-\Psi_{t-s})] = \lg\left[2\sin^2\left(\frac{\Psi_t-\Psi_{t-s}}{2}\right)\right]$$
$$\to \lg\left[2\left(\frac{\Psi_t-\Psi_{t-s}}{2}\right)^2\right] = 2\lg(|\Psi_t-\Psi_{t-s}|) - \lg 2$$
$$\to 2\lg[e^{-h(t-s)+d} - e^{-ht+d}] - \lg 2 = 2\lg[e^{-ht+d}(e^{hs}$$
$$-1)] - \lg 2 = Ht + D', \quad t \to \infty, \tag{17}$$

where $H=-2h\lg e$ and $D'=2d\lg e+2\lg(e^{hs}-1)-\lg 2$. Equation (17) shows that, when $t\to\infty$, the curve RC(s) will also converge to a straight line, whose rate of slope H depends on h which is the decline rate of relaxation angle Ψ_t, and whose intercept D' depends on h and s. In other words, two curves RC(s) and RC(s') will converge to a pair of parallel straight lines, and the distance between them—which equals $2\lg(e^{hs}-1)-2\lg(e^{hs'}-1)$—depends on, besides the identical h, the different number s and s', but NOT, interestingly, the number d. This conclusion also holds for the discrete form RC(k) and RC(k') of course, as being proved by the computer simulation results.

For example, we call the three straight lines L1, L2, and L3 to which RC(1), RC(5) and RC(10) curves converge in the later phase of the sound decay in Case 1, as shown in Fig. 12. Through linear regression (using the least-square method) of the seemingly straight part of the RC(k) curves, we obtained the rates of slope of these straight lines. The results show they are the same number: −0.0022. That means these three straight lines really are parallel. Then, by use of Eq. (17), the decline rate h of the relaxation angle of Case 1 was obtained, following which the difference between intercepts of L1 and L2 was calculated as 1.4046, and also the

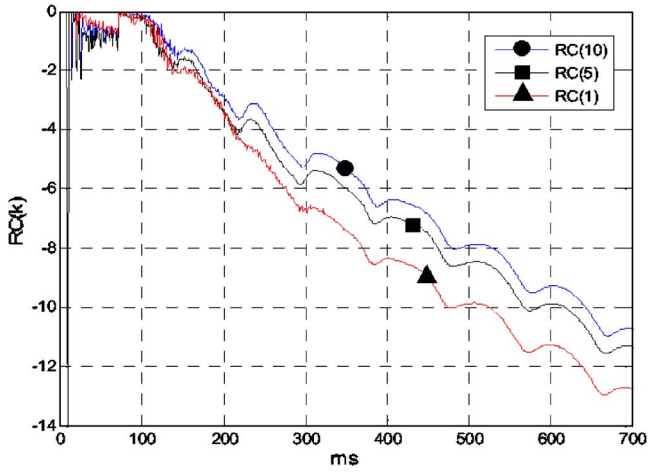


FIG. 17. (Color online) The parameter RC(k) in cube 1, α=0.15, k=1, 5, and 10.

difference between intercepts of L2 and L3 was calculated as 0.6077. These calculated results by Eq. (17) are exactly equal to the results "measured" from the simulated data in Fig. 12. Similar checks were done for other cases. Even in Case 9, where periodic jumps exist in the RC(k) curves, the calculated differences between the intercepts of the straight lines and to the "measured" results were quite small, even with the jump data being included in the linear regression.

In addition, we must report that, while it is true—as we mentioned above—that when the room shape becomes more regular the relaxation speed becomes faster and the μ curve seems smoother in the later phase of sound decay, the RC(k) curve turns out to be less straight and oscillates periodically, as seen, for example, in Fig. 17, which shows RC(1), RC(5), and RC(10) curves of Case 10 (cube 1, α=0.15). It must be noted that, although in the later part these RC(k) curves oscillate periodically, the linear trend is still obvious, and these curves still separate from each other at a constant distance.

We also performed linear regression on the data of later part, for example, the data from 400 to 700 ms, and used Eq. (17) to calculate the distances between these curves. The results correspond with the "measured" distances very well. A possible explanation for the periodic oscillations is that when the room shape becomes more regular, the radiosity in later phase becomes more evenly distributed on all wall elements because of the symmetry of room shape, which causes RC(k) to become more vulnerable to the perturbing energy in the system or the effect of perturbing energy on the system becomes more significant. On the other hand, if the radiosity is unevenly distributed on all the boundary elements, the RC(k) value will be determined more by those boundary elements with larger radiosity values which make the perturbing energy comparatively trivial. (Remember the inner product involved for relaxation angle calculation.)

We also found that in spherical geometries, the oscillations of the RC(k) curves appear more pronounced. Nonetheless, the curves still show obvious linear and parallel trends in the later phase of sound decay.

Finally, while the relationship of RC(k) curves in the early phase of sound decay may be complex, if we compensate for the initial delay of the RC(k) curves, such that the value of RC(k) is zero for the initial k steps, we often observe that the RC(k) curves are very close to each other in the early phase of sound decay, especially when the shape of the room is regular, and then separate from each other in the later phase.

Now we can also use the relationship of several RC(k) curves to measure the relaxation of sound fields. We determine that the sound decay is entering its later phase when the RC(k) curves begin to separate from each other at a gradually stable distance and with a linear trend. These curves can be visualized during the process of radiosity computation such that we can terminate the radiosity process manually according to the observation of the curves, or develop an algorithm for the computer to do it automatically.

VI. CONCLUSION

In this paper, we introduced a concept of relaxation of sound field in rooms with diffusely reflecting boundaries. Parameters used to measure the degree of relaxation of the sound field of the room are suggested. We also studied relaxation properties of rooms ranging from irregularly shaped to regularly shaped with different boundary absorption levels. We found that if the room is more regularly shaped and boundaries more reflective, the relaxation speed is generally faster.

New criteria based on the relaxation of the sound field, which we call *relaxation criteria*, are suggested to reduce the expensive radiosity calculations. We tested the performance of one relaxation criterion and found it to be efficient with regularly shaped rooms having moderate boundary absorption. Tactful application of the relaxation criterion combined with the energy criterion is a good strategy to balance the efficiency and precision of the radiosity calculations.

We assume that the logarithm of relaxation angle tends to decay linearly because the relaxation angle itself tends to decay exponentially, and that the normalized radiosity vector on the boundaries will converge to its limit along a fixed direction. These two assumptions are based on the observations of our computer simulation results in several rooms. Future effort should be given to the theoretical proof of these assumptions.

ACKNOWLEDGMENT

We express deep gratitude to Professor Shuoxian Wu in South China University of Technology for his valuable suggestions for this paper. We also express our deep gratitude to Mr. Ian Ashdown who provided a great deal of assistance in editing this version of the manuscript and also provided many valuable suggestions.

[1]C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile, "Modelling the interaction of light between diffuse surfaces," Comput. Graphics **3**, 213–222 (1984).
[2]R. Miles, "Sound field in a rectangular enclosure with diffusely reflecting boundaries," J. Sound Vib. **92**, 203–226 (1984).
[3]A. Le Bot and A. Bocquillet, "Comparison of an integral equation on energy and the ray-tracing technique for room acoustics," J. Acoust. Soc. Am. **108**, 1732–1740 (2000).
[4]H. Kuttruff, *Room Acoustics*, 4th ed. (Spon Press, London, 2000).

[5]H. Kuttruff, "Nachhall und effektive Absorption in Räumen mit diffuser Wandreflexion," [Reverberation and effective absorption in rooms with diffuse wall reflections], Acustica **35**, 141–153 (1976).
[6]H. Kuttruff, "Simulierte Nachhallkurven in Rechteckräumen mit diffusem Schallfeld," Acustica **25**, 333–342 (1971).
[7]M. M. Carroll and C. F. Chien, "Decay of reverberant sound in a spherical enclosure," J. Acoust. Soc. Am. **62**, 1442–1446 (1977).
[8]M. M. Carroll and R. N. Miles, "Steady-state sound in an enclosure with diffusely reflecting boundary," J. Acoust. Soc. Am. **64**, 1424–1428 (1978).
[9]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).
[10]H. Kuttruff, "Energetic sound propagation in rooms," Acust. Acta Acust. 83, 622–628 (1997).
[11]H. Kuttruff, "Stationäre Schallausbreitung in Flachräumen," [Stationary sound propagation in flat enclosures], Acustica **57**, 62–70 (1985).
[12]T. Lewers, "A combined beam tracing and radiant exchange computer model of room acoustics," Appl. Acoust. **38**, 161–178 (1993).
[13]Jian Kang, "Sound propagation in street canyons: Comparison between diffusely and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[14]Jian Kang, "Comparison of Sound Fields in Regularly-shaped, Long and Flat Enclosures with Diffusely Reflecting Boundaries," International Journal of Acoustics and Oscillation **7**, 165–171 (2002).
[15]J. Kang, "Reverberation in rectangular long enclosures with diffusely reflecting boundaries," Acust. Acta Acust. 88, 77–87 (2002).
[16]J. Shi, A. Zhang, J. Encarnação, and M. Göbel, "A modified radiosity algorithm for integrated visual and auditory rendering," Comput. Graphics **17**, 633–642 (1993).
[17]G. Rougeron, F. Gaudaire, Y. Gabillet, and K. Bouatouch, "Simulation of the indoor propagation of a 60 GHz electromagnetic wave with a time dependent radiosity algorithm," Comput. Graphics **26**, 125–141 (2002).
[18]E. N. Gilbert, "An iterative calculation of auditorium reverberation," J. Acoust. Soc. Am. **69**, 178–184 (1981).
[19]H. Kuttruff, "A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries," J. Acoust. Soc. Am. **98**, 288–293 (1995).
[20]C. Cianfrini, M. Corcione, and D. M. Fontana, "A method for predicting nonuniform steady sound fields within spaces bounded by diffusive surfaces," Appl. Acoust. **54**, 305–321 (1998).
[21]E. M. Nosal, M. Hodgson, and I. Ashdown, "Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms," J. Acoust. Soc. Am. **116**, 970–980 (2004).
[22]I. Ashdown, *Radiosity: A Programmer's Perspective* (Wiley, New York, 1994).
[23]M. F. Cohen and J. R. Wallace, *Radiosity and Realistic Image Synthesis* (Academic, Boston, 1993).
[24]J. R. Howell, *A Catalog of Radiation Configuration-Factors* (McGraw-Hill, New York, 1982).
[25]U. Gross, K. Spindler, and E. Hahne, "Shape-factor equations for radiation heat transfer between plane rectangular surfaces of arbitrary position and size with rectangular surfaces of arbitrary position and size with rectangular boundaries," Lett. Heat Mass Transfer **8**, 219–227 (1981).
[26]M. F. Cohen and D. P. Greenberg, "The hemi-cube: a radiosity solution for complex environments," Comput. Graphics **19**, 31–40 (1985).
[27]J. R. Wallace, K. A. Elmquist, and E. A. Haines, "A Ray tracing algorithm for progressive radiosity," Comput. Graphics **23**, 315–324 (1989).
[28]D. R. Baum, H. E. Rushmeire, and J. M. Winget, "Improving radiosity solutions through the use of analytically determined form-factors," Comput. Graphics **23**, 325–334 (1989).
[29]M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am. **37**, 409–412 (1965).

Effects of the inclination of a rigid wall on the free vibration characteristics of acoustic modes in a trapezoidal cavity

K. S. Sum and J. Pan
School of Mechanical Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, Western Australia 6009

(Received 7 October 2005; revised 19 December 2005; accepted 20 December 2005)

The coupling between rigid-walled modes of a rectangular cavity (RC modes) is used to obtain the shapes and resonance frequencies of rigid-walled modes of a trapezoidal cavity (TC modes) with an inclined rigid wall. A method is established to identify the TC modes, where the modes can be defined to evolve from individual RC modes. The wall inclination generates two coupling mechanisms, namely, the local coupling where the RC modes couple at the inclined wall, and the global coupling where the RC modes couple throughout the trapezoidal volume. The latter arises from the nonorthogonality of the RC modes in the trapezoidal volume. Both couplings are selective that only RC modes with the same number of nodes in the direction perpendicular to the inclination are coupled to each other. For small inclinations, each TC mode possesses the distorted shape of the RC mode that evolves it. When the inclination is increased, the TC-mode shape becomes complicated and unrecognizable, and extrema can also exist in the resonance frequency of the TC mode. These behaviors are determined by the behaviors of the local and global couplings of the RC mode. This paper provides an understanding of how the free vibration characteristics of TC modes change with the inclination and what determines these changes. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2166707]

PACS number(s): 43.55.Br, 43.55.Ka, 43.20.Ks [NX] Pages: 2201–2210

I. INTRODUCTION

A large amount of published literature in acoustics of enclosed spaces were concerned with rectangular and cylindrical enclosures. The shapes and resonance frequencies of rigid-walled modes of these regular enclosures are well defined analytically, and have been used in many previous studies for predicting acoustical properties of enclosed sound fields. These studies include the sound transmission through structures and radiation from structures into rooms, the vibration response of structures backed by air cavities, the sound field response in rooms, and the active control of noise and vibration in enclosures. The relevant works have been reviewed in detail where comprehensive lists of bibliography were given.[1,2] Particularly, the study of free vibration behaviors of sound fields in rectangular[1,3,4] and cylindrical spaces[5–9] has provided an insight into many basic acoustical characteristics of regular enclosures, which include the pressure responses, resonance frequencies, and decay times of their acoustic modes.

However, irregular enclosed spaces are often encountered in practice and these can range from tiny cavities in miniaturized transducers to small size passenger compartments of cars, medium size passenger cabins of boats, and large size halls with complicated geometry. Unlike regular enclosures, the shapes and resonance frequencies of rigid-walled modes of irregular enclosures are not definable analytically. In this case, methods based on direct numerical calculations or finite element and boundary element are usually used to obtain the details. The relevant previous works were devoted to the formulation and validation of numerical models that predict the resonance frequencies and pressure responses of acoustic modes in the enclosures.[10–12]

As sound fields in irregular spaces can be more explicitly quantified and described by analytical means, some research in this area has also been conducted. In an early mathematical work, an analytical technique that uses rigid-walled modes of rectangular enclosures to predict acoustic modes of irregular enclosures was developed.[13] In this technique, each acoustic mode of an irregular enclosure is expressed as the normal expansion of rigid-walled modes of the rectangular enclosure that bounds the irregular enclosure. The technique has been employed in later investigations of acoustic fields in a semicircular cavity and a simplified aircraft cabin,[14] and in trapezoidal cavities with a tilted wall of varying inclination angle[15,16] and with four opposite tilted walls of a fixed inclination angle.[17] However, Refs. 14, 15, and 17 were concerned with the formulation and validation of analytical models for predicting the resonance frequencies of acoustic modes and the pressure responses in the irregular enclosures. In addition, Ref. 15 focused mainly on whether or not the sound field would become diffuse when the wall inclination is increased, and Ref. 16 focused on the total pressure change on a given plane in the cavity when the inclination is altered.

As there is only little reported analytical work on irregular spaces and most of the relevant previous works dealt primarily with the development of methodology as reviewed above, there is still no significant understanding of how acoustic modes in the spaces are defined and how the irregularity affects the acoustical behavior of the spaces. Since the fundamental characteristics of sound fields in irregular enclosures have an enormous practical significance and many

existing analytical models use rigid-walled modes of rectangular enclosures as basis functions, it is essential to first understand two basic acoustical aspects. First, one needs to know how these modes are changed by the irregularity. Second, it is then necessary to understand the way that these modes determine the free vibration behavior of each rigid-walled mode of an irregular enclosure, and how this behavior is related to the sound pressure distribution in the enclosure. The free vibration characteristics are the mode shape and the resonance frequency.

In this paper, a trapezoidal cavity with an inclined rigid wall is considered. All other noninclined walls are also rigid. This irregular shape is chosen because of its wide applications, which include miniaturized transducers.[12,17] Even the design of perforated panel absorbers has adopted the profile of anechoic chamber walls where the perforated panel is inclined so that the air space behind the panel is trapezoidal.[18–20] Similar to previous analytical works, the response of the trapezoidal cavity is composed by rigid-walled modes of its bounding rectangular cavity, and this technique has been validated[15,17] and widely used.[13–17] The bounding cavity is defined in such a way that all its edges enclose exactly the trapezoidal shape. Rigid-walled modes of the rectangular cavity will be called RC modes and rigid-walled modes of the trapezoidal cavity will be called TC modes. A method is first established to identify the TC modes, where each TC mode can then be defined to evolve from each RC mode. The reason for this definition is explained and the evolution is illustrated graphically. Effects of the wall inclination on the shapes and resonance frequencies of the TC modes are studied when the inclination is altered under the condition of a constant volume of the trapezoidal cavity. As the trapezoidal cavity volume does not change, the number of TC modes is about the same for any inclination, and results for different inclinations can then be compared for the same modal density of the cavity. It is found that the inclination couples the RC modes together via two mechanisms, namely, the local coupling and the global coupling. Explanations are given to the operation of each mechanism and the types of RC modes that they can couple and cannot couple. Both couplings are then analytically related to the shapes and resonance frequencies of the TC modes. When the inclination is increased, it is shown that a given TC mode changes from the distorted shape of the RC mode that evolves it to a complicated and unrecognizable shape. A maximum or minimum or both can also exist in the resonance frequency of the TC mode. Computational examples are used to show how the local and global couplings determine the shapes of TC modes, and how these couplings control the occurrence of extrema in the resonance frequencies of various TC modes that evolve from different types of RC modes. The study provides a basic understanding of the free vibration behavior of the sound field in a trapezoidal space in terms of the way that the field reacts to the change in the inclination of its boundary and the physical mechanisms that are involved.

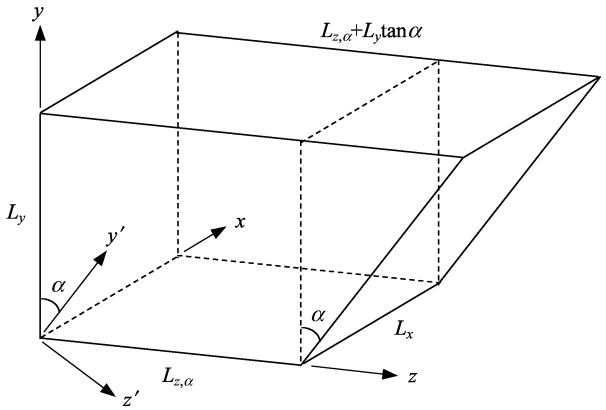


FIG. 1. Schematic illustration of the trapezoidal cavity with an inclined wall.

II. FREE VIBRATION OF THE TRAPEZOIDAL SOUND FIELD

A. Solutions to eigenvalues and modal pressure amplitudes

The trapezoidal cavity used is shown in Fig. 1, where the inclination of the tilted wall is represented by an angle, α. At the inclined wall, $z=L_{z,\alpha}+y\tan\alpha$. In a free vibration, the sound pressure at any position in the cavity, p, is described by the homogeneous wave equation:

$$\nabla^2 p + k^2 p = 0. \tag{1}$$

The wave number is $k=\omega/c_0$, where ω is the angular frequency and c_0 is the speed of sound in air. If Φ_{ai}^T is the shape function and k_{ai}^T is the wave number of the ith rigid-walled mode of the trapezoidal cavity (i.e., the ith TC mode) that satisfy Eq. (1), then

$$\nabla^2 \Phi_{ai}^T + (k_{ai}^T)^2 \Phi_{ai}^T = 0. \tag{2}$$

The resonance frequency of the mode is then $\omega_{ai}^T=k_{ai}^T c_0$ (rad/s), where $f_{ai}^T=\omega_{ai}^T/2\pi$ (Hz).

Consider a rectangular cavity of dimensions $L_x\times L_y\times(L_{z,\alpha}+L_y\tan\alpha)$ that encloses exactly the trapezoidal cavity in Fig. 1. Let P_i denotes the complex pressure amplitude of the ith rigid-walled mode of the rectangular cavity (i.e., the ith RC mode). If Φ_{ai} is the shape function of the mode that is given by

$$\Phi_{ai} = \cos(l_i\pi x/L_x)\cos(m_i\pi y/L_y) \times\cos[n_i\pi z/(L_{z,\alpha}+L_y\tan\alpha)], \tag{3}$$

then the corresponding wave number is

$$k_{ai}^2 = (l_i\pi/L_x)^2 + (m_i\pi/L_y)^2 + [n_i\pi/(L_{z,\alpha}+L_y\tan\alpha)]^2, \tag{4}$$

where l_i, m_i, and n_i are the indices of the RC mode, which represent the number of nodes in the x, y, and z directions, respectively. When Φ_{ai} is used as a basis function in the normal expansion of the sound pressure in the trapezoidal cavity and N RC modes are employed, the pressure can be written as

$$p = \sum_{i=1}^{N} P_i \Phi_{ai}. \tag{5}$$

Also, Φ_{ai} replaces Φ_{ai}^T, and k_{ai} replaces k_{ai}^T in Eq. (2).

By multiplying Eq. (2) with p after Φ_{ai}^T and k_{ai}^T are replaced, multiplying Eq. (1) with Φ_{ai}, subtracting the former from the latter, and integrating the result over the trapezoidal volume, V_T,

$$\int_{V_T} (\Phi_{ai} \nabla^2 p - p \nabla^2 \Phi_{ai}) dV = (k_{ai}^2 - k^2) \int_{V_T} p \Phi_{ai} dV. \tag{6}$$

The use of the second Green's formula converts the volume integral on the left-hand side of Eq. (6) into a surface integral,[21] and allows the boundary conditions of the trapezoidal cavity to be introduced into the solution of the pressure:

$$\int_A \left(\Phi_{ai} \frac{\partial p}{\partial n} - p \frac{\partial \Phi_{ai}}{\partial n} \right) dA = (k_{ai}^2 - k^2) \int_{V_T} p \Phi_{ai} dV, \tag{7}$$

where A denotes the area of the surface under consideration, and n denotes the unit vector for the direction normal to the surface (positive outward). The boundary conditions are described by the two differentials in the equation.

At the inclined wall, $\partial \Phi_{ai}/\partial n$ does not necessarily equal to zero but at all other noninclined walls, $\partial \Phi_{ai}/\partial n = 0$. When all the walls are rigid, $\partial p/\partial n = 0$. By substituting Eq. (5) into Eq. (7) and then using these boundary conditions, it can be shown that

$$(\omega_{ai}^2 - \omega^2) \sum_{k=1}^{N} M_{i,k} P_k = -\rho_0 c_0^2 A_w \sum_{k=1}^{N} I_{i,k} P_k, \tag{8}$$

$$M_{i,k} = \rho_0 \int_{V_T} \Phi_{ai} \Phi_{ak} dV, \tag{9}$$

$$I_{i,k} = \frac{1}{A_w} \int_{A_w} \Phi_{ak} \frac{\partial \Phi_{ai}}{\partial n} dA, \tag{10}$$

where ρ_0 is the air density, A_w is the surface area of the inclined wall, and the resonance frequency of the ith RC mode is $\omega_{ai} = k_{ai} c_0$. $I_{i,k}$ can be analytically obtained after a coordinate transformation from y and z to y' and z' (see Fig. 1). As described below, Eq. (8) can be converted into an eigenvalue equation, where ω is an unknown to be solved that gives an eigenvalue of the trapezoidal cavity, and P_k's are also unknowns that correspond to the free vibration for this eigenvalue. [Note: In a forced vibration by a harmonic sound source with a volume velocity per unit volume of the cavity, q, the right-hand side of Eq. (1) would have the $-j\rho_0 \omega q$ term where the wave equation becomes inhomogeneous. In this case, it can be shown from the similar derivation as above that Eqs. (6)–(8) would have an extra term, which is a function of ωq. ω in Eq. (8) is then the forcing frequency rather than an eigenvalue solution, and P_k's are unknowns that correspond to the forced vibration at the forcing frequency and affected by q].

A substitution of $\lambda = -j\omega$ is made in Eq. (8), where λ denotes the eigenvalue of the trapezoidal cavity. For N eigenvalues associated with N TC modes, the equation can be expressed as a matrix equation:

$$(\lambda^2 \mathbf{M} + \mathbf{S}) \mathbf{P} = 0. \tag{11}$$

$\mathbf{M}$ and $\mathbf{S}$ are the mass and stiffness matrices:

$$\mathbf{M} = \begin{bmatrix} M_{1,1} & \cdots & M_{1,N} \\ \vdots & \ddots & \vdots \\ M_{N,1} & \cdots & M_{N,N} \end{bmatrix}, \tag{12}$$

$$\mathbf{S} = \begin{bmatrix} \omega_{a1}^2 M_{1,1} + \rho_0 c_0^2 A_w I_{1,1} & \cdots & \omega_{a1}^2 M_{1,N} + \rho_0 c_0^2 A_w I_{1,N} \\ \vdots & \ddots & \vdots \\ \omega_{aN}^2 M_{N,1} + \rho_0 c_0^2 A_w I_{N,1} & \cdots & \omega_{aN}^2 M_{N,N} + \rho_0 c_0^2 A_w I_{N,N} \end{bmatrix}. \tag{13}$$

Equation (11) is a standard eigenvalue equation, which can be directly solved for the eigenvalues of the TC modes. For each eigenvalue, there is a corresponding set of modal pressure amplitudes given in $\mathbf{P}$, where $\mathbf{P_j} = [P_1^{(j)} P_2^{(j)} \ldots P_N^{(j)}]^T$ is for the jth eigenvalue, λ_j. $[\ ]^T$ denotes the matrix transpose. Since Eq. (11) uses a finite number of RC modes, N, a truncation error exists in the calculation, and the accuracy of λ_j's and $\mathbf{P_j}$'s of the TC modes depends on the value of N. From previous works on irregular cavities, it is known that truncation errors increase with the irregularity and more rigid-walled modes of the bounding rectangular cavities are required for the errors to become negligibly small.[14,16,17] This means that as α increases, the trapezoidal cavity deviates more from a rectangular shape, and the size of its bounding rectangular cavity increases. Thus, N used in the calculation must also increase if λ_j's and $\mathbf{P_j}$'s of the TC modes are to be accurately predicted. However, as N does not only increase with α and the rectangular cavity dimensions but generally also, with the highest resonance frequency of the TC modes of interest, its chosen value is specific only to the particular problem under consideration. For example, the work in Ref. 17 only employed up to 20 RC modes, but the study in Ref. 16 used a large number of RC modes in the order of 40,000.

B. Local and global couplings

Equation (8) can be rewritten as a quadratic equation of ω:

$$\omega^2 - \omega_{ai}^2 - \rho_0 c_0^2 A_w \sum_{k=1}^{N} I_{i,k} P_k \Big/ \sum_{k=1}^{N} M_{i,k} P_k = 0. \tag{14}$$

As Eq. (11) originates from Eq. (14), the trapezoidal cavity eigenvalues that have been obtained from Eq. (11) are actually the solutions of Eq. (14) when the values of P_k's for each eigenvalue are given. It can be seen from Eq. (14) that the wall inclination affects the solutions of the equation (i.e., the eigenvalues) via the two summations in the last term, which correspond to two different physical mechanisms. The numerator part is called the local coupling and the denominator part is called the global coupling, and these mecha-

nisms couple each RC mode to other RC modes. The local coupling of the ith RC mode for the jth eigenvalue is given by

$$L_{ai}^{(j)} = \rho_0 c_0^2 A_w \sum_{k=1}^{N} I_{i,k} P_k^{(j)}, \tag{15}$$

where $L_{ai}^{(j)}$ describes the coupling between the ith RC mode and itself and other RC modes at the inclined wall (i.e., surface coupling). On the other hand, the global coupling of the ith RC mode for the jth eigenvalue is given by

$$G_{ai}^{(j)} = \sum_{k=1}^{N} M_{i,k} P_k^{(j)}, \tag{16}$$

where $G_{ai}^{(j)}$ describes the coupling between the ith RC mode and itself and other RC modes throughout the trapezoidal cavity (i.e., volume coupling). The intercoupling of RC modes in this mechanism arises from the nonorthogonality of the RC-mode shapes in the trapezoidal volume, which can be verified by substituting Eq. (3) into Eq. (9) and solving the integral.

As far as the local and global couplings are concerned, the RC modes can be categorized into three types. In the first type, the RC modes graze the inclined wall in the x direction (i.e., perpendicular to the inclination), and have indices $(*,0,0)$ where $*$ is a non-negative integer. In the second type, the RC modes still graze the inclined wall but not perpendicular to the inclination, and have indices $(*,\#,0)$ where # is a nonzero positive integer. In the third type, the RC modes do not graze the inclined wall and have indices $(*,*,\#)$. The three types of modes will be called perpendicular grazing RC modes, nonperpendicular grazing RC modes, and nongrazing RC modes. As described in the following, the local and global couplings are selective that only certain types of RC modes are coupled together.

The wall inclination is perpendicular to the x direction and it does not affect the RC-mode shape in this direction. Thus, as far as the local coupling is concerned, there are two consequences. First, $\partial\Phi_{ai}/\partial n=0$ for a perpendicular grazing ith RC mode. So, Eq. (10) suggests that $I_{i,k}=0$ regardless of the type of the kth RC mode because Φ_{ak} is multiplied to $\partial\Phi_{ai}/\partial n$. Thus, $L_{ai}^{(j)}=0$ from Eq. (15). In other words, a perpendicular grazing ith RC mode has a zero local coupling since it does not excite itself or other RC modes. Second, for a nonperpendicular grazing or nongrazing ith RC mode, $\partial\Phi_{ai}/\partial n\neq 0$, and $\cos(l_i\pi x/L_x)$ is orthogonal to $\cos(l_k\pi x/L_x)$ over the cavity width, L_x, because it is unaffected after solving $\partial\Phi_{ai}/\partial n$. Therefore, from Eq. (10), $I_{i,k}\neq 0$ when $l_k=l_i$ (for $k=i$ or $k\neq i$), and $I_{i,k}=0$ when $l_k\neq l_i$ (for $k\neq i$), regardless of m_i, m_k, n_i, and n_k. In other words, a nonperpendicular grazing or nongrazing ith RC mode only excites itself or other RC modes with the same number of nodes in the x direction (i.e., perpendicular to the inclination). As a result, it can be shown from matrix **P** after solving Eq. (11) that if the jth eigenvalue corresponds to the RC modes with indices $(l,*,*)$, then $P_k^{(j)}\neq 0$ for $l_k=l$, and $P_k^{(j)}=0$ for $l_k\neq l$. Hence, from Eq. (15), $L_{ai}^{(j)}\neq 0$ for $l_i=l$, and $L_{ai}^{(j)}=0$ for $l_i\neq l$.

As far as the global coupling is concerned, the $\cos(l_i\pi x/L_x)$ term is also unaffected by the inclination and it is then orthogonal to $\cos(l_k\pi x/L_x)$ over L_x. So, from Eq. (9), $M_{i,k}\neq 0$ when $l_k=l_i$ (for $k=i$ or $k\neq i$), and $M_{i,k}=0$ when $l_k\neq l_i$ (for $k\neq i$), regardless of m_i, m_k, n_i, and n_k. In other words, an ith RC mode only couples with itself or other RC modes with the same number of nodes in the x direction. Thus, similar to the case of the local coupling, it can be shown from Eq. (16) that if the jth eigenvalue corresponds to the $(l,*,*)$ modes, then $G_{ai}^{(j)}\neq 0$ for $l_i=l$, and $G_{ai}^{(j)}=0$ for $l_i\neq l$, since $P_k^{(j)}\neq 0$ for $l_k=l$, and $P_k^{(j)}=0$ for $l_k\neq l$. In addition, the positions of Φ_{ai} and Φ_{ak} are exchangeable in the integral in Eq. (9) and thus, matrix **M** is symmetric along its diagonal.

C. Definition, resonance frequencies, and shapes of TC modes

Unlike the rectangular cavity, the trapezoidal cavity is irregular that TC modes are not recognizable because their shapes and resonance frequencies are not definable analytically. As a result, the TC modes are not distinguishable from one another. So, it is essential to first establish a way to identify individual TC modes before their shapes and resonance frequencies are studied. As N eigenvalues for N TC modes are generated by N RC modes from Eq. (11), it is convenient to assign individual TC modes to individual RC modes by defining the evolution of each TC mode from each RC mode. The definition is established by linking the resonance frequencies of the TC modes to the indices of the RC modes because these frequencies are the only identity of the TC modes after Eq. (11) is solved. This is explained as follows. Let $\omega_a^{T(j)}$ be the resonance frequency for the jth eigenvalue obtained by solving Eq. (11). Therefore, $\omega=\omega_a^{T(j)}$ is a solution of Eq. (14) which can then be reexpressed as

$$(\omega_a^{T(j)})^2 = \omega_{ai}^2 + R_{ai}^{(j)}; \quad R_{ai}^{(j)} = L_{ai}^{(j)}/G_{ai}^{(j)}, \tag{17}$$

where $R_{ai}^{(j)}$ is the local-to-global coupling ratio.

However, since $L_{ai}^{(j)}=0$ and $G_{ai}^{(j)}=0$ when the jth eigenvalue corresponds to the $(l,*,*)$ modes and the indices of the ith RC mode are such that $l_i\neq l$ as described in Sec. II B, Eq. (17) is only valid for the $(l_i,*,*)$ modes with $l_i=l$. In the latter, the ω_{ai} and $R_{ai}^{(j)}$ of any of such RC modes give $\omega_a^{T(j)}$ from Eq. (17). In other words, the unknown l associated with a given $\omega_a^{T(j)}$ obtained from Eq. (11) can be deduced as $l=l_i$ when the $(l_i,*,*)$ modes yield the same $\omega_a^{T(j)}$ from Eq. (17). For the other $\omega_a^{T(j)}$'s obtained from Eq. (11), the same way is used to determine the corresponding l's. Subsequently, all the $\omega_a^{T(j)}$'s for the α under consideration are sorted in ascending order, and all the RC modes are sorted in the ascending order of their ω_{ai}'s for $\alpha=0°$. The sorting of the RC modes is based on $\alpha=0°$ because only the rectangular cavity volume for $\alpha=0°$ is the same as the trapezoidal cavity volume for $\alpha>0°$. Then, each $\omega_a^{T(j)}$ is successively allocated to each $(l_i,*,*)$ mode with the aid of matching the l's to l_i's, and thus, individual TC modes are associated with individual RC modes. Since the former modes are now uniquely identified using the latter modes, $\omega_a^{T(j)}$ can be re-

placed by $\omega_{a(i)}^T$ where subscript (i) indicates that the TC mode evolves from the ith RC mode (i.e., $f_{a(i)}^T = \omega_{a(i)}^T/2\pi$). Hence, Eq. (17) can be rewritten as

$$(\omega_{a(i)}^T)^2 = \omega_{ai}^2 + R_{ai}^{(j)}. \tag{18}$$

[Note: The definitions of ω_{ai}^T and f_{ai}^T given below Eq. (2) in Sec. II A refer to the ith TC mode, and they are different from those of $\omega_{a(i)}^T$ and $f_{a(i)}^T$ here. The ith TC mode for a given α does not necessarily evolve from the ith RC mode]. Equation (18) shows that the wall inclination modifies the resonance frequency of the ith RC mode to the resonance frequency of its evolved TC mode via the local-to-global coupling ratio. As described in Sec. II B, $L_{ai}^{(j)}=0$ for a perpendicular grazing RC mode. Thus, $R_{ai}^{(j)}=0$ from Eq. (17) and $\omega_{a(i)}^T=\omega_{ai}$ from Eq. (18). Since ω_{ai} of this type of RC mode is not affected by α because its $n_i=0$ [it can be verified by Eq. (4)], $\omega_{a(i)}^T$ does not change with the wall inclination.

By using Eqs. (15) and (16), Eq. (18) can be rearranged to obtain $P_i^{(j)}$ at $\omega_{a(i)}^T$:

$$P_i^{(j)} = \sum_{k\neq i}^{N} \{\rho_0 c_0^2 A_w I_{i,k} - [(\omega_{a(i)}^T)^2 - \omega_{ai}^2] M_{i,k}\} \times P_k^{(j)} \Big/ \{[(\omega_{a(i)}^T)^2 - \omega_{ai}^2] M_{i,i} - \rho_0 c_0^2 A_w I_{i,i}\}. \tag{19}$$

Equation (19) indicates that at $\omega_{a(i)}^T$, $P_i^{(j)}$ is dependent on the pressure amplitudes of other RC modes through the local and global couplings. This means that the couplings distribute the acoustic energy in the ith RC mode among other RC modes. As both couplings vary with α, $P_i^{(j)}$ also changes. Again, for a perpendicular grazing ith RC mode, $I_{i,k}=0$ and $L_{ai}^{(j)}=0$, and Eq. (19) can then be simplified as

$$P_i^{(j)} = -\sum_{k\neq i}^{N} M_{i,k} P_k^{(j)} / M_{i,i}. \tag{20}$$

For this type of the ith RC mode, Eq. (20) suggests that only the global coupling is at work since the local coupling is zero. Thus, although $\omega_{a(i)}^T$ does not change with the wall inclination as explained in the above, $P_i^{(j)}$ will change because the global coupling still distributes the acoustic energy in the ith RC mode among other RC modes. However, for a nonperpendicular grazing or nongrazing ith RC mode, both the local and global couplings are at work. Therefore, Eqs. (18) and (19) suggest that both $\omega_{a(i)}^T$ and $P_i^{(j)}$ will change with the inclination.

The trapezoidal cavity pressure for the jth eigenvalue, $p^{(j)}$, evaluated at $\omega_{a(i)}^T$ is only due to the TC mode that evolves from the ith RC mode. By employing Eq. (5), this pressure can be expressed as

$$p^{(j)} = p_{(i)}^T = P_i^{(j)} \Phi_{ai} + \sum_{k\neq i}^{N} P_k^{(j)} \Phi_{ak}, \tag{21}$$

where $p_{(i)}^T$ is the spatial distribution of pressure of the TC mode. Equation (21) suggests that $p_{(i)}^T$ is a function of the local and global couplings as the pressure amplitude of each RC mode depends on both couplings [see Eq. (19)]. It is obvious from Eq. (21) that $p_{(i)}^T$ has the shape of Φ_{ai} combined with the shapes of Φ_{ak}'s. The extent of the shape distortion of Φ_{ai} by Φ_{ak}'s depends on the magnitude of $P_i^{(j)}$ relative to the magnitudes of $P_k^{(j)}$'s and thus, the amount of the couplings. The larger the couplings, the more important the influence of Φ_{ak}'s when the shape of $p_{(i)}^T$ can become complicated.

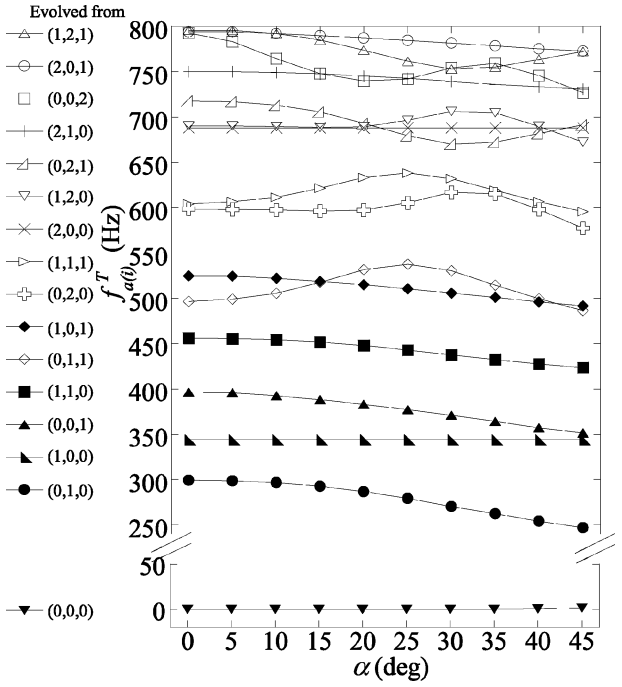


FIG. 2. The resonance frequencies of the first 16 TC modes and the RC modes that evolve the TC modes.

III. RESULTS AND DISCUSSION

Examples are used to illustrate the effects of the wall inclination of the trapezoidal cavity on the shapes and resonance frequencies of some selected TC modes. The width and height of the cavity are $L_x=0.5$ m and $L_y=0.575$ m. Values of α that range from 0° to 45° are considered in 5° increments. As the cavity volume is maintained the same when the inclination is varied, the length of the cavity at $y=L_y$ (i.e., $L_{z,\alpha}+L_y \tan\alpha$) increases with α, and the length of the cavity at $y=0$ m (i.e., $L_{z,\alpha}$) decreases with α (see Fig. 1). For $\alpha=(0°,5°,10°,15°,20°,25°,30°,35°,40°,45°)$, the corresponding values of both lengths are $L_{z,\alpha}+L_y \tan\alpha = (0.434,\ 0.459,\ 0.485,\ 0.511,\ 0.539,\ 0.568,\ 0.6,\ 0.635,\ 0.675,\ 0.722)$ m and $L_{z,\alpha}=(0.434,\ 0.409,\ 0.383,\ 0.357,\ 0.329,\ 0.3,\ 0.268,\ 0.233,\ 0.193,\ 0.147)$ m. For these values of α and trapezoidal cavity dimensions, the number of RC modes that has been employed, N, increases with α. Since only the first 22 TC modes are studied in the following and their highest resonance frequency is less than 930 Hz, values of N of several hundreds have been used where the truncation errors in these TC modes can be reasonably ignored.

Figure 2 depicts the $f_{a(i)}^T$'s of the first 16 TC modes as a

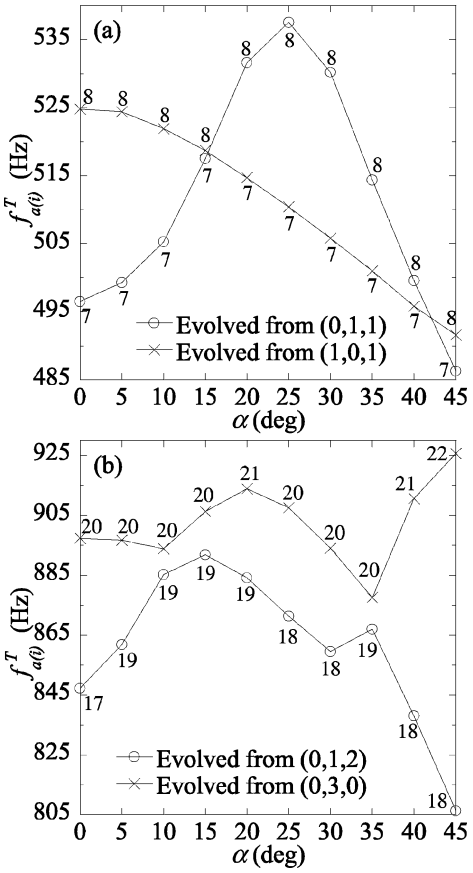


FIG. 3. The resonance frequencies of the TC modes that evolve from (a) (0,1,1) and (1,0,1), and (b) (0,1,2) and (0,3,0).

function of α. The RC modes from which the TC modes evolve, are shown. Three specific features can be obtained from the figure. First, the $f^T_{a(i)}$'s of the TC modes that evolve from (0,0,0), (1,0,0), and (2,0,0) do not change with α. Second, the $f^T_{a(i)}$'s of the other TC modes exhibit variations with α, where a maximum or minimum or both exist. These extrema can occur at either small values or large values of α. For example, the minima for the TC modes that evolve from (0,1,0), (0,0,1), (1,1,0), (1,0,1), (2,1,0), and (2,0,1) occur at $\alpha > 45°$ and are not shown in Fig. 2. Third, the TC modes are sparsely distributed in the frequency range below 500 Hz and each of the modes has the same mode number for the values of α considered. For example, the TC mode that evolves from (0,1,0) is always the second mode and the TC mode that evolves from (1,1,0) is always the fifth mode. However, the distribution of the TC modes is denser above 500 Hz and on the frequency rank, adjacent TC modes can exchange positions or mode numbers when α is increased. This third feature and the way that RC modes compose TC modes are illustrated below. The first and second features are explained in the later part of the discussion.

Figure 3 presents the $f^T_{a(i)}$'s of two pairs of adjacent TC modes as a function of α. For each value of α, the TC-mode numbers are shown at the markers of the curves. In Fig. 3(a), only the seventh and eighth TC modes exist for the values of α and the frequency range shown, and the TC modes that evolve from (0,1,1) and (1,0,1) are considered. Thus, only the TC-mode numbers of 7 and 8 are of concern in the ex-

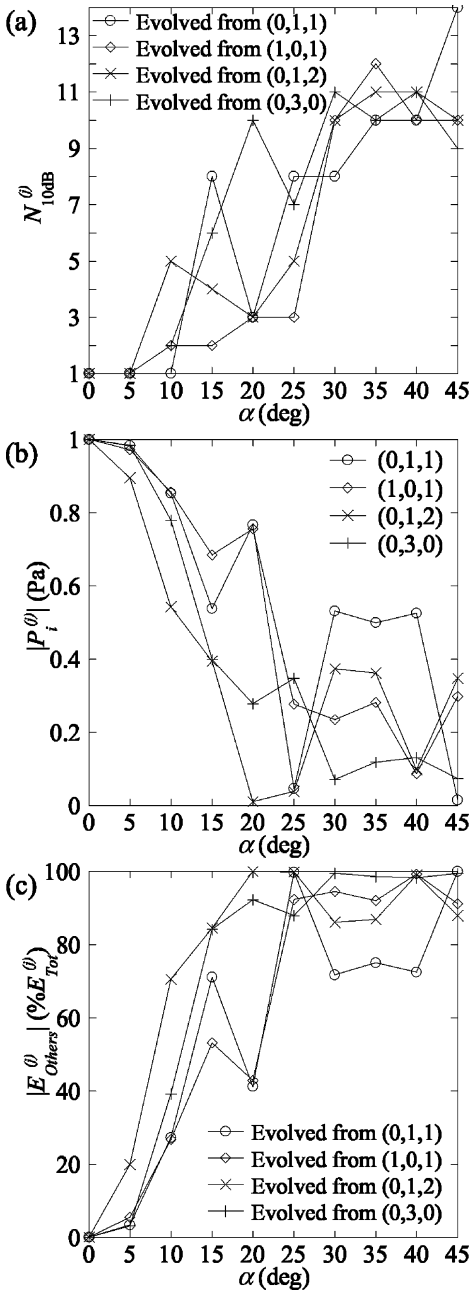


FIG. 4. (a) The number of RC modes with the highest 10 dB of pressure amplitudes for the TC modes that evolve from (0,1,1), (1,0,1), (0,1,2), and (0,3,0), (b) the pressure amplitudes of (0,1,1), (1,0,1), (0,1,2), and (0,3,0), and (c) the total energy of the other RC modes.

change of the TC-mode positions on the frequency rank when α is increased. For $\alpha = 0° - 15°$, the $f^T_{a(i)}$ of the seventh TC mode has an associated $l=0$ and it evolves from (0,1,1), and the $f^T_{a(i)}$ of the eighth TC mode has an associated $l=1$ and it evolves from (1,0,1). The $f^T_{a(i)}$'s approach each other at $\alpha = 15°$ after which the $f^T_{a(i)}$ of the seventh TC mode has an associated $l=1$ and the $f^T_{a(i)}$ of the eighth TC mode has an associated $l=0$, where both modes exchange positions. In other words, the seventh TC mode becomes the mode that evolves from (1,0,1) and the eighth TC mode becomes the mode that evolves from (0,1,1). At $\alpha = 45°$, both TC modes change back to their original positions since their associated l's are, respectively, 0 and 1. In Fig. 3(b), the 17th to 22nd

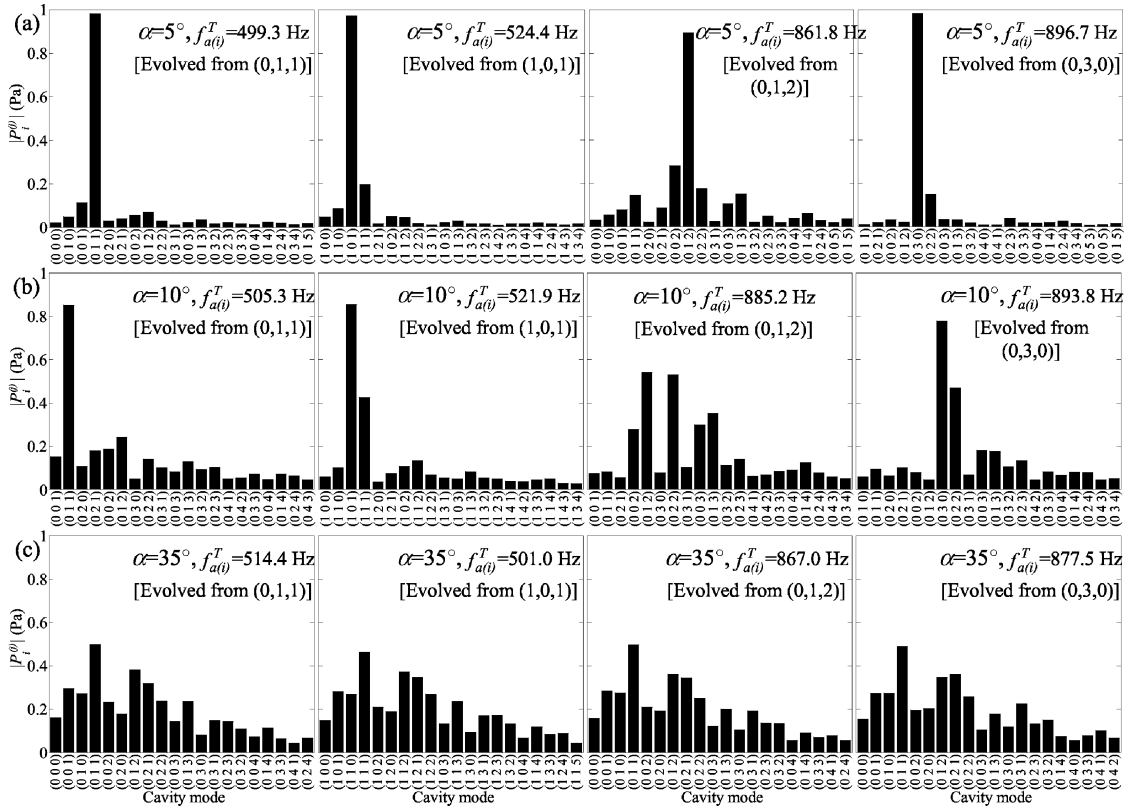


FIG. 5. The distributions of pressure amplitude of the top 20 RC modes with the highest amplitudes for the TC modes that evolve from (0,1,1), (1,0,1), (0,1,2), and (0,3,0) at (a) $\alpha=5°$, (b) $\alpha=10°$, and (c) $\alpha=35°$.

TC modes exist where two of their $f^T_{a(i)}$'s have the associated $l=0$. The TC modes that evolve from (0,1,2) and (0,3,0) are considered, and their positions can change between 17 and 22 when α is increased since there are six TC modes in the frequency range shown. At $\alpha=0°$, (0,1,2) ranks lower than (0,3,0) (i.e., 17th and 20th, respectively) because the former has a lower f_{ai} than the latter. Based on this order and that the $f^T_{a(i)}$'s of both the TC modes that evolve from these RC modes have the same associated $l=0$, the two TC modes do not exchange positions after their $f^T_{a(i)}$'s approach each other at $\alpha=10°$ and 35° [see Fig. 3(b)].

Figure 4 depicts the RC-mode contribution at the $f^T_{a(i)}$'s of the TC modes that evolve from (0,1,1), (1,0,1), (0,1,2), and (0,3,0). At each $f^T_{a(i)}$, $|P^{(j)}_i|$'s of all RC modes can be obtained for a given value of α after Eq. (11) is solved. By converting the $|P^{(j)}_i|$'s into dB, the number of RC modes with the highest 10 dB of pressure amplitudes, $N_{10\ \mathrm{dB}}$, is counted and presented in Fig. 4(a). In general, it can be seen that for $\alpha<10°$, each TC mode is composed by only one RC mode but for $\alpha\geqslant 10°$, many RC modes are involved. From the trend of Fig. 4(a), the number of RC modes that are significant to the composition of the TC modes increases with α. Figure 4(b) presents the $|P^{(j)}_i|$'s of the four RC modes of interest at the $f^T_{a(i)}$'s of the evolved TC modes, and Fig. 4(c) presents the total acoustic energy in the other RC modes, $E^{(j)}_{\mathrm{Others}}$, expressed as the percentage of the total energy in all RC modes, $E^{(j)}_{\mathrm{Tot}}$. $E^{(j)}_{\mathrm{Others}}$ and $E^{(j)}_{\mathrm{Tot}}$ are calculated from the acoustic energies in the RC modes, where the energy in the ith RC mode is given by $|P^{(j)}_i|^2 V_T/\rho_0 c_0^2$. From Fig. 4(b), the $|P^{(j)}_i|$'s of (0,1,1), (1,0,1), (0,1,2), and (0,3,0) have a maximum value at $\alpha=0°$ when they are the only RC modes that individually compose the TC modes at the respective $f^T_{a(i)}$'s. As the other RC modes are not excited, $E^{(j)}_{\mathrm{Others}}=0\%$ of $E^{(j)}_{\mathrm{Tot}}$ from Fig. 4(c). For $\alpha>0°$, the local and global couplings exist, and the energies in the four RC modes are distributed among the other RC modes that are excited into motion by the couplings. Consistent with the increase of $N_{10\ \mathrm{dB}}$ with α in Fig. 4(a), Figs. 4(b) and 4(c) indicate that when α increases, $|P^{(j)}_i|$ has a decreasing trend and $E^{(j)}_{\mathrm{Others}}$ has an increasing trend (i.e., $|P^{(j)}_i|$ and $E^{(j)}_{\mathrm{Others}}$ are related in a reciprocal manner). This suggests that, in general, the influence of the local and global couplings on the energy distribution increases with α, where the energy is distributed among more RC modes. The details of the energy distribution can be obtained from the spread of $|P^{(j)}_i|$'s among the RC modes (see Fig. 5 which illustrates the top 20 RC modes with the highest $|P^{(j)}_i|$'s at $\alpha=5°$, 10°, and 35°). At $\alpha=5°$, the dominance of single RC modes is obvious [i.e., (0,1,1), (1,0,1), (0,1,2), and (0,3,0) in Fig. 5(a)]. As a result, the TC modes still possess the shapes of the individually dominant RC modes [e.g., at $\alpha=5°$, Fig. 6(a) shows that the TC modes that evolve from (0,1,2) and (0,3,0) have the shapes of these RC modes]. At $\alpha=10°$, Fig. 5(b) shows that (0,1,1), (1,0,1), (0,1,2), and (0,3,0) still have a great influence but there are more than one RC mode that contribute significantly to the composition of the TC modes. This results in the shapes of these four RC modes to distort in the evolved TC modes [e.g., see Fig. 6(b) for the TC modes that have the distorted shapes of (0,1,2)

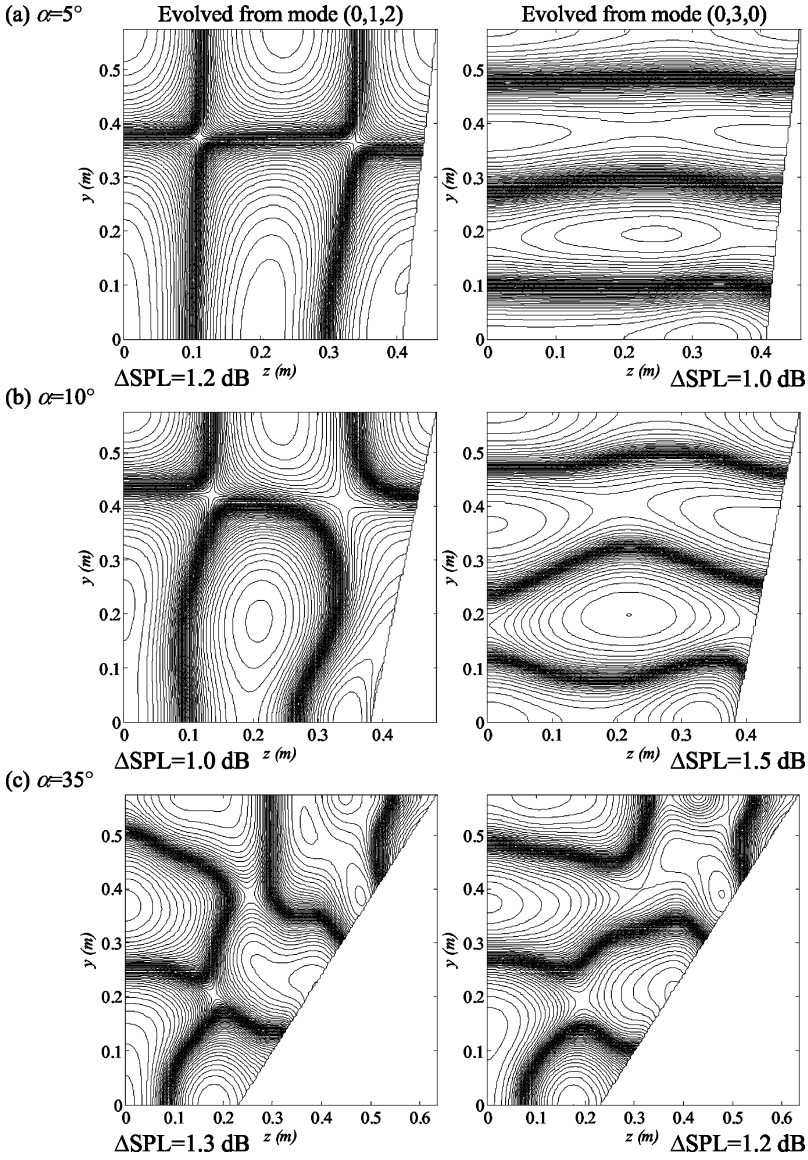


FIG. 6. The spatial distributions of pressure of the TC modes that evolve from (0,1,2) and (0,3,0) across the $x=0.0$ m plane at (a) $\alpha=5°$, (b) $\alpha=10°$, and (c) $\alpha=35°$. The dark regions indicate the nodal lines. ΔSPL is the sound pressure level difference between contours.

and (0,3,0) at $\alpha=10°$]. For $\alpha>10°$, many RC modes are involved and there is almost an equal energy distribution among the modes [e.g., $\alpha=35°$ in Fig. 5(c)]. Consequently, the evolved TC modes have complicated shapes that are not analytically definable because they are composed by many significant RC modes [e.g., see Fig. 6(c) for the shapes of the TC-modes that evolve from (0,1,2) and (0,3,0) at $\alpha=35°$]. This justifies the need to establish a way to define the TC modes as described in Sec. II C. For $\alpha>0°$, the selective nature of the local and global couplings can also be seen in Fig. 5, where only those RC modes with the same l_i's as the individual RC modes that evolve the TC modes are excited (i.e., all the RC modes that compose a given TC mode have the same l_i's).

Figure 7 shows the $f^T_{a(i)}$'s of the TC modes that evolve from (0,2,0), (1,0,0), (0,2,1), and (1,1,1). The first has a minor minimum at $\alpha=15°$ followed by an obvious maximum at $\alpha=30°$, the second does not change with α, the third has a minimum at $\alpha=30°$, and the fourth has a maximum at $\alpha=25°$. These features can be related to the local-to-global coupling ratio, $R^{(j)}_{ai}$, of each of the four RC modes as suggested by Eq. (18). The local and global couplings (i.e., $L^{(j)}_{ai}$'s and $G^{(j)}_{ai}$'s) of these RC modes are presented in Fig. 8, and the corresponding $R^{(j)}_{ai}$'s are presented in Fig. 9. When $L^{(j)}_{ai}$ and $G^{(j)}_{ai}$ are either simultaneously positive or simultaneously negative, the couplings are in-phase and $R^{(j)}_{ai}$ is positive. Otherwise, they are out-of-phase and $R^{(j)}_{ai}$ is negative (see Figs. 8 and 9). Based on Eq. (18), a positive $R^{(j)}_{ai}$ has an effect of increasing $\omega^T_{a(i)}$ and a negative $R^{(j)}_{ai}$ has an opposite effect, but these depend also on ω_{ai}. Therefore, the way in which the behavior of $\omega^T_{a(i)}$ is affected by $R^{(j)}_{ai}$ depends on whether a given TC mode evolves from a grazing or nongrazing RC mode.

For the case of grazing RC modes, Figs. 7(a) and 7(b) show that the f_{ai}'s of (0,2,0) and (1,0,0) do not vary with α. This can be verified by Eq. (4) that k_{ai} is independent of α when $n_i=0$. So, Eq. (18) suggests that the $R^{(j)}_{ai}$ behavior of the RC mode directly controls the $\omega^T_{a(i)}$ behavior of the evolved TC mode. This means that when the $R^{(j)}_{ai}$ has a maximum, the $\omega^T_{a(i)}$ will have a maximum, and when the $R^{(j)}_{ai}$ has a minimum, the $\omega^T_{a(i)}$ will have a minimum [e.g., Figs. 7(a), 7(b),

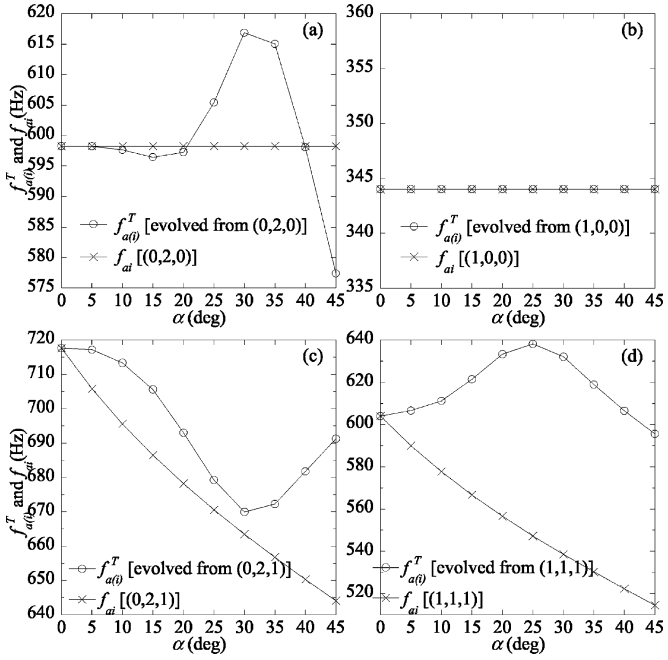


FIG. 7. The resonance frequencies of the TC modes that evolve from (a) (0,2,0), (b) (1,0,0), (c) (0,2,1), and (d) (1,1,1), and the resonance frequencies of the RC modes.

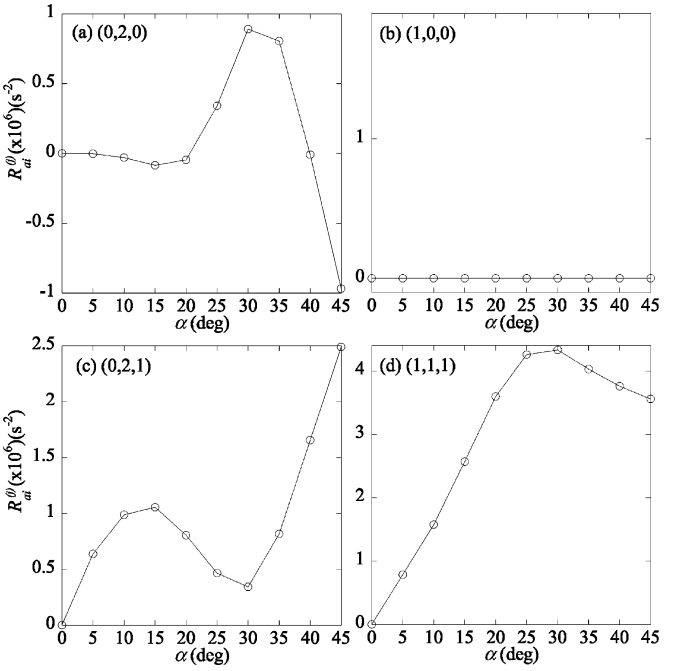


FIG. 9. The local-to-global coupling ratios of (a) (0,2,0), (b) (1,0,0), (c) (0,2,1), and (d) (1,1,1).

8(a), and 8(b) show that for the respective TC modes that evolve from (0,2,0) and (1,0,0), the characteristics of the $f^T_{a(i)}$'s and the $R^{(j)}_{ai}$'s are similar]. The $f^T_{a(i)}$ of the TC mode that evolves from (1,0,0) does not have any extrema and is exactly the same as the f_{ai} of the RC mode which does not vary with α. The reason is because (1,0,0) is a perpendicular grazing RC mode where its $L^{(j)}_{ai}=0$ [see Fig. 8(b)] and thus, its $R^{(j)}_{ai}=0$ [see Fig. 9(b)]. The $L^{(j)}_{ai}=0$ is due to $\partial\Phi_{ai}/\partial n=0$ at the inclined wall as has been explained in Sec. II B. The results for the TC modes that evolve from (0,2,0) and (1,0,0) suggest that the existence and the locations of extrema in their $\omega^T_{a(i)}$'s are controlled only by the $R^{(j)}_{ai}$'s of the individual RC modes.

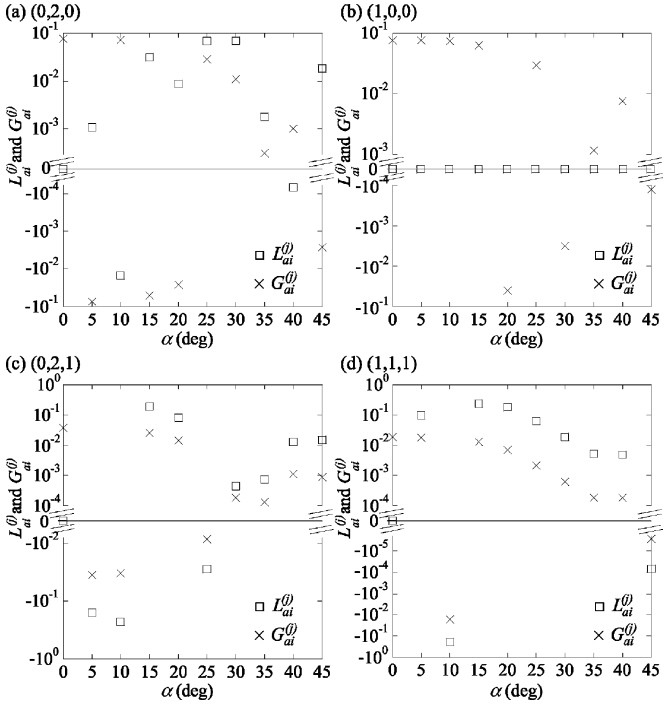


FIG. 8. The local and global couplings of (a) (0,2,0), (b) (1,0,0), (c) (0,2,1), and (d) (1,1,1). The local coupling has been divided by $\rho_0 c_0^2$.

For the case of nongrazing RC modes, comparisons between Figs. 7(c) and 9(c) that concern (0,2,1), and between Figs. 7(d) and 9(d) that concern (1,1,1), reveal that the $f^T_{a(i)}$ behavior is not necessarily similar to the $R^{(j)}_{ai}$ behavior (i.e., the occurrence of extrema in the $f^T_{a(i)}$'s is not directly controlled by the $R^{(j)}_{ai}$'s). For example, the $R^{(j)}_{ai}$ in Fig. 9(c) has a minimum and a maximum but the $f^T_{a(i)}$ in Fig. 7(c) only has a minimum and no maximum. In addition, the $R^{(j)}_{ai}$ in Fig. 9(d) has a maximum at $\alpha=30°$ but the maximum in the $f^T_{a(i)}$ in Fig. 7(d) is at $\alpha=25°$. As the f_{ai}'s of nongrazing RC modes decrease with α [e.g., see Figs. 7(c) and 7(d)], Eq. (18) suggests that the characteristics of both the ω_{ai} and $R^{(j)}_{ai}$ have to be considered. The decrease of the ω_{ai} with α can be verified by Eq. (4), where k_{ai} is a reciprocal function of α when $n_i>0$. Since an extremum in the $\omega^T_{a(i)}$ occurs at $\partial(\omega^T_{a(i)})/\partial\alpha=0$, Eq. (18) implies that $\partial\omega^2_{ai}/\partial\alpha+\partial R^{(j)}_{ai}/\partial\alpha=0$ and will change sign when one passes the extremum. In other words, the total rate of change of the ω^2_{ai} and $R^{(j)}_{ai}$ with α determines the behavior of the $\omega^T_{a(i)}$. The discrete total rate, $\Delta\omega^2_{ai}/\Delta\alpha+\Delta R^{(j)}_{ai}/\Delta\alpha$, is depicted in Fig. 10 for (0,2,1) and (1,1,1). Since the gradient of the ω^2_{ai} and $R^{(j)}_{ai}$ is computed from the ω^2_{ai}'s and $R^{(j)}_{ai}$'s at each pair of adjacent values of α, it is plotted at the midpoints of adjacent angles. For (0,2,1), Fig. 10(a) shows that the total rate changes sign from negative to positive between $\alpha=25°$ and 35°, which corresponds to a minimum in $f^T_{a(i)}$ at $\alpha=30°$ [see Fig. 7(c)]. For (1,1,1), Fig. 10(b) shows that the total rate changes sign from positive to negative between $\alpha=20°$ and 30°, which corresponds to a

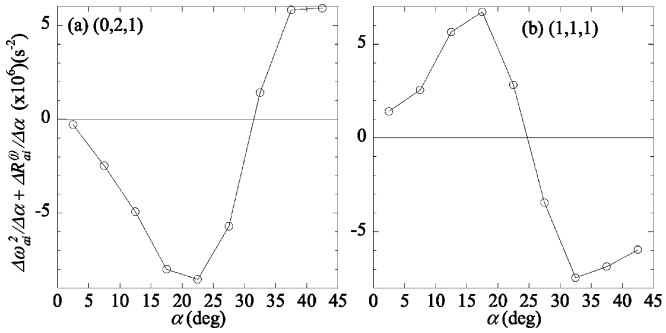


FIG. 10. The total rates of change with α of the resonance frequency square and the local-to-global coupling ratio for (a) (0,2,1) and (b) (1,1,1).

maximum in $f_{a(i)}^T$ at $\alpha=25°$ [see Fig. 7(d)]. In both cases, an extremum occurs when the ω_{ai}^2 and $R_{ai}^{(j)}$ have an equal and opposite rate of change with α.

IV. CONCLUSIONS

The coupling between rigid-walled modes of a rectangular cavity (RC modes) is used to study the shapes and resonance frequencies of rigid-walled modes of a trapezoidal cavity (TC modes) with an inclined rigid wall. This wall is characterized by an inclination angle, α, that is varied by retaining a constant trapezoidal cavity volume. As the TC modes are not definable analytically, a method to identify these modes is first established. It is based on matching the indices of the sorted RC modes for $\alpha=0°$ in the direction perpendicular to the inclination (i.e., the x direction in this paper), l_i's, to the indices in the same direction, l's, which are associated with the sorted resonance frequencies of the TC modes for a given $\alpha>0°$. Each TC mode can then be defined to evolve from each RC mode. The inclination generates two coupling mechanisms, namely, the local coupling where the RC modes couple at the inclined wall, and the global coupling where the RC modes couple throughout the trapezoidal volume. The latter arises from the nonorthogonality of the RC modes in the trapezoidal volume. Both couplings are selective that only RC modes with the same indices in the x direction are coupled to each other, where the acoustic energy in the RC mode that evolves a given TC mode is distributed among the other RC modes. Generally, at $\alpha<10°$, a TC mode is composed by only one significant RC mode, and it has the shape of this RC mode. At $\alpha\geqslant 10°$, many significant RC modes are involved and the TC-mode shape becomes complicated and unrecognizable. As far as the local and global couplings are concerned, the RC modes can be categorized into three types, namely, nonperpendicular grazing RC modes, perpendicular grazing RC modes, and nongrazing RC modes. When α is varied, the couplings change, and extrema can exist in the resonance frequencies of TC modes. For a TC mode that evolves from a nonperpendicular grazing RC mode, the existence and positions of the extrema are similar to those of the local-to-global coupling ratio, $R_{ai}^{(j)}$, of the RC mode. For a TC mode that evolves from a perpendicular grazing RC mode, there are no extrema and the resonance frequency of the TC mode is independent of α because the local coupling of the RC mode is zero, which implies $R_{ai}^{(j)}=0$. For a TC mode that evolves from a nongrazing RC mode, the occurrence of the extrema is determined by the total rate of change with α of the resonance frequency of the RC mode, ω_{ai}^2, and $R_{ai}^{(j)}$ (i.e., $\partial\omega_{ai}^2/\partial\alpha+\partial R_{ai}^{(j)}/\partial\alpha$). When this total rate changes sign from positive to negative, a maximum occurs, and vice versa for a minimum.

ACKNOWLEDGMENT

This work is funded by the Australia Research Council through the ARC Discovery Project Grant.

[1] 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).
[2] 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).
[3] J. Pan and D. A. Bies, "The effect of fluid-structural coupling on sound waves in an enclosure—Experimental part," J. Acoust. Soc. Am. **87**, 708–717 (1990).
[4] E. H. Dowell, G. F. Gorman, and D. A. Smith, "Acoustoelasticity: General theory, acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiment," J. Sound Vib. **52**, 519–542 (1977).
[5] T. C. Lin and G. W. Morgan, "Wave propagation through fluid contained in a cylindrical elastic shell," J. Acoust. Soc. Am. **28**, 1165–1176 (1956).
[6] C. R. Fuller and F. J. Fahy, "Characteristics of wave propagation and energy distribution in cylindrical elastic shells filled with fluid," J. Sound Vib. **81**, 501–518 (1982).
[7] M. El-Raheb, "Acoustic propagation in finite length elastic cylinders. Part I: Axisymmetric excitation," J. Acoust. Soc. Am. **71**, 296–306 (1982).
[8] M. El-Raheb, "Acoustic propagation in finite length elastic cylinders. Part II: Asymmetric excitation," J. Acoust. Soc. Am. **71**, 307–317 (1982).
[9] F. J. Fahy, "Acoustic coupling between structures and enclosed volumes of fluid," in *Sound and Structural Vibration—Radiation, Transmission, and Response* (Academic, London, 1987), Chap. 6, 241–269.
[10] M. Petyt, J. Lea, and G. H. Koopman, "A finite element method for determining the acoustic modes of irregular shaped cavities," J. Sound Vib. **45**, 497–502 (1976).
[11] S. H. Sung and D. J. Nefske, "A coupled structural-acoustic finite element model for vehicle interior noise analysis," J. Vib. Acoust. Stress Rel. Design **106**, 314–318 (**1984**).
[12] P. Herzog and M. Bruneau, "Shape perturbation and inertial mode coupling in cavities," J. Acoust. Soc. Am. **86**, 2377–2384 (1989).
[13] P. M. Morse and H. Feshbach, "Approximate methods," in *Methods of Theoretical Physics—Part II* (McGraw–Hill, New York, 1953), Chap. 9, 999–1064.
[14] J. Missaoui and L. Cheng, "A combined integro-modal approach for predicting acoustic properties of irregular-shaped cavities," J. Acoust. Soc. Am. **101**, 3313–3321 (1997).
[15] I. J. Cooper and H. F. Pollard, "Low frequency resonances in unsymmetrical enclosures," Acustica **41**, 86–93 (1978).
[16] Y. Y. Li and L. Cheng, "Modifications of acoustic modes and coupling due to a leaning wall in a rectangular cavity," J. Acoust. Soc. Am. **116**, 3312–3318 (2004).
[17] P. Hamery, P. Dupire, and M. Bruneau, "Acoustic fields in trapezoidal cavities," Acust. Acta Acust. 83, 13–18 (1997).
[18] F. J. Fahy, D. G. Ramble, J. G. Walker, and M. Sugiura, "Development of a novel modular form of sound absorbent facing for traffic noise barriers," Appl. Acoust. **44**, 39–51 (1995).
[19] M. J. Kingan and J. R. Pearse, "Sound absorption of porous material in combination with perforated facings," *Proceedings of Acoustics 2004*, Gold Coast, Australia, 83–87 (2004).
[20] J. Pan, R. Ming, and J. Guo, "Wave trapping barriers," *Proceedings of Acoustics 2004*, Gold Coast, Australia, 283–287 (2004).
[21] E. Kreyszig, "Vector integral calculus. Integral theorems," in *Advanced Engineering Mathematics* (Wiley, New York, 1999), 8th ed., Chap. 9, 505–513.

Mechanism and calculation of the niche effect in airborne sound transmission

Roman Vinokur[a)]
Vinokur VibroAcoustics, Woodland Hills, California 91367

(Received 12 October 2005; revised 1 February 2006; accepted 2 February 2006)

The goal is to interpret and calculate the "niche effect" for the airborne sound transmission through a specimen mounted inside an aperture in the wall between the source and receiving reverberation rooms. The low-frequency sound insulation is known to be worse for the specimen placed at the center than for the specimen mounted at either edge of the aperture. As shown, the aperture with a tested specimen can be simulated at low frequencies as a triple partition where the middle element is the specimen and the role of the edge leaves is played by the air masses entrained at the aperture edges. With a centrally located specimen, such a triple system is symmetric and has two main natural frequencies close together. In this case, the resonant transmission is higher than for the edge arrangement simulated as a double system with one natural frequency. Analogous resonant phenomena are known to reduce the low-frequency transmission loss for symmetric triple windows or solid walls with identical air gaps and lightweight boards on both sides. The theoretical results obtained for the mechanical and acoustical models are in a good agreement with the experimental data. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2179656]

PACS number(s): 43.55.Rg, 43.55.Nd, 43.55.Ti, 43.20.Bi [NX] Pages: 2211–2219

I. INTRODUCTION

The low-frequency sound insulation is of importance because in this range the typical city noise sources (automobiles, building equipment, trains, etc.) are very loud, but the transmission loss of walls and windows is commonly lower than at high frequencies. The theoretical concepts of sound insulation, first derived for infinite walls under free field conditions on both sides,[1–5] revealed such important sound transmission phenomena as the mass law, coincidence effect for single partitions, and mass-spring-mass resonances for double and triple partitions. More complicated mathematical models were developed to calculate the airborne sound transmission through a partition in an infinite soundproof baffle of the same thickness[6–9] or through a wall separating two adjacent rooms.[10–15] Statistical energy analysis (SEA),[16–18] finite element analysis (FEA), and boundary element analysis (BEA) produce numeric results rather than analytical relationships and are utilized mainly for computer modeling; besides, the SEA techniques are applicable only at relatively high frequencies. However, some low-frequency effects observed experimentally are still waiting for interpretation, in particular, the "niche effect"[19–24] (or the "tunneling effect"[25]) that influences the sound transmission loss of a test panel mounted inside an aperture in the common wall between the source and receiving rooms (according to the conventional method of two reverberation rooms[26,27]). The niche effect is the dependence of the measured transmission loss on the specimen position in the aperture. The sound transmission loss $R_{\text{edge}}(f)$, measured for the panel mounted at either edge of the aperture [Figs. 1(b) and 1(c)], can be notably higher than the sound transmission loss $R_{\text{center}}(f)$ measured for the centrally located test panel [Fig. 1(a)]; here f is the sound frequency (commonly, the center frequency of an octave or one-octave band). The difference

$$\Delta R_{\text{niche}}(f) = R_{\text{edge}}(f) - R_{\text{center}}(f) \tag{1}$$

is defined here as the niche effect. The niche effect may be responsible for the variability of low-frequency sound transmission measurements[28,29] (dissimilar results were obtained at different laboratory facilities, even though the same panels were tested and all flanking paths are cut). The other key reason for such discrepancies is the fact that at low frequencies the acoustic fields in the reverberation rooms are dominated by a few normal modes and therefore are not sufficiently diffuse.[30,31] The computer modeling study[25] confirms the experimentally observed tendencies, but does not explain the physical mechanism of the phenomenon.

II. PECULIARITIES OF RESONANT RESPONSE FOR LUMPED TRIPLE SYSTEM

A. Mathematical model for a triple system with similar and constant loss factors

1. General analytical solution

Consider mechanical triple system [Fig. 2(a)] of three masses m_1, m_2, and m_3 connected with two spring-damper sets. Harmonic force $F=F_0e^{-i\omega t}$ is applied to the first mass. Calculate the mechanical impedance $Z=F/v_3$, where v_3 is the velocity of the third mass. For a forced harmonic motion, each spring-damper set can be simulated by the imaginary spring with complex spring constant $K_j=k_j(1-i\eta)$, where k_j is the actual spring constant and η is the loss factor ($j=1, 2$; $i=\sqrt{-1}$). Here, both loss factors are considered to be similar and constant. The differential equations of the forced harmonic motion can be written as

[a)]Electronic mail: rvinokr@aol.com

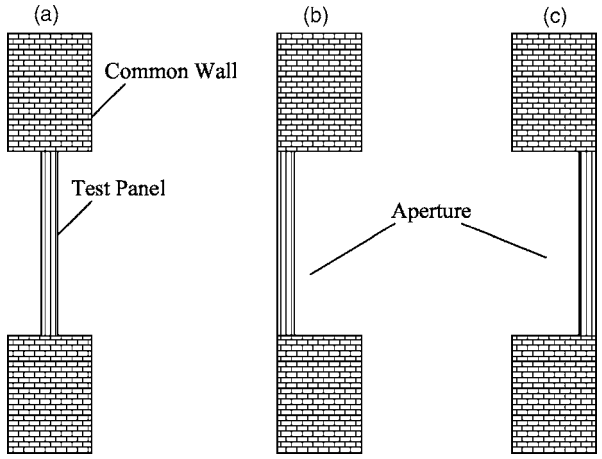


FIG. 1. Test panel mounted at the center (a) and edge (b, c) locations.

$$m_1\ddot{X}_1 + K_1(X_1 - X_2) = F_0 e^{-i\omega t},$$

$$m_2\ddot{X}_2 + K_1(X_2 - X_1) + K_2(X_2 - X_3) = 0,$$

$$m_3\ddot{X}_3 + K_2(X_3 - X_2) = 0,$$

where $X_p = X_{p0}e^{-i\omega t}$ is the displacement of the pth mass ($p = 1, 2, 3$).

Using the previous equations, obtain the velocity $v_3 = \dot{X}_3$ and therefore the mechanical impedance

$$Z_M = \frac{F}{v_3} = \frac{-im_\Sigma\omega[1 - (\omega/\omega_1)^2 - i\eta][1 - (\omega/\omega_2)^2 - i\eta]}{(1 - i\eta)^2}. \tag{2}$$

Here, the total mass $m_\Sigma = m_1 + m_2 + m_3$, and the natural angular frequencies of the system,

$$\omega_{1,2} = \sqrt{A(1 \mp \sqrt{1 - B})}, \tag{3}$$

where

$$A = \frac{1}{2}\left[k_1\left(\frac{1}{m_1} + \frac{1}{m_2}\right) + k_2\left(\frac{1}{m_3} + \frac{1}{m_2}\right)\right],$$

$$B = \frac{k_1 k_2 m_\Sigma}{m_1 m_2 m_3 A^2}. \tag{4}$$

2. Ratio of two natural frequencies and its maximum value

Using Eq. (3), determine the ratio of the fundamental natural frequency ω_1 to the second natural frequency ω_2:

$$\Omega = \frac{\omega_1}{\omega_2} = \frac{\sqrt{1 - \sqrt{1 - B}}}{\sqrt{1 + \sqrt{1 - B}}} = \frac{1}{\sqrt{B}} - \sqrt{\frac{1}{B} - 1}. \tag{5}$$

The ratio Ω expressed by Eq. (5) is a monotonically decreasing function of the parameter **B** and therefore attains its maximum at the peak value of this parameter. If p and q are two positive numbers, then $(p+q)^2 > 4pq$ if $p \neq q$ and $(p+q)^2 = 4pq$ if $p = q$. Using these supporting relationships in Eqs. (4), obtain that the parameter **B** attains its maximum value

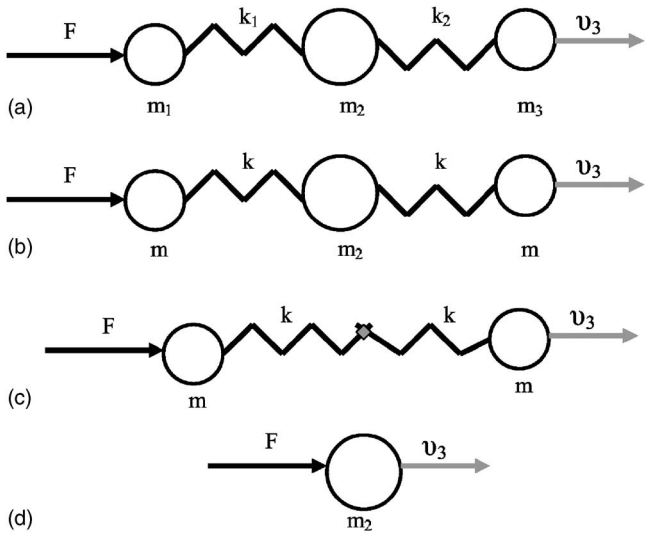


FIG. 2. Simplified mechanical models (the dampening elements are not shown).

$$B_{\max} = \frac{m_\Sigma}{m_1 m_2 m_3\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\left(\frac{1}{m_1} + \frac{1}{m_2}\right)} = \frac{1}{1 + \mu}, \tag{6}$$

provided that

$$\frac{k_1}{k_2} = \frac{\frac{1}{m_3} + \frac{1}{m_2}}{\frac{1}{m_1} + \frac{1}{m_2}}. \tag{7}$$

Here, parameter

$$\mu = \frac{m_1 m_2}{m_\Sigma m_2}. \tag{8}$$

From Eqs. (5) and (6), the maximum ratio of the natural frequencies is derived as

$$\Omega_{\max} = \sqrt{1 + \mu} - \sqrt{\mu}. \tag{9}$$

3. Natural frequencies and their ratio for a symmetric triple system

For a symmetric triple system [Fig. 2(b)],

$$k_1 = k_2 = k,$$

$$m_1 = m_3 = m. \tag{10}$$

From Eqs. (3), (4), (9), and (10), the fundamental natural frequency and the ratio of the natural frequencies for a symmetric system are given by

$$\omega_1 = \sqrt{\frac{k}{m}},$$

$$\Omega = \Omega_{\max}\sqrt{\frac{m_2}{m_\Sigma}} = \sqrt{1 - \frac{2m}{m_\Sigma}}. \tag{11}$$

From Eqs. (11), the fundamental frequency of a symmetric triple system does not depend on the middle mass, but the

natural frequencies come close together as the middle mass, increases. From Eq. (2), the closer are the natural frequencies, the lower is the mechanical impedance at each natural frequency.

B. Effect of the middle mass on the resonant response of symmetric triple system

Substituting Eqs. (11) into Eq. (2), calculate the module of the mechanical impedance for a symmetric triple system at the fundamental frequency ω_1 as

$$|Z_M(\omega_1)| = m_\Sigma \omega_1 \eta \frac{\sqrt{(1-\beta^2)^2+\eta^2}}{(1+\eta^2)} = 2m\omega_1\eta \frac{\sqrt{1+\left(\frac{m_\Sigma \eta}{2m}\right)^2}}{(1+\eta^2)}. \quad (12)$$

If the loss factor is low enough so that

$$\eta \ll \min\left\{\frac{2m}{m_\Sigma}, 1\right\}, \quad (13)$$

Eq. (12) can be brought to simplified form,

$$|Z_M(\omega_1)| \approx 2m\omega_1\eta. \quad (14)$$

From Eq. (14), the mechanical impedance of a symmetric triple system with small loss factors does not depend on the middle mass at the fundamental natural frequency. Hence, at the fundamental frequency a symmetric triple system with small loss factors has the same mechanical impedance as the "degenerate" system [the double system with the same edge masses and springs but without the middle mass, as shown in Fig. 2(c)].

This result is of importance for triple and double windows at low frequencies, where the airborne sound transmission is affected by the resonances of masses (panes) and springs (air gaps).[5,32–36] For instance, the symmetric triple glazing 3+20+10+20+3 (this descriptive code shows the pane and air gap thicknesses, mm, from left to right) proved not effective from the viewpoint of sound insulation; the low-frequency transmission loss of the asymmetric triple glazing 10+20+3+20+3, made just by simple transposition of the massive 10 mm middle pane and the 3 mm left edge pane, was up to 7 dB higher.[5]

For elevated loss factors,

$$\eta \gg \frac{2m}{m_\Sigma}, \quad (15)$$

Eq. (12) turns to the form

$$|Z_M(\omega_1)| \approx m_\Sigma \omega_1 \frac{\eta^2}{(1+\eta^2)}. \quad (16)$$

In this case, the mechanical impedance increases with the middle mass but if $\eta \leqslant 1$ and $m_2 \gg m$, it is still much less by the module than the mechanical impedance of the middle mass [Fig. 2(d)]

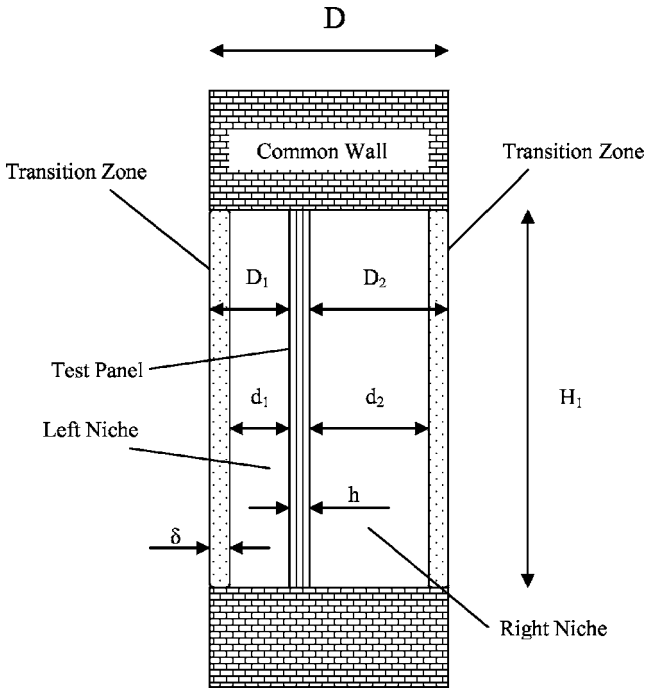


FIG. 3. Aperture with a test panel and the schematically shown transition zones.

$$|Z_2(\omega_1)| = m_2\omega_1. \quad (17)$$

So, at the fundamental frequency ω_1 the middle mass has a higher mechanical impedance than the whole symmetric triple system. This explains why the sound transmission loss of a concrete wall with gypsum boards attached on both sides by wood strapping can be less at low frequencies than the transmission loss of the individual concrete wall.[37] As shown in the next section, a similar phenomenon is also responsible for the "niche" effect.

III. THEORY OF THE "NICHE EFFECT"

A. Mathematical model

1. Basic assumptions

Consider a test panel mounted in a rectangular aperture (Fig. 3) where D and h, are, respectively, the thickness of the wall and panel, D_1 and D_2 is the depth of the left and right niche, the aperture height and width is H_1 and H_2, respectively. The sidewall surface of the aperture is rigid, so the left and right niches can be considered like two segments of a rectangular acoustical waveguide.

a. Upper frequency bound Only the plane-front sound wave (zero mode or mode 0) is not evanescent in the axial direction for the rectangular waveguides with the rigid sidewalls[38,39] if

$$f < \frac{c}{2H},$$

where

$$H = \max\{H_1, H_2\} \approx \sqrt{H_1 H_2} = \sqrt{S}, \quad (18)$$

where S is the aperture area. In practice, the aperture height H_1 and width H_2 are about similar, so, the approximation (18) is quite feasible.

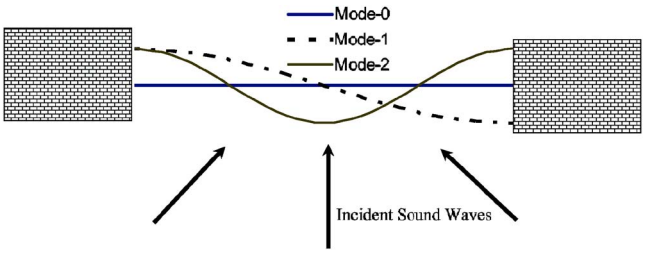


FIG. 4. (Color online) Fronts of the main acoustical waveguard modes.

At frequencies

$$f < f_{m2} = \frac{c}{H} \approx \frac{c}{\sqrt{S}}, \tag{19}$$

where f_{m2} is the cutoff frequency of the symmetrical mode 2, both mode 0 and mode 1 (Fig. 4) are not evanescent. However, the propagating magnitude of mode 1 is expected to be low in comparison to the magnitude of mode 0 and therefore can be neglected for simplicity. The reason is that commonly in both reverberation rooms the acoustical fields near the aperture are about symmetrical relative to the aperture axis, because the apertures are made in the centers of the common walls and the acoustical fields in the rooms are almost diffuse at frequencies over 100 Hz. Such a symmetrical field should not generate the modes with asymmetrical fronts inside the aperture. So, condition (1) defines the upper frequency bound for the tube[26] model, where only the zero modes are not evanescent.

b. Transition zones. However, the symmetrical modes of higher orders (in particular, mode 2) can be generated in the transition zones on both edges of the aperture (Fig. 3). The effective thickness of each zone can be evaluated[39] as $\delta \approx H/2\pi$. Here, we consider

$$\delta \approx \frac{H}{3\pi}, \tag{20}$$

where H is defined by Eq. (18). At frequencies $f \ll f_{m2}$, the δ determined by Eq. (20) is the "relaxation" distance for the magnitude of an evanescent mode 2 to reduce by half. The evanescent modes of higher orders vanish much faster. So, the effective depth of each niche can be determined as

$$d_{1,2} = \begin{cases} D_{1,2} - \delta, & \text{if } D_{1,2} > \delta, \\ 0, & \text{if } D_{1,2} \leqslant \delta. \end{cases} \tag{21}$$

c. Lower-frequency bound. The test panel is considered to consist of one or more thin panels with relatively low fundamental natural frequencies of bending vibration. According to the asymptotic Shoch's law,[7] a finite thin plate excited by high-frequency harmonic sound waves vibrates as the solid piston of the identical surface density. In practice, the same is approximately true for plain waves with the white noise spectrum in a relatively wide (octave or even third-octave) frequency band if the central frequency of the band exceeds the fundamental natural frequency of the plate at least two to three times. It is noteworthy that Shoch's law relates to the normally incident waves. Commonly, the fundamental natural frequencies of ordinary window panes, standard gypsum boards, and many other single partitions fall below 50 Hz, so, the solid piston model asserts at frequencies over 100 Hz. In this case, the single and multi-

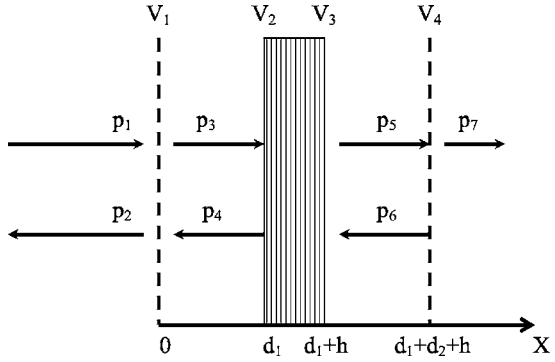


FIG. 5. Acoustic fields in four plain-wave regions inside the aperture.

player (in particular, double and triple) test panels can be described by the impedance $Z(\omega)$ calculated in the infinite size approximation.[1–5] However, this condition is not radically important because the final result (as will be shown further) does not depend notably on the test panel parameters.

d. Simulation of the reverberation rooms by the half-infinite spaces. At low frequencies, the measured transmission loss depends on the geometrical and sound absorption parameters of the reverberation rooms. However, the niche effect defined by Eq. (1) is the difference of two transmission losses measured for the same test panel. Since a shift of the test panel inside the aperture is minor relative to the room dimensions, the room effect on the transmission loss is about similar in both cases and should be eliminated by the subtraction in Eq. (1). In this case, the reverberation rooms do not affect the final result and can be simulated by the half-infinite spaces, as shown in Fig. 3.

2. Matching boundary conditions for the sound transmission

The one-degree-of-freedom (1-DOF) acoustic fields arising in the plane-wave region from an incident plane wave $p_1 e^{ikx}$ are schematically shown in Fig. 5 ($k = \omega/c$ is the wave number; the harmonic term $e^{-i\omega t}$ is omitted for simplicity). At the left edge ($x=0$) of the plane-wave region,

$$\begin{aligned} p_1 + p_2 - p_3 - p_4 &= i\rho c Y_a V_1, \\ p_3 - p_4 &= \rho c X_a V_1, \\ p_3 - p_4 &= \rho c V_1. \end{aligned} \tag{22}$$

Here, the total acoustic field at $x \leqslant 0$ consists of the incident wave $p_1 e^{ikx}$ and the reflected wave $p_1 e^{-ikx}$, and the total acoustic field at $0 \leqslant x \leqslant d_1$ is composed of the waves $p_3 e^{ikx}$ and $p_4 e^{-ikx}$; V_1 is the air velocity; $X_a(\beta)$ and $Y_a(\beta)$ are the real and imaginary parts of the normalized impedance of the air load on each aperture side that can be approximated by

the appropriate values for an equivalent plane piston[38] of the same area as

$$X_a(\beta) = 1 - \frac{J_1(2\beta)}{\beta} = \frac{\beta^2}{2}\sum_{n=1}^{\infty} a_n \beta^{2(n-1)},$$

$$Y_a(\beta) = -\frac{H_1(2\beta)}{\beta} = -\frac{8}{3}\frac{\beta}{\pi}\sum_{n=1}^{\infty} b_n \beta^{2(n-1)}, \tag{23}$$

where parameter

$$\beta = kr \tag{24}$$

and the coefficients in Eqs. (23) are defined as

$$a_1 = b_1 = 1;$$

$$a_n = -\frac{a_{n-1}}{n(n+1)}; \quad b_n = -\frac{4b_{n-1}}{4n^2-1}.$$

The radius of the equivalent round piston is given by

$$r = \frac{\sqrt{S}}{\pi}. \tag{25}$$

Functions $X_a(\beta)$ and $Y_a(\beta)$ are related, respectively, to the radiation and inertial effects; the minus sign for $Y_a(\beta)$ is introduced because of the negative exponent utilized in this paper for the harmonic term $e^{-i\omega t}$.

At the left edge of the test panel ($x=d_1$);

$$p_3 e^{i\alpha_1} - p_4 e^{-i\alpha_1} = \rho c V_2. \tag{26}$$

Here, V_2 is the velocity of the left edge of the specimen, $\alpha_1 = kd_1$. At the right edge of the specimen ($x=d_1+h$),

$$p_5 - p_6 = \rho c V_3,$$

$$p_3 e^{i\alpha_1} + p_4 e^{-i\alpha_1} - p_5 - p_6 = Z V_3. \tag{27}$$

Here, the total acoustic field at $x \geqslant d_1+h$ consists of the waves $p_5 e^{ik(x-d_1-h)}$ and $p_6 e^{ik(x-d_1-h)}$, V_3 is the velocity of the right edge of the specimen, and $Z=Z(\omega)$ is the acoustical impedance of the test panel. If the test panel is a single solid partition, the velocities of its left and right edges coincide ($V_2=V_3$).

At the right edge of the plane-wave region ($x=d=d_1+h+d_2$),

$$p_5 e^{i\alpha_2} + p_6 e^{-i\alpha_2} - p_7 = i\rho c Y_a V_4,$$

$$p_7 = \rho c X_a V_4. \tag{28}$$

$$p_5 e^{i\alpha_2} - p_6 e^{-i\alpha_2} = \rho c V_4.$$

Here, V_4 is the air velocity, the total acoustic field at $x \geqslant d$ is formed by the transmitted wave $p_7 e^{ik(x-d)}$, $\alpha_2 = kd_2$.

3. Equation for the sound transmission coefficient

Solving Eqs. (22) and (25)–(27), obtain the relationship between the amplitudes of the incident and transmitted waves as

$$p_1 - \rho c V_2 \Phi(\beta, \alpha_1) = \frac{p_7}{X_a(\beta)\Psi(\beta, \alpha_1)} \times \left(\frac{Z(\omega)}{2\rho c \Psi(\beta, \alpha_2)} + \Phi(\beta, \alpha_2) \right). \tag{29}$$

Here, the dimensionless functions

$$\psi(\beta, \alpha_{1,2}) = \frac{1}{\cos(\alpha_{1,2}) - iZ_a(\beta)\sin(\alpha_{1,2})},$$

$$\Phi(\beta, \alpha_1) = \frac{Z_a(\beta)\cos(\alpha_1) - i\sin(\alpha_1)}{2}, \tag{30}$$

where $Z_a(\beta) = X_a(\beta) + iY_a(\beta)$.

For real solid test panels

$$|p_1| \gg \rho c |V_2|,$$

$$\left| \frac{Z(\omega)}{2\rho c \Psi(\beta, \alpha_2)} \right| \gg |\Phi(\beta, \alpha_2)|,$$

where V_2 is the velocity of the left edge of the test panel, and therefore Eq. (29) can be reduced to the form

$$\frac{p_1}{p_7} \approx \frac{Z(\omega)}{2\rho c X_a \Psi(\alpha_1)\Psi(\alpha_2)}. \tag{31}$$

So, the coefficient of the sound transmission through the specimen in the aperture is given by

$$\tau(f) = \left| \frac{p_7}{p_1} \right|^2 \approx \tau_0(f) Q(\beta, \alpha_1) Q(\beta, \alpha_2), \tag{32}$$

where

$$\tau_0(f) = \left| \frac{2\rho c X_a(\beta)}{Z(\omega)} \right|^2 \tag{33}$$

is the coefficient of the sound transmission through in case $d_{1,2}=0$ (for the specimen mounted in the soundproof baffle of the same thickness), and function

$$Q(\beta, \alpha) = |\Psi(\beta, \alpha)|^2 = \frac{1}{\cos^2(\alpha) + Y_a(\beta)\sin(2\alpha) + [X_a^2(\beta) + Y_a^2(\beta)]\sin^2(\alpha)} \tag{34}$$

describes the acoustical effects of the niches. Here, parameter $\alpha = kd$, where d stands for d_1 or d_2, and parameter $\beta = kr = \gamma\alpha$, where

$$\gamma = \frac{r}{d}. \tag{35}$$

If $\alpha \to \infty$, then $\beta \to \infty$, $X_a(\beta) \to 1$, $Y_a(\beta) \to 0$, and therefore $Q(\beta, \alpha) \to 1/[\cos^2(\alpha) + \sin^2(\alpha)] = 1$. If $\alpha \to 0$, then β, $X_a(\beta)$, $Y_a(\beta) \to 0$ and therefore $Q(\beta, \alpha) \to 1$. For α between 0 and ∞, function $Q(\beta,\alpha)$ has multiple maximums and minimums, but the first maximum is of major importance. Evaluate the first natural frequency and the loss factor in the practical case $\alpha \ll 1$, so that $\sin(\alpha) \approx \alpha$, $\cos(\alpha) \approx 1 - (\alpha^2/2)$, $X_a(\beta) \approx \beta^2/2 = (\gamma\alpha)^2/2$, $Y_a(\beta) \approx -8\beta/3\pi = -8\gamma\alpha/3\pi$.

Substituting the above expressions into Eq. (34), obtain

$$Q(\beta,\alpha) \approx \frac{1}{1-\left(\frac{f}{f_n}\right)^2+\eta(f)}, \tag{36}$$

where the natural frequency and the effective loss factor are given by

$$f_n = \frac{c}{2\pi\sqrt{\frac{8rd}{3\pi}+\frac{d^2}{2}}},$$
$$\eta(f) = \frac{4\pi^2 dSf^3}{c^3}. \tag{37}$$

Substituting Eq. (36) into Eq. (32) produces the relationship similar to Eq. (2) for a lumped triple system, but with the loss factor growing with frequency; in this case, the first resonant frequency (the frequency of maximum forced amplitude) is below the natural frequency.[40] The loss factor η defined by Eq. (37) is determined by the sound radiation from the aperture. If $d=0.2$ m and $S=1$ m^2, the calculated loss factor is relatively high; in particular, it equals 0.2, 0.4, and 0.8 at frequencies 100, 125, and 160 Hz. This indicates that function $Q(\beta,\alpha)$ does not have pronounced peaks at frequencies over 100 Hz.

B. Results

1. Derivation of the niche effect equation

The transmission loss determined for a one-third octave frequency band with the center frequency f_c is given by equation

$$R(f_c) = -10\,\mathrm{Log}\left(\frac{1}{f_2-f_1}\int_{f_1}^{f_2}\tau(f)df\right)$$
$$= -10\,\mathrm{Log}\left(\frac{1}{f_2-f_1}\int_{f_1}^{f_2}\tau_0(f)P(\beta,\alpha_1,\alpha_2)df\right), \tag{38}$$

where $f_1=f_c/2^{1/6}$ and $f_2=f_c\cdot 2^{1/6}$ are, respectively, the lower and upper bounds of the one-third octave band, the coefficients of the sound transmission $\tau(f)$ and $\tau_0(f)$ are determined by Eqs. (32) and (33), and the function

$$P(\beta,\alpha_1,\alpha 2) = Q(\beta,\alpha_1)Q(\beta,\alpha_2). \tag{39}$$

Since $\tau_0(f)$ and $P(\beta,\alpha_1,\alpha 2)$ are continuous functions and $\tau_0(f)>0$, the integral in Eq. (38) can be transformed to the form

$$\int_{f_1}^{f_2}\tau_0(f)P(\beta,\alpha_1,\alpha_2)df = P(\tilde{\beta},\tilde{\alpha}_1,\tilde{\alpha}_2)\int_{f_1}^{f_2}\tau_0(f)df, \tag{40}$$

where parameters $\tilde{\beta}$, $\tilde{\alpha}_1$, and $\tilde{\alpha}_2$ are calculated for frequency $\tilde{f}\in(f_1,f_2)$.[41]

Since a one-third octave band is relatively narrow, it suggests that

$$P(\tilde{\beta},\tilde{\alpha}_1,\tilde{\alpha}_2) \approx \bar{P}(f_1,f_2), \tag{41}$$

where

$$\bar{P}(f_1,f_2) = \frac{1}{f_2-f_1}\int_{f_1}^{f_2}P(\beta,\alpha_1,\alpha_2)df \tag{42}$$

is the average value of function $P(\beta,\alpha_1,\alpha_2)$ over the interval $[f_1,f_2]$, Substituting Eqs. (40)–(42) into Eq. (38), obtain

$$R(f_c) = R_0(f_c) - 10\,\mathrm{Log}[\bar{P}(f_1,f_2)], \tag{43}$$

where

$$R_0(f_c) = -10\,\mathrm{Log}\left(\frac{1}{f_2-f_1}\int_{f_1}^{f_2}\tau_0(f)df\right) \tag{44}$$

is a function of the test panel parameters and does not depend on the aperture parameters. Using Eqs (1) and (43), calculate the niche effect as

$$\Delta R_{\mathrm{niche}}(f_c) \approx 10\,\mathrm{Log}\left(\frac{\bar{P}_{\mathrm{center}}(f_1,f_2)}{\bar{P}_{\mathrm{edge}}(f_1,f_2)}\right), \tag{45}$$

where

$$\bar{P}_{\mathrm{center}}(f_1,f_2) = \frac{1}{f_2-f_1}\int_{f_1}^{f_2}[Q(\beta,\alpha_{\mathrm{center}})]^2\,df,$$
$$\bar{P}_{\mathrm{edge}}(f_1,f_2) = \frac{1}{f_2-f_1}\int_{f_1}^{f_2}Q(\beta,\alpha_{\mathrm{edge}})df. \tag{46}$$

Here, variables

$$\alpha_{\mathrm{center}} = k\left(\frac{D_{\mathrm{net}}}{2}-\delta\right),$$
$$\alpha_{\mathrm{edge}} = k(D_{\mathrm{net}}-\delta), \tag{47}$$

where the net aperture depth $d_{\mathrm{net}}=D\text{-}h$. The niche effect calculated by Eq. (45) does not depend on the test panel parameters but depends on frequency and the aperture area S (since the transition zone depth δ and effective radius r are the functions of area S).

2. Calculation for typical cases

The depth of typical test apertures is about 0.6 m. If the thickness of the test panel is 0.05 m, the net aperture depth $D_{\mathrm{net}}=0.55$ m. Consider that the aperture and test panel has the same area. The calculation is done for three aperture areas: 1.0, 2.0, and 5.5 m^2. The value of 5.5 m^2 (60 ft^2) is recommended by the ASTM Standard,[27] but areas of 1–2 m^2 are common if windows are tested. As follows from the plotted results (Fig. 6), the niche effect is considerable (up to 5 dB) for $S=1.0$ m^2, notable (up to 2 dB) for $S=2.0$ m^2, and negligible for $S=5.5$ m^2.

The numeric estimation proves that the maximum error of the approximation (41) does not exceed 2 dB for $S=1$ m^2 and 1 dB for $S\geqslant 2$ m^2. So, the maximum "niche effect" may reach 5–7 dB for $S=1.0$ m^2, 2–3 dB for $S=2.0$ m^2, and reduces next to zero with the aperture area.

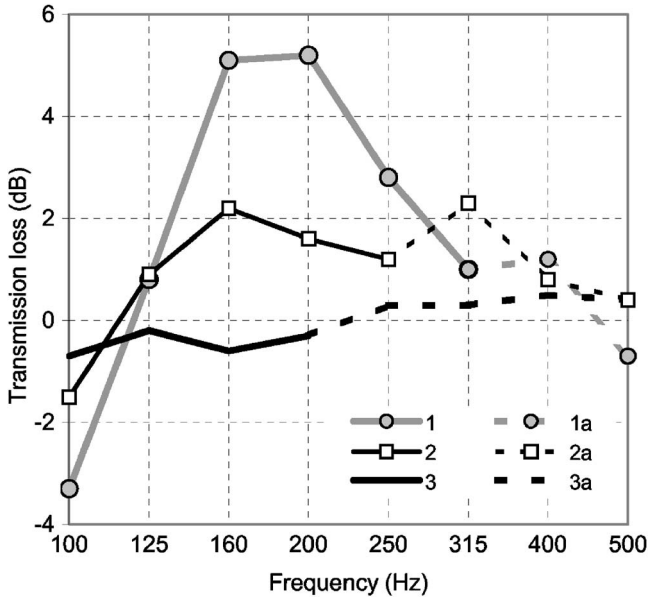


FIG. 6. The "niche effect" for aperture areas of 1 (1), 2 (2), and 5.5 m^2 (3) calculated by Eq. (45) in case the net aperture depth is 0.55 m. The trends (1a–3a) at frequencies exceeding the upper bound (19) are dotted.

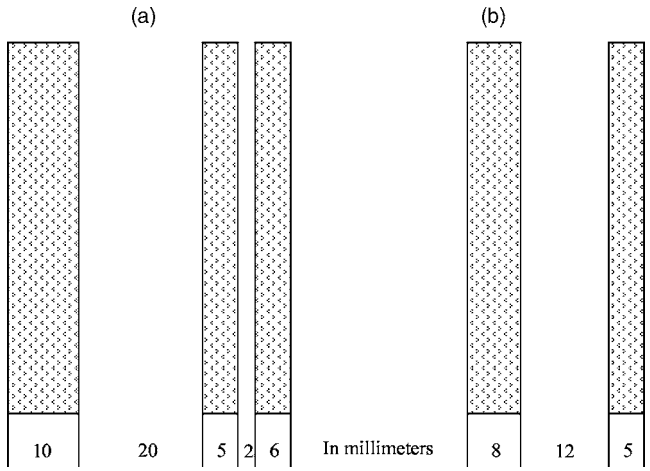


FIG. 8. Experimental specimens: (a) triple glazing; (b) double glazing. The structural links connecting the panes over their perimeter are not shown.

IV. EXPERIMENTAL RESULTS

A. Experimental setup

The experimental setup is shown in Fig. 7. The width of the source and receiving reverberation room was, respectively, 3.3 and 4.5 m; the length and height of both rooms was 6.0 and 3.0 m. So, the volume of the source and receiving rooms was about 60 and 80 m^3. The average sound absorption coefficient in each room varied from 0.07 at 63 Hz to 0.15 at 5000 Hz. The pink noise loudspeaker was located in the corner away from the test aperture. There were seven fixed microphone positions in each room. The microphone was installed at a height of 1.5 m, the distance between the microphone positions and adjacent walls was at least 1.2 m. The main specimen was the triple glazing with the descriptive code 10+20+5+2+6 (glass layer+air gap+glass layer +air gap+glass layer in mm) schematically shown in Fig. 8, its height H_1 and width H_2 were, respectively, 1.3 and 0.9 m. The pane edges were connected along the whole perimeter by a hollow 20 mm aluminum profile or a 2 mm layer of mastic, functioning as a firm frame and sealed with thiocol. The specimen was mounted in a wood frame in the middle [Fig. 1(a)] or edge sections [Figs. 1(b) and 1(c)] of the test aperture made in the brick wall separating the source and receiving rooms. The wall thickness D was 0.6 m. The aperture and specimen had similar dimensions because the wood frame was built in the test aperture walls.

B. Comparison of the experimental and calculation results for the niche effect

The airborne sound transmission loss was measured in one-third octave bands with the center frequencies 100–5000 Hz. The experimental results are plotted for comparison in Fig. 9. For both edge arrangements, the transmission loss data are almost similar and notably exceed the data for the center specimen location at low frequencies. However, all three curves approximately coincide at frequencies over 630 Hz. The difference between the average transmission loss for the edge locations and the transmission loss measured for the center location [the niche effect defined by Eq. (1)] is plotted to the first decimal place in Fig. 10. The difference is notable the frequency range 125–400 Hz, the maximum value of about 5 dB is attained at frequency 200 Hz. These experimental results were obtained at the Laboratory of Building Acoustics in Moscow.[24] A similar experimental study was done by the author at the same testing facility on a double glazing with the descriptive code 8 +12+5 [8 and 5 mm panes separated by a 12 mm air gap (Fig. 8(b)]. The measured niche effect proved very close to

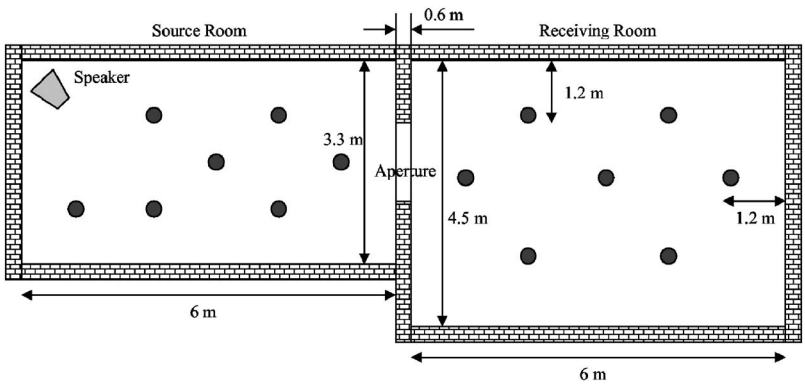


FIG. 7. Experimental setup for sound transmission loss measurement.

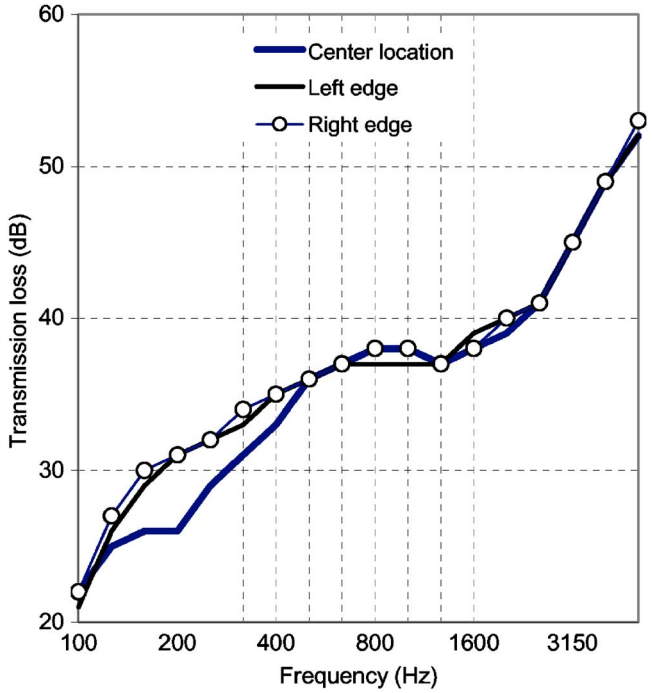


FIG. 9. (Color online) The transmission loss measured for the specimen in Fig. 8(a) at the center, left edge, and right edge locations.

the appropriate data obtained for the triple glazing. It is noteworthy that the curve calculated by Eq. (45) closely matches the experimental curve (Fig. 10).

V. CONCLUSION

A physical mechanism of the niche effect was studied. At low frequencies, the aperture with a specimen (single or multilayer test panel) can be simulated by a triple partition, where the specimen stands for the middle leaf and the air masses entrained at the aperture edges work as the edge leaves. If such a triple system is symmetric (the edge leaves are identical and the air gaps are similar), its two natural frequencies are close together because the masses of the edge leaves are much smaller than the mass of the middle leaf. If the middle leaf is shifted to one of the edge leaves, such an arrangement can be simulated by a double system with only one natural frequency. The resonant sound transmission through a system with two nearby natural frequencies used to be intensive, the known practical examples are symmetric triple windows or solid walls with identical air gaps and lightweight boards on both sides. The relationship (45) was derived to calculate the niche effect as the difference in the sound transmission loss for the center and edge locations of the specimen. The frequency domain includes one-third octave bands with the center frequencies between 100 Hz and the upper frequency bound defined by Eq. (19). To a reasonable approximation, the niche effect does not depend on the specimen parameters but depends on the aperture dimensions and on the sound frequency. As the sound frequency and/or the aperture area increases, the inertial impedance to the sound waves at the aperture edges degrades and the niche effect vanishes. If the area S of the apertures and test panels exceeds 60 ft^2 (5.5 m^2—as recommended by the E 336-97 ASTM standard), the niche effect is negligible. However, the aperture area is commonly about $1-2$ m^2 if windows or small wall fragments are tested; in this case, the calculated niche effect can be up to 5–7 dB for $S=1$ m^2 and 2–3 dB for $S=2$ m^2. A close agreement between the calculated and experimental data indicates that the mechanism mentioned above plays a major role in the "niche effect," although secondary phenomena may also occur.

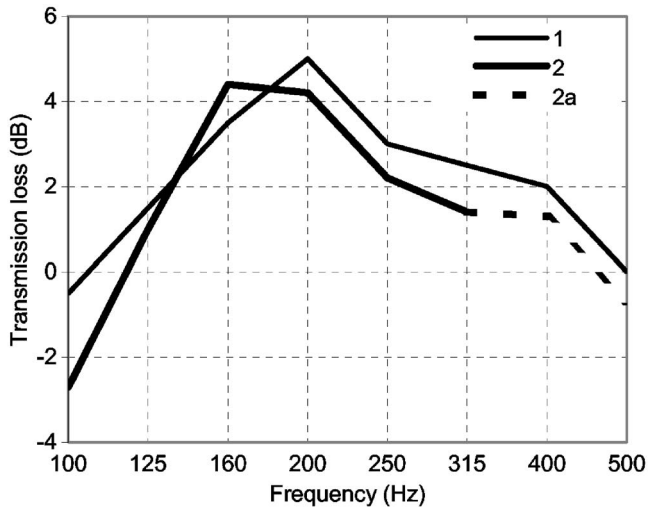


FIG. 10. The niche effect measured (1) and calculated (2–2a) for the specimen sketched in Fig. 8(a). The aperture area is 1.17 m^2 and the net aperture depth is 0.55 m. The trend at frequencies exceeding the upper bound (19) is dotted.

[1]H. Reissner, "Der senkrechte und shräge durchtutt einer in einem flüssingen medium erzeugten ebenen dilatationswelle durch eine in diesem medium befindliche planparallele, feste platte ("Normal and oblique transmission of a plain wave propagating in fluid through a flat solid panel submerged into this fluid")," Helv. Phys. Acta **11**, 140–155 (1938).
[2]L. Cremer, "Theorie der schalldämmung wände bei schrägen einfall ("Theory of sound insulation of a panel at oblique sound incidence")," Akust. Zeits. **7**, 81–104 (1942).
[3]A. London, "Transmission of reverberant sound through double walls," J. Acoust. Soc. Am. **22**, 270–279 (1950).
[4]K. A. Mulholland, H. D. Parbrook, and A. Cummings, "The transmission loss of double panels," J. Sound Vib. **6**, 324–334 (1967).
[5]R. Vinokur, "Transmission loss of triple partitions at low frequencies," Appl. Acoust. **29**, 15–24 (1990).
[6]A. Cummings and K. A. Mulholland, "The transmission loss of finite size double panels in a random incidence sound field," J. Sound Vib. **8**, 126–133 (1968).
[7]A. Schoch "Über ein asymptotishe verhalten von erzwungenen plattenschwingungen bei hohen frequenzen ("On asymptotic law of forced vibration of panel at high frequencies")," Akust. Z. **2**, 113–128 (1937).
[8]E. S. Sewell, "Transmission of reverberant sound through a single-leaf partition surrounded by an infinite rigid baffle," J. Sound Vib. **12**, 21–32 (1970).
[9]E. L. Shenderov, *Wave Problems of Hydroacoustics* (Sudostroenie, Leningrad, 1972) (in Russian).
[10]P. Josse and R. H. Lamure, "Transmission du son par une paraisimple ("Sound transmission through a single wall")," Acustica **14**, 266–280 (1964).
[11]A. C. Nilsson, "Reduction index and boundary conditions for a wall between two rectangular rooms. Part I: Theoretical results," Acustica **26**, 1–18 (1972).
[12]K. A. Mulholland and R. H. Lyon, "Sound insulation at low frequencies," J. Acoust. Soc. Am. **54**, 867–878 (1973).
[13]H. Romilly, "Exact solution for guided sound transmission through a simply supported plate," Acustica **28**, 234–237 (1973).
[14]H. Romilly, "Exact solution for guided sound transmission through a double partition," J. Sound Vib. **48**, 243–249 (1976).

[15]R. Vinokur, "Low-frequency increase of the acoustic noise reduction for lightweight panel," Sov. Phys. Acoust. **29**, 239–240 (1983).
[16]G. Maidanik, "Response of ribbed panels to reverberant acoustic fields," J. Acoust. Soc. Am. **34**, 809–826 (1962).
[17]M. J. Crocker and A. J. Price, "Sound transmission using statistical energy analysis," J. Sound Vib. **9**, 469–486 (1969).
[18]A. J. Price and M. J. Crocker, "Sound transmission through double panels using statistical energy analysis," J. Acoust. Soc. Am. **47**, 683–693 (1970).
[19]M. Heckl, "The tenth Sir Richard Fairey Memorial lecture: sound transmission in building," J. Sound Vib. **77**, 165–189 (1981).
[20]A. C. C. Warnock, "Influence of specimen frame on sound transmission loss measurement," Appl. Acoust. **15**, 307–314 (1982).
[21]R. W. Guy and P. Sauer, "The influence of sills and reveals on sound transmission loss," Appl. Acoust. **17**, 453–476 (1984).
[22]A. Cops and M. Minten, "Comparative study between the sound intensity method and the conventional two-room method to calculate the sound transmission loss of wall constructions," Noise Control Eng. J. **22**, 104–111 (1984).
[23]R. E. Halliwell and A. C. C. Warnock, "Sound transmission loss: Comparison of conventional techniques with sound intensity techniques," J. Acoust. Soc. Am. **77**, 2094–2103 (1985).
[24]R. Vinokur, *Sound Insulation by Windows* Ph.D. dissertation, Central Research Institute of Building Physics, Moscow, Russia, 1986.
[25]B.-K. Kim, H.-J. Kang, J.-S. Kim, H.-S. Kim, and S.-R. Kim, "Tunneling effect in sound transmissionloss determination: Theoretical approach," J. Acoust. Soc. Am. **115**, 2100–2109 (2004).
[26]L. L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1996), pp. 324–327.
[27]E 336-97 Standard Test Method for Measurement of Airborne Sound Insulation in Buildings, Annual Book of ASTM Standards **04.06**, 2004, pp. 885–894.
[28]R. E. Jones, "Intercomparisons of laboratory determinations of airborne sound transmission loss," J. Acoust. Soc. Am. **66**, 148–164 (1979).
[29]D. B. Pedersen, J. Roland, G. Raabe, and W. Maysenholder, "Measurement of the low-frequency sound insulation of building components," Acust. Acta Acust. 86, 495–505 (2000).
[30]W. A. Utley, "Single leaf transmission loss at low frequencies," J. Sound Vib. **8**, 256–261 (1968).
[31]T. Bravo and S. J. Elliot, "Variability of low frequency sound transmission measurements," J. Acoust. Soc. Am. **115**, 2986–2997 (2004).
[32]W. A. Utley and B. L. Fletcher, "Influence of edge conditions on the sound insulation of windows," Appl. Acoust. **2**, 131–136 (1969).
[33]J. A. Marsh, "The airborne sound insulation of glass: Parts 1, 2, and 3," Appl. Acoust. **4**, 55–70 (1971; **4**, 131–154 (1971). **4**, 175–195 (1971).
[34]A. Cops and H. Myncke, "Sound insulation of glass by means of scale models," Acustica **31**, 143–149 (1974).
[35]J. D. Quirt, "Sound transmission through windows I. Single and double glazing," J. Acoust. Soc. Am. **72**, 834–844 (1982).
[36]J. D. Quirt, "Sound transmission through windows II. Double and triple glazing," J. Acoust. Soc. Am. **74**, 532–542 (1983).
[37]A. C. C. Warnock and J. D. Quirt, "Airborne sound insulation" in *Handbook on Acoustical Measurements and Noise Control*, edited by C. M. Harris (Acoustical Society of America, New York, 1998), p. 31.16.
[38]E. Skudrzyk, *The Foundations of Acoustics* (Springer-Verlag, New York, 1971), Chap. 17.
[39]M. Isacovitch, *The General Acoustics* (Nauka, Moscow, 1973), in Russian.
[40]R. Vinokur, "The relationship between the resonant and natural frequency for non-viscous systems," J. Sound Vib. **267**, 187–189 (2003).
[41]G. A. Korn and T. M. Korn, *Mathematical Book for Scientists and Engineers* (McGraw-Hill, New York, 1961), Chaps. 4 and 17.

Automatic determination of the number of targets present when using the time reversal operator

Angela Quinlan
E.N.S. de Cachan, 61, Avenue du President Wilson 94235, Cachan Cedex, France and Department Electronic and Electrical Engineering, Trinity College, Dublin, Ireland

Jean-Pierre Barbot and Pascal Larzabal
1E.N.S. de Cachan, 61, Avenue du President Wilson 94235, Cachan Cedex, France

(Received 21 April 2005; revised 30 January 2006; accepted 3 February 2006)

Acoustical time reversal mirrors have been shown to provide a highly accurate means of studying and focusing on acoustical sources. The DORT method is a derivation of the time reversal process, which allows for focusing on multiple targets. An important step in this process is the determination of the number of targets or sources present. This is achieved by examining the eigenvalues of the time reversal operator (TRO). The number of significant eigenvalues is then chosen as the number of sources present. However, as mentioned in [N. Mordant, C. Prada, and M. Fink, J. Acoust. Soc. Am. **105**, 2634–2642 (1999) and C. Prada, M. Tanter, and M. Fink, in Proceedings of the IEEE Symposium, 1997, pp. 679–683], factors such as low signal to noise ratio (SNR), small data sample, array configuration and the target location may result in the eigenvalues corresponding to the targets no longer being distinguishable from the background noise eigenvalues. This paper proposes a robust method of automatically determining the number of targets even in the presence of a small number of snapshots. For white Gaussian noise, the profile of the ordered eigenvalues is seen to fit an exponential law. The observed eigenvalues are then compared to this model and a mismatch is detected between the observed profile and the noise-only model. The index of the mismatch gives the number of scatterers present. [DOI: 10.1121/1.2180207]

PACS number(s): 43.60.Ac, 43.38.Hz [DRD] Pages: 2220–2225

I. INTRODUCTION

In recent years, the use of acoustical time reversal mirrors and the DORT method have been shown to be highly accurate methods of focusing a sound field on either a single, or else multiple targets.[1,2] The use of such techniques has been demonstrated in a wide variety of applications, for example, underwater acoustics,[3] focusing a sound field in a reverberant room[4] and lithotripsy.[5] The DORT method, which consists of the eigendecomposition of the time reversal operator (TRO) (Refs. 6 and 7) is a variation of the original time reversal process. Assuming the scatterers are pointlike, the number of significant eigenvalues of the TRO then corresponds to the number of scatterers (or targets) and the associated eigenvectors give the phase that must be applied to the transducer array elements in order to focus on the targets.

The tests available in the literature act as a rank determination based on the evaluation of the multiplicity of the smallest eigenvalue. The underlying principles commonly used for this are only asymptotically valid. However, a small number of snapshots, the presence of defects with low reflectivity, or defects located near an interface[6] will result in a very small difference between the noise and target eigenvalues leading to nondetection. The key point of the approach proposed in this paper is that it directly takes into account the situation of a small number of snapshots, which leads to the failure of traditional tests.

The proposed algorithm is based on the analysis of the ordered noise eigenvalue profile. The observed eigenvalues are then compared to this theoretical noise-only profile and the number of eigenvalues that differ significantly from the profile by more than a threshold is equal to the number of scatterers present. While the test is based on the eigenvalue profile of Gaussian white noise, experimental results reported by various groups of researchers indicate its potential interest in practical situations, for examples, see Refs. 8 and 9. From these references it can be clearly seen that the noise eigenvalues found experimentally also follow an approximately exponential profile. This is one of the principal motivations of our work.

The layout of the paper is as follows: Sec. II gives the notation used throughout the paper. Section III formulates the basic problem of multiple point source detection and explains why the tests traditionally used in array signal processing are not retained in this work. The model for the noise-only eigenvalue profile is introduced in Sec. IV. Section V proposes an exponential fitting test (EFT) exploiting this profile. Conclusions concerning the method are drawn in Sec. VI.

II. NOTATION USED

d	=	number of scatterers
$\hat{d}$	=	estimated number of targets
M	=	number of sensors
N	=	number of transmit vectors
$\mathbf{r}(t)=[r_1(t)\cdots r_M(t)]^T$	=	vector of received signals
$\mathbf{e}(t)=[e_1(t)\cdots e_d(t)]^T$	=	vector of transmit signals

 0001-4966/2006/119(4)/2220/6/$22.50

$\mathbf{K}(t)$ = interelement transfer matrix
$\mathbf{e}(\omega)$ = Fourier transform of $\mathbf{e}(t)$
$\mathbf{r}(\omega)$ = Fourier transform of $\mathbf{r}(t)$
$\mathbf{K}(\omega)$ = Fourier transform of $\mathbf{K}(t)$
$\mathbf{R}_K$ = covariance of $\mathbf{K}(\omega)$
$\hat{\mathbf{R}}_K$ = estimated covariance of $\mathbf{K}(\omega)$
λ_m = mth eigenvalue of $\mathbf{R}_K$
δ_{ij} = Kronecker delta function

III. PROBLEM FORMULATION

The application of a time reversal mirror (TRM) is known to be a highly accurate method of focusing an array of transducers on a scatterer (or source) located in an inhomogeneous material.[1] Discussions and applications occur widely in literature, for example, see Refs. 2 and 4. Iteration of the time reversal process has been shown to result in convergence on the most reflective of the targets (or the source with the highest amplitude).[10] However, in some situations it may be desirable to focus on less reflective targets, or on numerous targets simultaneously. In such cases the DORT (french acronym for decomposition of the time reversal operator) technique, which is based on the manipulation of the TRO can be used.[6,7]

The first step in deriving the TRO is to find the interelement response matrix $\mathbf{K}(t)$. This matrix is measured by emitting a short pulse from each array element in turn and measuring the resulting response across the array. Then for transmit signal vector $\mathbf{e}(t)$, the output signal vector is $\mathbf{r}(t)$,

$$\mathbf{r}(t) = \mathbf{K}(t) \otimes \mathbf{e}_m(t), \tag{1}$$

where $\otimes$ is the convolution operator.

Transformation of Eq. (1) into the frequency domain gives

$$\mathbf{r}(\omega) = \mathbf{K}(\omega)\mathbf{e}(\omega). \tag{2}$$

$\mathbf{r}(\omega)$, $\mathbf{K}(\omega)$, and $\mathbf{e}(\omega)$ are now evaluated at a single frequency, usually the center frequency of the transducers. For simplicity of notation, the ω dependency will be omitted in the sequel.

This leads to the definition of the TRO,[11]

$$\mathrm{TRO} = \mathbf{K}^H\mathbf{K}. \tag{3}$$

If the scatterers are assumed to be pointlike, then the number of significant eigenvalues of the TRO is equal to the number of scatterers present. The eigenvalue associated with a scatterer is proportional to the reflectivity of the scatterer and its position relative to the array.[12] The eigenvector $\mathbf{v}_i$ corresponding to the eigenvalue λ_i gives the phase information necessary to focus on the associated target.

An acquisition of the TRO operator requires N snapshots thereby providing a set of measurement vectors: $\{\mathbf{r}_i, 1 \leqslant i \leqslant N\}$ and as mentioned by Prada *et al.* in Ref. 11, the TRO operator can be considered as an estimate $\mathbf{R}_K$ of the covariance matrix of the kind introduced in passive source detection,

$$\mathbf{R}_K = \frac{1}{N}\sum_{i=1}^{N} \mathbf{r}_i\mathbf{r}_i^H. \tag{4}$$

Finding the eigenvalues of the TRO is equivalent to finding the eigenvalues of $\mathbf{R}_K$. If the eigenvalues of $\mathbf{R}_K$ are arranged in order of decreasing value, in the presence of d targets or sources, $\mathbf{R}_K$ can be rewritten as[9]

$$\mathbf{R}_K = \sum_{i=1}^{d} \lambda_i\mathbf{v}_i\mathbf{v}_i^H + \sum_{i=d+1}^{M} \lambda_i\mathbf{v}_i\mathbf{v}_i^H. \tag{5}$$

When there is no noise present, the $(M-d)$ smallest eigenvalues equal zero. However, in practice during measurement of the transfer matrix, noise (electronic and quantization or natural) will be present. Because of this noise the $(M-d)$ smallest eigenvalues are no longer equal to zero. In the asymptotic case (very large N) they are almost equal to σ^2.

The multiplicity of the smallest eigenvalue is $(M-d)$. When the SNR is high or N is large a threshold T can be applied to the ordered eigenvalues of $\mathbf{R}_K$. The number of targets can be determined as $\hat{d}$, the number of eigenvalues greater than this threshold; $\hat{d}$ is the number of elements of λ_i such that $\lambda_i > T$. This approach is classically used in time reversal literature, e.g., Refs. 6 and 13. In this paper it is referred to as the threshold based test (TBT).

In the field of high resolution array signal processing a lot of work has been published concerning the model order selection problem. Anderson[14] gave a hypothesis testing procedure based on the confidence interval of the noise eigenvalue, in which a threshold value must be assigned subjectively. This test is closely connected with the previously mentioned TBT. In Ref. 15, Chen *et al.* proposed a method, based on an *a priori* determination of the observation probability density function, that detects this number by setting an upper bound on the value of the eigenvalues. For 30 years information theoretic criteria (ITC) have been widely suggested approaches for this multiple detection problem.[16] The best known of this test family are the Akaike Information Criterion (AIC) (Ref. 17) and the Minimum Description Length (MDL).[18–20] An alternative using the Bayesian methodology was proposed by Djurić.[21,22] Some authors have also investigated the possible use of eigenvectors for model order selection.[23,24]

Time reversal experimentation is not performed in such ideal asymptotic conditions and there are several factors that will result in the smallest eigenvalues no longer being equal. In Ref. 8 Prada *et al.* show that reduction of the data length by the use of subarrays can, in some cases, lead to difficulties in distinguishing between noise and signal eigenvalues, which may result in nondetection of some target. Low SNR will also result in a small difference between the signal and noise eigenvalues. In Ref. 6 it is shown that if a target is placed near an interface, the resulting eigenvalue will be reduced, once again making it difficult to distinguish from noise eigenvalues.

It appears that at low SNR and/or small number of snapshots, the results of ITC based tests degrade rapidly,[25] therefore such a strategy is not applicable any longer in our difficult scenario. Few works have been concerned with model order selection under such constraints. An interesting approach at low SNR, based on the ratio of eigenvalues, has

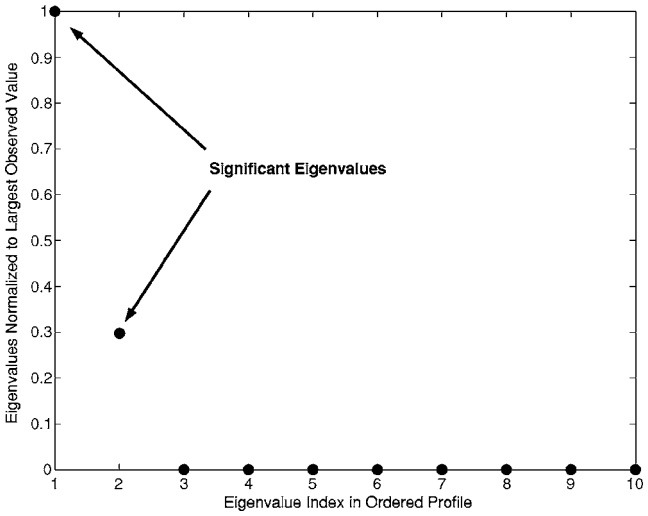


FIG. 1. Profile of ordered eigenvalues for SNR of 15 dB. It can be seen that the TBT will correctly determine the number of targets or scatterers, as will the EFT.

been proposed by Liavas *et al.*[26] for determining the length of the impulse response in the case of a multipath propagation channels.

Investigation of the scenario of two targets, located at angles of 0° and 10°, impinging on a uniform linear array with $M=10$ half-wavelength separated sensors, illustrates the influence of the low SNR and/or small number of snapshots on the behavior of the ordered normalized eigenvalues. The correlation matrix will be estimated using $N=10$ snapshots. Figure 1 shows the eigenvalue profile for SNR=15 dB. It can be seen that a TBT approach will correctly determine the number of targets in such an easy scenario. Figure 2 shows the normalized eigenvalue profile for SNR=0 dB. In this case the TBT will no longer work. In order to overcome the limitations of the TBT, which is the most commonly used test in time reversal literature, this paper proposes a new strategy capable of handling this difficult scenario, leading to the EFT.

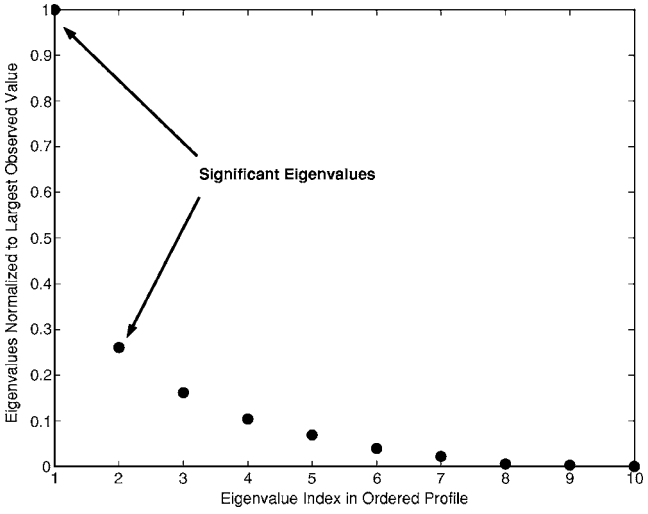


FIG. 2. Profile of ordered eigenvalues for SNR of 0 dB. It can now be seen that the TBT is no longer able to effectively discern between the noise and target eigenvalues, whereas the EFT continues to work.

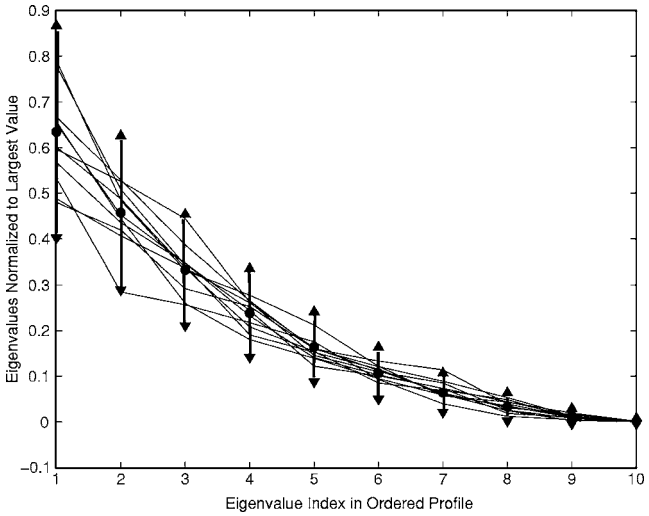


FIG. 3. Profile of ordered noise eigenvalues for several realizations. The circles through the center show the mean value for each eigenvalue. The distance between the upper and lower triangles is the spread of the eigenvalue and the chosen threshold is equal to half this distance.

IV. EIGENVALUE PROFILE UNDER NOISE ONLY ASSUMPTION

In the presence of zero-mean white Gaussian noise with power σ^2 the correlation matrix $\mathbf{R}_K$, as defined in Eq. (3), has a Wishart distribution with N degrees of freedom. It is a multivariate generalization of the χ^2 distribution and depends on M, N, σ^2.

The eigenvalue profile of the noise only Wishart covariance matrix $\hat{\mathbf{R}}_K$ will be approximated. Figure 3 shows the profiles of the ordered noise eigenvalues for numerous realizations for an array of $M=10$ sensors and using $N=10$ snapshots. A continuously decreasing profile appears resulting in failure of the TBT strategy. For this reason the decreasing profile is modeled with the help of first and second order moments of the eigenvalues.

The error of the covariance matrix is denoted by ψ, where $\psi=\hat{\mathbf{R}}_K-\mathbf{R}_K=\hat{\mathbf{R}}_K-E\{\hat{\mathbf{R}}_K\}=\hat{\mathbf{R}}_K-\sigma^2 I$.

The noise eigenvalue profile can be found by considering the first and second moments of $\mathrm{tr}[\psi]$ and from $E\{\mathrm{tr}[\psi]\}=0$, it follows that

$$M\sigma^2=\sum_{i=1}^{M}\lambda_i. \tag{6}$$

When the noise is white Gaussian complex and circular it can be shown that

$$E\{\|\psi_{ij}\|^2\}=\frac{\sigma^4}{N}. \tag{7}$$

The trace of a matrix remains unchanged when the base changes and therefore it can be seen that

$$E\{\mathrm{tr}(\hat{\mathbf{R}}_K-\mathbf{R}_K)^2\}=\sum_{ij}E\{\|\psi_{ij}\|^2\}=M^2\frac{\sigma^4}{N} \tag{8}$$

with

$$M^2\frac{\sigma^4}{N}=\sum_{i=1}^{M}(\lambda_i-\sigma^2)^2. \tag{9}$$

The profile of the ordered eigenvalues, as shown in Fig. 3, may then be approximated using the decreasing model

$$\lambda_i=\lambda_1 r^{i-1}. \tag{10}$$

Therefore the combination of Eqs. (6) and (10) results in

$$M\sigma^2=\sum_{i=1}^{M}\lambda_1 r^{i-1},$$

$$\lambda_1=M\left(\frac{1-r}{1-r^M}\right)\sigma^2=MJ\sigma^2, \tag{11}$$

where

$$J=\frac{1-r}{1-r^M}.$$

It can now be seen that

$$(\lambda_i-\sigma^2)=(MJr^{i-1}-1)\sigma^2, \tag{12}$$

which, combined with Eq. (9), results in

$$\frac{M+N}{MN}=\frac{(1-r)(1+r^M)}{(1-r^M)(1+r)}. \tag{13}$$

By substituting $r=e^{-2a}$, this reduces to

$$\frac{M\tanh(a)-\tanh(a)}{M\tanh(a)}=\frac{1}{N}, \tag{14}$$

where tanh is the hyperbolic tangent function.

An order 4 expansion of Eq. (14) produces the following biquadratic equation:

$$\sigma^4-\frac{15}{M^2+2}a^2+\frac{45M}{N(M^2+1)(M^2+2)}=0. \tag{15}$$

Only the positive solution of Eq. (15) is of interest, and this is given by

$$a=\sqrt{\frac{15}{2(M^2+2)}\left[1-\sqrt{1-\frac{4M(M^2+2)}{5N(M^2-1)}}\right]}. \tag{16}$$

V. A RECURSIVE EXPONENTIAL FITTING TEST

In the presence of $d<M$ noncoherent signals the profile of the noise eigenvalues discussed in Sec. IV can now be applied to the smallest $Q=(M-d)$ eigenvalues. The ordered profile of the normalized eigenvalues of $\hat{\mathbf{R}}_K$ can then be compared with the theoretical profile of the ordered noise eigenvalues. The signal eigenvalues will cause a break in the distribution as seen in Fig. 4.

The number of pointlike scatterers or targets present is equal to the eigenvalue index where the observed profile breaks from the theoretical profile. This is shown in Fig. 4, which demonstrates a simulation of the normalized eigenvalue profile for an array of 10 half wavelength separated sensors, using 10 snapshots in the presence of 2 targets located at 0° and 10° from the array. The break in the profile can be determined by performing a recursive test on the eigenvalue of $\hat{\mathbf{R}}_K$.[27]

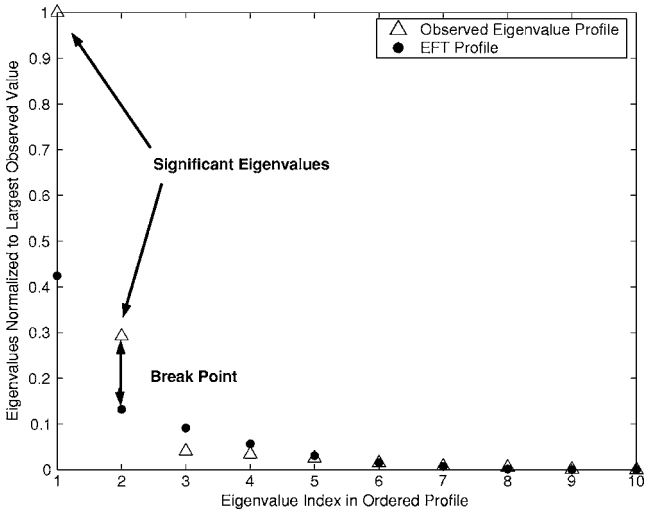


FIG. 4. Profile of ordered eigenvalues in simulations using two targets located at angles 0° and 10° from the array. The target eigenvalues cause a break in the predicted noise profile. The index of the eigenvalue where the break occurs is the number of significant eigenvalues, and therefore, the number of targets present assuming the targets are pointlike.

The initial estimation of the noise subspace dimension (P) is 1. The smallest P eigenvalues are assumed to be noise eigenvalues. The previous eigenvalue in the observed profile λ_{M-P}, is then tested to see if it corresponds to noise or if it corresponds to a scatterer. The outcome of the test is decided by using the previous P noise eigenvalues to predict the next value in the noise eigenvalue profile. Equation (10) is used to predict $\hat{\lambda}_{M-P}$, the next eigenvalue based the value for P retained at the previous step. Therefore r, and J will now be dependent on P, and will be written as r_P and J_P, respectively,

$$\hat{\lambda}_{M-P}=(P+1)J_{P+1}\hat{\sigma}^2, \tag{17}$$

with

$$J_{P+1}=\frac{1-r_{P+1}}{1-r_{P+1}^{P+1}}, \tag{18}$$

$$\hat{\sigma}^2=\frac{1}{P+1}\sum_{i=0}^{P}\lambda_i. \tag{19}$$

The combination of Eqs. (18) and (19) leads to the prediction equation

$$\hat{\lambda}_{M-P}=J_{P+1}\sum_{i=0}^{P}\lambda_{M-i}. \tag{20}$$

r_{P+1} is found by calculating a using Eq. (16), where $(P+1)$ is used instead of M and $r=e^{-2a}$.

Now, at step P the absolute difference between λ_{M-P} and $\hat{\lambda}_{M-P}$ is computed. This difference is then compared with a threshold value η_P, where the subscript P shows explicitly the dependence of the threshold on P. The following two hypotheses can now be defined as:

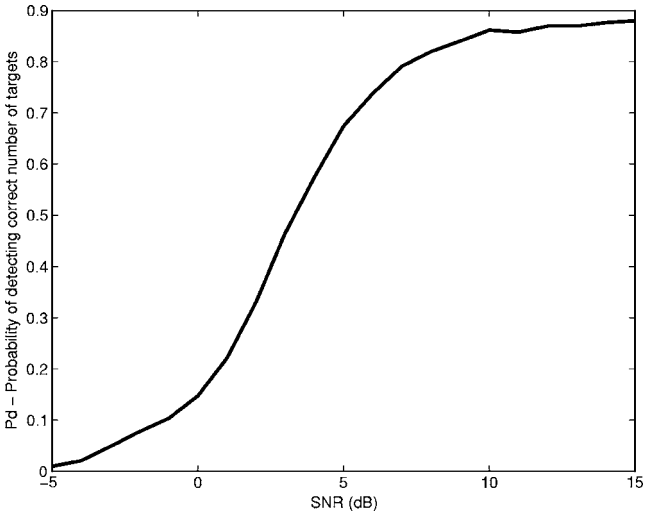


FIG. 5. Probability of correctly determining the number of targets or scatterers present using the EFT with increasing SNR. Simulations are run using number of snapshots $N=10$, the number of sensors $M=10$, and two targets located at angles 0° and 10°.

$$H_{P+1}:|\lambda_{M-P}-\hat{\lambda}_{M-P}| \leqslant \eta_P, \tag{21}$$

$$\bar{H}_{P+1}:|\lambda_{M-P}-\hat{\lambda}_{M-P}| > \eta_P. \tag{22}$$

If the error is less than η_P, the eigenvalue is classified as a noise eigenvalue and hypothesis H_{P+1} is chosen. If the error is greater than η_P, the eigenvalue is considered to be a target eigenvalue and therefore hypothesis $\bar{H}_{P+1}$ is selected. The estimated dimension of the noise subspace $\hat{P}$, is therefore the value of P, for which H_{P+1} is chosen over $\bar{H}_{P+1}$. The number of scatterers is then given by $\hat{d}=M-\hat{P}$.

The empirical distribution of the noise-only eigenvalue profile can be used to select a suitable threshold, η_P for each step of the test. This is emphasized in Fig. 3, where the circles denote the mean value of each ordered eigenvalue empirically obtained over 10 000 realizations. Assuming a normal distribution for each ordered eigenvalue, this figure reports the width corresponding to a six standard deviation truncation. For clarity of presentation, the figure only shows a few realizations, linking the corresponding eigenvalues. The threshold for each step P is chosen as half the spreading of the corresponding eigenvalue in the noise-only distribution. It should be noted that these relative η_P are independent of the noise power contrary to the choice of the TBT threshold, which remains subjective because it is SNR dependent.

The performance of the EFT is evaluated by considering the scenario of 2 targets located at angles of 0° and 10° from the array broadside impinge on a uniform linear array with $M=10$ sensors. There are $N=10$ snapshots available, and the SNR increases from −5 dB to 15 dB. The probability of determining the correct number of scatterers has been computed over 1000 realizations and is plotted against SNR in Fig. 5. There is no sense in comparing these results with the TBT, as the threshold of the latter must be subjectively adapted for each SNR value. It is worth remarking that a threshold determined for a given case will result in the TBT failing for almost all other cases.

VI. CONCLUSION

Low SNR or small numbers of snapshots give rise to a noise eigenvalue profile that is no longer almost horizontal, thereby resulting in failure of the traditional tests used by the array signal processing community. For example, this results in the TBT no longer correctly determining the number of scatterers present.

This paper has proposed an exponential profile of the ordered noise eigenvalues, which is based on examining the first and second order moments of the eigenvalues. Using this profile, a recursive test that has the advantage of being fully automatic, has been presented. It has been shown that while the proposed test is based on an assumption of white Gaussian noise, results reported by several groups indicate that this exponential profile is also observed experimentally. One of the important advantages of the proposed test is that it continues to work when the number of snapshots is small, unlike traditional model order selection tests which are only accurate asymptotically.

ACKNOWLEDGMENTS

This work is part funded by an IRCSET (Irish Research Council for Science Engineering and Technology) postgraduate scholarship.

[1]M. Fink, "Time reversal of ultrasonic fields. Part I: Basic theory," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[2]C. Dorme and M. Fink, "Ultrasonic beam steering through inhomogeneous layers with a time reversal mirror," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 167–175 (1996).
[3]G. F. Edelmann, T. Akal, W. S. Hodgkiss, S. Kim, W. A. Kupermann, and H. C. Song, "An initial demonstration of underwater acoustic communication using time reversal," IEEE J. Ocean. Eng. **27**, 602–609 (2002).
[4]S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach," J. Acoust. Soc. Am. **113**, 1533–1543 (2003).
[5]F. Wu, J. L. Thomas, and M. Fink, "Time reversal of ultrasonic fields—Part II: Experimental results," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 567–578 (1992).
[6]N. Mordant, C. Prada, and M. Fink, "Highly resolved detection and selective focusing in a waveguide using the D.O.R.T. method," J. Acoust. Soc. Am. **105**, 2634–2642 (1999).
[7]E. Kerbert, C. Prada, D. Cassereau, and M. Fink, "Ultrasonic nondestructive testing of scattering media using the decomposition of the time reversal operator," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1103–1113 (2002).
[8]C. Prada, M. Tanter, and M. Fink, "Flaw detection in solid with the D.O.R.T. method," in *Proceedings of the IEEE Ultrasonics Symposium*, 1997, pp. 679–683.
[9]H. C. Song, S. Kim, W. S. Hodgkiss, and W. A. Kuperman, "Environmentally adaptive reverberation nulling using a time reversal mirror," J. Acoust. Soc. Am. **116**, 762–768 (2004).
[10]N. Chakroun, M. Fink, and F. Wu, "Time reversal processing in ultrasonic nondestructive testing," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 1087–1098 (1995).
[11]C. Prada and J. L. Thomas, "Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix," J. Acoust. Soc. Am. **114**, 235–243 (2003).
[12]E. Kerbert, C. Prada, D. Cassereau, and M. Fink, "Imaging in the presence of grain noise using the decomposition of the time reversal operator," J. Acoust. Soc. Am. **113**, 1230–1240 (2003).
[13]T. Folegot, C. Prada, and M. Fink, "3D spatial resolution enhancement through environmental effects with the time reversal operator decomposition," *Oceans 2002 MTS/IEEE*, 2004, Vol. **1**, pp. 65–71.
[14]T. W. Anderson, "Asymptotic theory for principal component analysis," Ann. Math. Stat. **34**, 122–138 (1963).
[15]W. Chen, K. M. Wong, and J. Reilly, "Detection of the number of signals: A predicted eigen-threshold approach," IEEE Trans. Signal Process. **39**,

1088–1098 (1991).
[16]S. Valaee and P. Kabal, "An information theoretic approach to source enumeration in array signal processing," IEEE Trans. Signal Process. **52**, 1190–1196 (2004).
[17]H. Akaike, "A new look at the statistical model identification," IEEE Trans. Autom. Control **19**, 1361–1373 (1974).
[18]J. Rissanen, "Modeling by shortest data description length," Automatica **14**, 465–471 (1978).
[19]M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," IEEE Trans. Acoust., Speech, Signal Process. **33**, 387–392 (1985).
[20]M. Wax and I. Ziskind, "Detection of the number of coherent signals by the MDL principle," IEEE Trans. Acoust., Speech, Signal Process. **37**, 1190–1196 (1989).
[21]P. M. Djurić, "Model selection based on asymptotic Bayes theory," in *Proceedings of the Seventh SP Workshop on Statistical Signal and Array Processing*, 1994, pp. 387–392.
[22]"A model selection rule for sinusoïds in white Gaussian noise," IEEE Trans. Signal Process. **44**, 1744–1751 (1996).
[23]H. L. V. Trees, *Optimum Array Processing, ser. Detection, Estimation and Modulation Theory* (Wiley-Interscience, New York, 2002), Vol. **4**.
[24]A. Di, "Multiple source location—a matrix decomposition approach," IEEE Trans. Acoust., Speech, Signal Process. **33**, 1086–1091 (1985).
[25]A. P. Liavas and P. A. Regalia, "On the behavior of information theoretic criteria for model order selection," IEEE Trans. Signal Process. **49**, 1689–1695 (2001).
[26]A. P. Liavas, P. A. Regalia, and J. P. Delmas, "Blind channel approximation: effective channel order determination," IEEE Trans. Signal Process. **47**, 3336–3344 (1999).
[27]J. Grouffaud, P. Larzabal, and H. Clergeot, "Some properties of ordered eigenvalues of a Wishart matrix: Application in detection test and model order selection," in *Proceedings of ICASSP*, 1996, pp. 2463–2466.

Resonant modes in transiently evoked otoacoustic emissions and asymmetries between left and right ear

W. Wiktor Jedrzejczak and Katarzyna J. Blinowska
Department of Biomedical Physics, Institute of Experimental Physics, Warsaw University, Hoza 69 st., 00-681 Warszawa, Poland

Wieslaw Konopka
Department of Otolaryngology, Medical University, Zeromskiego 113 st., 90-549 Lodz, Poland

(Received 14 September 2005; revised 27 January 2006; accepted 28 January 2006)

A number of single-frequency resonant modes in click evoked otoacoustic emissions (OAEs) was investigated. The OAE modes were identified by means of an adaptive approximation method based on the matching pursuit (MP) algorithm. The signals were decomposed into basic waveforms coming from a very large and redundant dictionary of Gabor functions. The study was performed on transiently evoked otoacoustic emissions (TEOAEs) from left and right ears of 108 subjects. The correspondence between waveforms found by the procedure and resonant modes was shown (both for simulated noisy data and for single-person TEOAEs). The decomposition of TEOAEs made distinction between short and long-lasting components possible. The number of main resonant modes was studied by means of different criteria and they all led to similar results, indicating that the main features of the signal are explained on average by 10 waveforms. The same number of resonant modes for the right ear accounted for more energy than for the left ear. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2178718]

PACS number(s): 43.64.Jb, 43.60.Hj [BLM] Pages: 2226–2231

I. INTRODUCTION

It has been pointed out by several investigators (e.g., Elberling *et al.*, 1985; Probst *et al.*, 1991) that evoked otoacoustic emissions (OAEs) spectra are characterized by the occurrence of characteristic peaks, which appear at the same frequencies after stimulation in different frequency bands. The frequencies of these repeating spectral peaks do not correspond exactly to the frequency of stimulation in that usually the one with the highest energy is the closest to the stimulation frequency. Also, in OAEs evoked by broadband stimuli the spectral components present in burst-evoked OAEs can be identified. By means of the method of adaptive approximations it was found that not only the frequencies of these preferred components are practically the same, but their latencies are as well (Jedrzejczak *et al.*, 2004). This evidence indicates the presence of resonance modes in OAEs, which can be associated with the sharply tuned emission generators situated in specific areas of the organ of Corti as suggested by Wit *et al.* (1981).

This report addresses the question of how many resonant modes can be distinguished in transiently evoked otoacoustic emissions (TEOAEs). Toward this end, the method of adaptive approximations by matching pursuit (MP) introduced in Jedrzejczak *et al.* (2004) will be applied to the analysis of OAEs. This method is the perfect tool to study resonant modes, since it provides high-resolution information on frequencies, energies, and latencies of signal components, which correspond to the resonant modes. The method allows for determination of the time span of the components, which together with high resolution in the frequency dimension creates the possibility of distinguishing long-lasting components in TEOAEs.

The problem of possible asymmetries between TEOAEs in the right and the left ear also will be addressed. Statistically significant differences in the response level and signal-to-noise ratio (SNR) between TEOAEs from right and left ears were found for school children (Driscoll *et al.*, 2002). Also, for newborns the amplitude for the right ear was reported to be greater than for the left ear (Aidan *et al.*, 1997; Newmark *et al.*, 1997). In a maximum length sequence (MLS) study of OAEs (technique relying on application of quasirandom trains of clicks interspersed with silences of various durations), a higher level of OAE response was found in the right ears of females (Ismail and Thornton, 2003). And, recently Sininger and Cone-Wesson (2004) reported, for a large pool of data obtained from infants, that the average TEOAE SNR generated by clicks was larger when elicited in right ears, and the average distortion product otoacoustic emission (DPOAE) SNR generated by tones was larger for left ears. These results were associated with the fact that the left hemisphere (connected mainly to the right ear) takes precedence over the right in the processing of speech-like sounds (e.g., clicks), whereas the right hemisphere (connected to the left ear) plays a primary role in processing tonal stimuli and music. These findings were interpreted as a tendency of the cochlea to provide greater amplification to stimuli that will be preferably processed in the auditory areas of the contralateral hemisphere.

In the present study a link is established between the resonant modes and the signal components identified by the matching pursuit algorithm and the percentage of energy accounted for by the given number of resonant modes ana-

 0001-4966/2006/119(4)/2226/6/$22.50

lyzed. Possible asymmetries between right and left ear are also identified by considering resonant modes in both ears for the group of healthy adult males.

II. MATERIALS AND METHODS

TEOAEs from left and right ears of 108 young males (aged 18–25 years) were measured. Subjects were healthy without any otoscopic ear abnormalities. Impedance audiometry tests for all subjects were performed and revealed normal Type A tympanograms and normal acoustic reflexes. A hearing threshold of equal to or less than 10–15 dB HL was required and confirmed by pure-tone audiometry at all octave intervals between 125 and 8000 Hz. OAEs were recorded using the ILO292 Echoport system (version 5) designed by Otodynamics Ltd. Responses to 260 repetitions of the standard click stimulus were averaged using the "nonlinear" mode of stimulation. The acquisition window had a standard onset at 2.5 ms with a cosine rise/fall of 2.26 ms and flat top up to 20.5 ms. Stimulus intensity was kept at 77–80 dB SPL.

A. Matching pursuit

The method of adaptive approximations is based on the decomposition of the signal into basic waveforms (called atoms) from a very large and redundant dictionary of functions. Finding an optimal approximation of the signal by selecting functions from a very large and redundant set is a computationally intractable problem. Therefore, suboptimal, iterative solutions were applied. In practice, Gabor functions (sine-modulated Gaussians) are used, since they provide the best time-frequency resolution. This method was described in detail in Jedrzejczak *et al.* (2004).

In the same paper (Jedrzejczak *et al.*, 2004) it was shown that Gabor functions approximate the OAE signal very well. Comparison with other time-frequency (t-f) methods demonstrated the superior resolution of the MP method. Figure 1 illustrates the main features of this approach. The TEOAE signal (Fig. 1—top) was decomposed into its basic components and five of the largest energy peaks are presented (Fig. 1—bottom). Their energy distribution in the t-f space is illustrated in the center panel of Fig 1. At the left, the amplitude spectrum constructed by summation of the atoms' energy in time is shown. One can see that the weaker components are hardly visible in the spectrum. However, in the case of the MP analysis, the spectrum need not be relied on, since the spectral components are uncovered by the procedure and parametrized. In Fig. 1 the frequencies of the components are marked as lines, and their heights are proportional to their energy. It is clear that from the spectra it is not possible to distinguish the components of close frequencies (e.g., third and fourth waveforms at the bottom of Fig. 1).

In the present research, a dictionary consisting of 10^6 Gabor functions was used, with these functions being described by the following parameters: frequency, latency, time span, amplitude, and phase. The MP decomposition was performed on buffers "a" and "b" of the ILO92 system, on the average of "a" and 'b,' and on "a" subtracted by "b" (which is commonly used as an estimate of noise in OAE measurements). The decomposition was performed until fitted waveforms explained 99.5% of the signal energy. The parameters of main interest were amplitude, frequency, latency, and time span. The amplitude was determined as the maximum of the modulus of the waveform of a certain frequency, whereas the latency was measured from the stimulus onset to the maximum of the envelope of the waveform. The time-span parameter was defined as half the width of the TEOAE envelope and can be understood as the atom's duration in time.

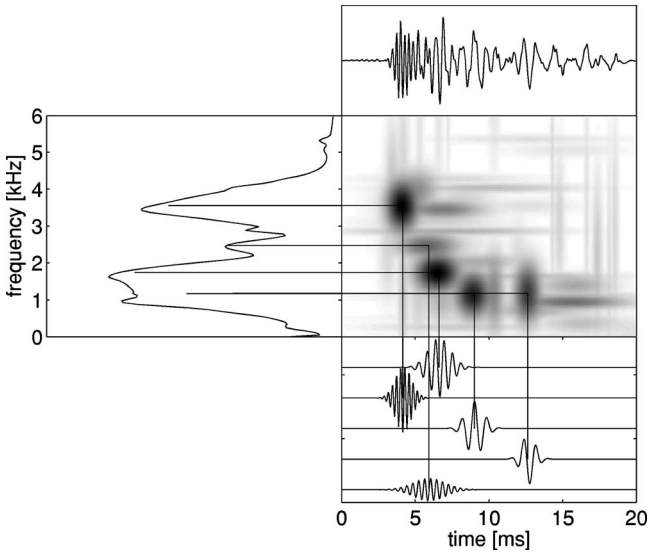


FIG. 1. Illustration of matching pursuit (MP) decomposition. Upper plot: OAE signal. Center plot: energy density (illustrated by shades of gray) in the time-frequency plane. Left: the spectrum. Bottom: waveforms fitted in the first five iterations, connected by lines with corresponding structures in the time-frequency plane and corresponding peaks in the spectrum. Please note that the third and fourth atoms are not resolved in the spectra.

B. Simulations

To demonstrate the performance of the MP procedure in finding resonant modes of TEOAEs, a test signal was constructed. This signal consisted of six gammatones (sine with gamma-function envelope) with frequencies 280, 550, 1100, 2200, 4400, and 8800 Hz, spaced 2 ms in time, and decaying at a constant 300 Hz, which gave descending slopes similar to those obtained experimentally. The spectra were calculated by means of fast Fourier transform (FFT) and MP for this simulated signal, and for the same signal with added noise as shown in Fig. 2. The spectral peaks in the case of MP are slightly broader, but in order to identify the components spectra are not needed, as the frequencies and energies of components are obtained directly as parameters. In the case of the FFT analysis, the distinction of gammatones from noise components is barely measurable as seen in Fig. 2(b). The MP method makes it possible to extract the signal from noise by neglecting low-energy atoms and constructing a spectrum only from the limited number of atoms found in the iterative procedure. For example, in Fig. 2(e) the spectrum was constructed from 10 atoms (corresponding to 80% of the energy). The result is quite similar to the original noise-free signal.

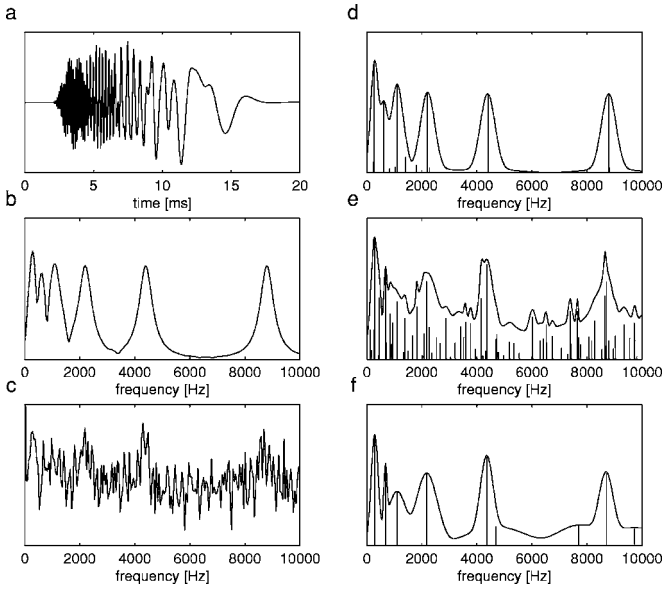


FIG. 2. Comparison of spectral properties of FFT (left) and MP (right) in application to the signal simulating an OAE. (a) Signal constructed from 6 gammatones of frequencies 280, 550, 1100, 2200, 4400, and 8800 Hz, spaced 2 ms in time and at a decay constant of 300 Hz; (b) FFT spectrum of signal (a); (c) FFT spectrum of signal (a) with added white noise (of variance twice the variance of the signal); (d) spectrum obtained by summation in time of atoms found by MP decomposition of signal (a); (e) spectrum obtained from MP decomposition of signal (a) with added noise; (f) spectrum obtained from 10 first atoms of MP decomposition performed on signal (a) with added noise. The spectral peaks in the case of MP analysis are broader, which follows from the Gaussian shape of atoms in time-frequency space. However, the resolution is better, since the frequencies of the components are determined by the MP procedure (they are marked by the vertical bars).

This simulation shows that even if the composition of the time series components is unknown, the first atoms of the MP decomposition represent the main features of the signal. The problem of how many TEOAE components are sufficient to describe the signal is considered below.

III. RESULTS

The simulation example shows that the noise level is important in the extraction of the basic components of the signal. Usually, in the recording procedure the two buffers a and b are stored and they provide information on the reproducibility of the signal and on the noise contribution. The level of noise can be estimated as the difference between buffers a and b. If this assumption is made, the number of atoms above a noise level can be calculated for each ear. The average number of frequency modes surpassing the noise level determined by the difference between a and b buffers was found as 9.7 ± 0.4.

One of the ways to identify resonant modes and extract them from the noise is to look for similar features in recording buffers a and b. Thus, each buffer was decomposed by the MP procedure, after which a search was undertaken for correlating components. That is, the correlation was computed between the atoms found for buffers a and b. The functions were qualified as the same modes when the correlation between them was above 0.75.

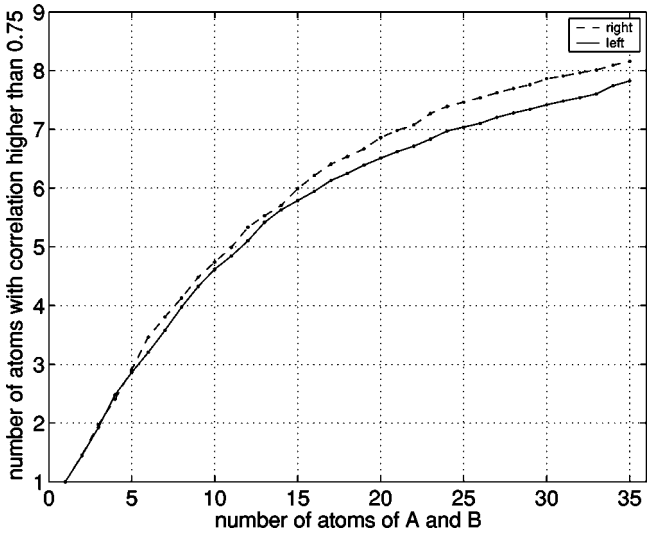


FIG. 3. Number of atoms from a and b averages with correlations higher than 75% plotted against the number of atoms of the whole decomposition for right (dashed line) and left (solid line) ears.

Figure 3 presents the number of the same modes from both buffers as a function of the number of atoms of the decomposition under consideration. First, when increasing the number of atoms, a rapid increase in the pairs of modes was observed, but then the curve became more flat and the increase of pairs of identical modes found was very small when more than 25 atoms were analyzed. For right ears the maximal number of resonant modes repeating in both a and b was 8.1 ± 0.4 and was slightly higher than in left ears at 7.8 ± 0.4. The number of resonant modes was also connected with the reproducibility parameter (correlation between a and b subaverages). For example, the higher the reproducibility, the more resonant modes were found in an ear. The high quality of measurement definitely allows for identification of a large number of frequency modes. Any disturbances in hearing caused a decrease in the values of parameters like reproducibility or response level (e.g., Lucertini *et al.*, 2002). The number of resonant modes was also lower in this case (Jedrzejczak *et al.*, 2005).

An alternative way of identifying the number of TEOAE components is consideration of the percentage of energy accounted for when increasing the number of atoms. In Fig. 4 the percentage of energy accounted for by a given number of atoms is shown. It is clear that the curve saturates around 10 atoms. In this case the sum of both buffers a and b, which leads to decrease of noise, is illustrated. And, as a result a slightly higher number of modes is obtained than for buffers a and b considered separately. The results of our simulations, and the fact that the principal waveforms described by the MP procedure occur to a large degree in both buffers, permits the 10 strongest atoms of the MP decomposition to be appreciated as the characteristic resonant modes of TEOAE.

In the literature resonant TEOAEs were sometimes associated with the long-lasting (slow decay) components appearing as spectral amplitude maxima (e.g., Talmadge *et al.*, 1998). Sisto *et al.* (2001) also defined the TEOAE components as "long-lasting," if their spectral-line amplitude exceeded the local noise for at least 40 ms after the click ad-

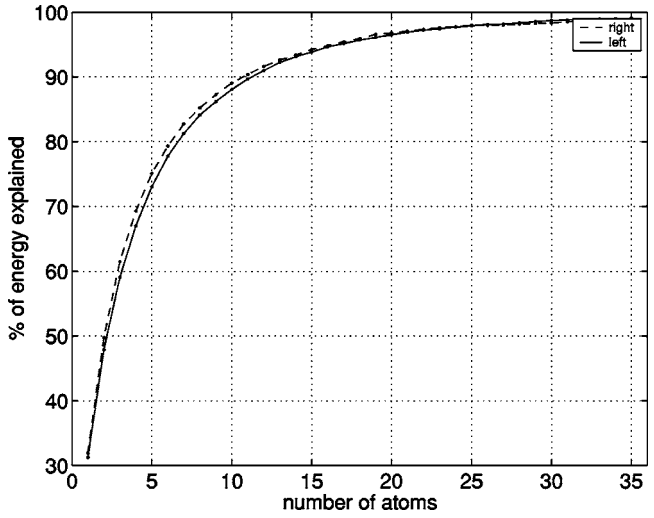


FIG. 4. The percentage of signal energy accounted for by the given number of waveforms (shown on horizontal axis) for right (dashed line) and left (solid line) ears. A significant difference between ears was found when the number of waveforms taken into account was between 3 and 7 (at level $p < 0.05$).

ministration, and 77% of these components were also identified as synchronized spontaneous otoacoustic emissions (SSOAEs).

Among resonance modes discovered here, there were also components of long duration and very narrow frequency span. Their characteristic shape in t-f maps made them easy to distinguish among the TEOAE components as illustrated in Fig. 5. Earlier, Jedrzejczak *et al.* (2004) showed that the histogram of the time spans of TEOAE atoms is bimodal with two peaks: the first at 2 ms, and the second at 14 ms, with a distinct minimum at 12 ms. This was also the case for the present data.

In this paper long-lasting components have been described as atoms longer than 12 ms. The definition of long-lasting components by Sisto *et al.* (2001) was different, which was connected with another experimental paradigm and signal-analysis method. Thus, the connection between long-lasting components defined here and SSOAEs cannot be established. However, it seemed relevant in the present analysis to take into account the contribution of long-lasting components. Among the 10 strongest atoms an average 0.38 ± 0.06 were connected with the long-lasting components. These were found to be among the 10 largest atoms in 32% of ears. The maximal number of long-lasting components in one ear was 5. When 35 atoms were taken into account, at least one long-lasting component was found in 81% of ears.

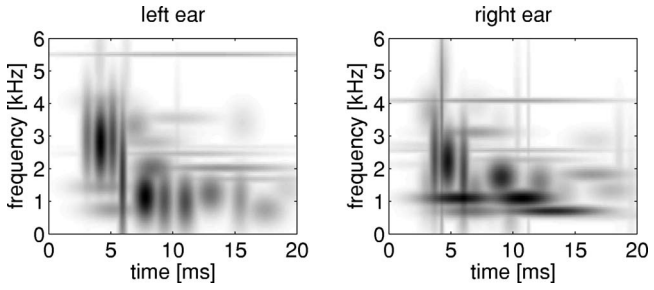


FIG. 5. Time-frequency distributions of energy density for TEOAEs from the right and left ears of the same subject. Please note distinct long-lasting components.

According to the general overview of the topic of ear asymmetries in TEOAEs, Robinette and Glattke (2002) concluded that the responses from the right ear were "more robust." In the present study the average energy of the TEOAE signals from both ears was considered as a function of t-f and revealed higher values for the right ear. However, these differences were not statistically significant. In addition, no correlations between TEOAE signals from the right and left ears of the same subject were found. When analyzing the t-f structures of responses for left and right ears of the subjects, different distributions of the frequency modes for each ear were observed for the same person. This finding is illustrated in Fig. 5, where TEOAE t-f maps for the right and left ears of the same subject are shown.

Next, the possibility of a functional dependence between a number of resonant modes in left/right ears for all subjects was investigated. Toward this end, the resonant modes were found in the same manner as in case of Fig. 3, namely by searching in both a and b buffers for waveforms with correlations higher than 0.75. The best fit was then achieved by a fitted linear function, indicating that more modes in one ear implies more modes in the second ear. However, the quality of the fit was not high ($R^2 = 0.33$), so it is difficult to draw conclusions concerning the relation between the number of resonant modes in both ears.

In search of specific differences between ears, also considered was the percentage of energy accounted for by the same number of resonant modes in the right and left ears. This dependence is shown in Fig. 4. Although both curves seem similar, the statistical significance of the differences between left and right ear was tested by means of the Wilcoxon rank sum test, which is an equivalent of the Student's t-test when analyzed populations do not have normal distributions. As a criterion of significance, a 95% confidence level ($p < 0.05$) was chosen. A significant difference between the TEOAEs of both ears was found when the number of waveforms taken into account was between 3 and 7 (at $p < 0.05$). To summarize, for TEOAEs for right ears, a given number of atoms contained a larger amount of energy as compared to the left ears.

Differences between ears were also examined by comparing contributions of long-lasting components in 10 principal modes. For the right ears the average number of long-lasting components was 0.4 ± 0.06 and 0.36 ± 0.06 in left ears, but this difference was not significant.

IV. DISCUSSION

In the present study, the resonant modes of TEOAE signal were represented by the first waveforms found by the adaptive approximations procedure. This assumption was supported by the following evidence:

(1) Simulations showing that the artificial TEOAE signal constructed from gamma-tones was well-reproduced by the atoms of the decomposition, even in the presence of noise;

(2) The simultaneous occurrence of the same waveforms (modes) in buffers a and b of the same person;
(3) The fact reported in the previous paper (Jedrzejczak *et al.*, 2004) that for a given ear, the same resonance modes identified as components of the decomposition occurred for different frequencies of stimulation and also in click-evoked OAEs.

The number of resonant modes in TEOAE is an individual feature of a subject and its identification depends also on the quality of the measurement—that is, the SNR. The present results indicate that the TEOAE signal is usually well-described by means of maximally 10 modes, and an average of 0.4 of them are long-lasting components.

Zimatore *et al.* (2002) found, using SVD (singular value decomposition), that four eigenvalues connected with the characteristic frequencies account for 90%–95% of TEOAE energy. However, the principal components described by them contained more than one frequency. Probst *et al.* (1991) stated that on average seven distinct frequencies are found in ears with SOAEs, and two or three of them were SOAEs. In ears without SOAEs, this number was decreased to four.

The higher number of resonant modes found here and the smaller contribution of long-lasting components (which might possibly be connected with SOAEs) to the principal modes than reported by Probst *et al.* (1991) may be explained by the different methods of analysis. In spectral analysis even important components placed on the slopes of the stronger spectral lines are not found, as illustrated in Fig. 2. On the other hand, spectral analysis favors long-lasting components of narrow frequency band, so even weaker components of this kind may be easily distinguished in the spectra. In the present study, the number of long-lasting components varied between subjects. That is, up to five long-lasting components in the first 10 atoms were identified in some data records, while for other subjects there were none.

Left-right asymmetry in TEOAEs with the prevalence of the right ear was found previously for newborns (e.g., Aidan *et al.*, 1997; Newmark *et al.*, 1997; Sininger and Cone-Wesson, 2004), children (Kei *et al.*, 1997; Driscoll *et al.*, 2002), and females (Ismail and Thornton, 2003). For adult males there was no clear evidence concerning left-right asymmetries in TEOAEs. In the investigation by Ismail and Thornton (2003) the response amplitude was found to be higher in the right ear, but this result was not statistically significant, which was also the case in the present study.

The significant differences concerning ear asymmetries in males were found here when considering resonant modes of TEOAEs for the left versus right ear. That is, the same number of resonant modes for the right ear accounted for more energy than in the left ear. The larger fraction of energy contained in smaller number of modes corresponds to higher SNRs. This result is compatible with the above-mentioned references concerning children and females reporting prevalence of high SNRs in the right ear.

Some authors connected the greater sensitivity in the right ears of females and children with the abundance of SOAEs found in females and in the right ears of both adults and full-term neonates (e.g., Morlet *et al.*, 1995). In Penner *et al.* (1993) the prevalence of multiple SOAEs in the right ears of females and males was reported. However, the study of Collet *et al.* (1993) revealed no correlation between SOAE prevalence and ear-side effects in adults. Sisto *et al.* (2001) reported that for males the fraction of ears with at least one measurable SSOAE was slightly higher for right ears, but this difference was not statistically significant. The number of long-lasting modes illustrated here in right ears was also slightly greater, but also not significant.

It seems that the role of SOAEs in right-left asymmetries and their contribution to the TEOAE merits further study. The application of high-resolution t-f analysis to TEOAE, SSOAE, and SOAE may bring the significant results. However, standard ILO-acquisition protocols do not allow such flexibility. If the correspondence between long-lasting atoms and SOAEs was established, the identification of SOAEs in the TEOAEs without the need of separate measurement would be possible by means of the MP algorithm, which would be beneficial for simplification of the measuring procedures.

V. CONCLUSIONS

The TEOAE signals can be considered to a large degree as a superposition of the resonant modes which are characteristic for each subject and for each ear. The identification of dominant frequencies is difficult for 20%–30% of ears according to Probst *et al.* (1991). The method of the MP analysis proposed here provides possible way to resolve this problem, which may also be important from the clinical point of view. Resonant modes can be connected with the sharply tuned emission generators situated in the cochlea (Wit *et al.*, 1981). The lack of resonance modes in a particular frequency band or their changed characteristics may provide more specific information on cochlear deficiencies than the amplitude of the OAE signal. For example, such information may be associated with distortion in a specific place in the cochlea, or as a particular kind of hearing disturbance. Among the resonant modes, the short and long-lasting components can be distinguished. Their simultaneous observation may bring additional valuable information about cochlear functionality, especially if the origin of the long-lasting modes was unequivocally established.

The existence of resonant modes of specific frequencies, latencies, and time spans gives support for oscillatory theories of OAE generation (e.g., Sisto and Moleti, 1999; Sisto *et al.*, 2001; Bell and Fletcher, 2004; Talmadge *et al.*, 1991). In particular, theories taking into account the morphological features of hair-cell arrangement in the cochlea and the presence of resonant cavities (Bell and Fletcher, 2004) seem relevant with respect to the generation of resonant modes. The problem of the modeling of mechanisms of OAE generation and auditory processing requires further studies based on an approach that considers not only spectral properties, but also the full t-f characteristics of OAE components.

ACKNOWLEDGMENTS

We would like to thank the two anonymous reviewers for their comments, which helped to make this a better paper.

This work was partly supported by a grant of the Polish Ministry of Scientific Research and Information Technology (No. 3 T11E 003 28).

Aidan, D., Lestang, P., Avan, P., and Bonfils, P. (**1997**). "Characteristics of transient-evoked otoacoustic emissions (TEOEs) in neonates," Acta Oto-Laryngol.. **117**, 25–30.

Bell, A., and Fletcher, N. H. (**2004**). "The cochlear amplifier as a standing wave: Squirting waves between rows of outer hair cells?," J. Acoust. Soc. Am. **116**, 1016–1024.

Collet, L., Gartner, M., Veuillet, E., Moulin, A., and Morgon, A. (**1993**). "Evoked and spontaneous otoacoustic emissions: A comparison of neonates and adults," Brain Dev. **15**, 249–252.

Driscoll, C., Kei, J., and McPherson, B. (**2002**). "Handedness effects on transient evoked otoacoustic emissions in schoolchildren," J. Am. Acad. Audiol **13**, 403–406.

Elberling, C., Parbo, N. J., Johnsen, N. J., and Bagi, P. (**1985**). "Evoked acoustic emissions: Clinical application," Acta Oto-Laryngol., Suppl. **421**, 77–85.

Ismail, H., and Thornton, A. R. (**2003**). "The interaction between ear and sex differences and stimulus rate," Hear. Res. **179**, 97–103.

Jedrzejczak, W. W., Blinowska, K. J., and Konopka, W. (**2005**). "Time-frequency analysis of transiently evoked otoacoustic emissions of subjects exposed to noise," Hear. Res. **205**, 249–255.

Jedrzejczak, W. W., Blinowska, K. J., Konopka, W., Grzanka, A., and Durka, P. J. (**2004**). "Identification of otoacoustic emission components by means of adaptive approximations," J. Acoust. Soc. Am. **115**, 2148–2158.

Kei, J., McPherson, B., Smyth, V., Latham, S., and Loscher, J. (**1997**). "Transient evoked otoacoustic emissions in infants: Effects of gender, ear asymmetry and activity status," Audiology **36**, 61–71.

Lucertini, M., Moleti, A., and Sisto, R. (**2002**). "On the detection of early cochlear damage by otoacoustic emission analysis," J. Acoust. Soc. Am. **111**, 928–978.

Morlet, T., Lapillonne, A., and Ferber, C. (**1995**). "Spontaneous otoacoustic emissions in preterm neonates: Prevalence and gender effect," Hear. Res. **90**, 44–54.

Newmark, M., Merlob, P., Bresloff, I., Olsha, M., and Attias, J. (**1997**). "Click evoked otoacoustic emissions: Inter-aural and gender differences in newborns," J. Basic Clin. Physiol. Pharmacol. **8**, 133–139.

Penner, M. J., Glotzbach, L., and Huang, T. (**1993**). "Spontaneous otoacoustic emissions: Measurement and data," Hear. Res. **68**, 229–237.

Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). "A review of otoacoustic emissions," J. Acoust. Soc. Am. **89**, 2027–2067.

Robinette, M. S., and Glattke, T. J. (**2002**). *Otoacoustic Emissions Clinical Applications*, 2nd ed. (Thieme, New York), Part II, Chap. 4, pp. 106–108.

Sininger, Y. S., and Cone-Wesson, B. (**2004**). "Asymmetric cochlear processing mimics hemispheric specialization," Science **305**, 1581.

Sisto, R., and Moleti, A. (**1999**). "Modelling otoacoustic emissions by active nonlinear oscillators," J. Acoust. Soc. Am. **106**, 1893–1906.

Sisto, R., Moleti, A., and Lucertini, M. (**2001**). "Spontaneous otoacoustic emissions and relaxation dynamics of long decay time OAEs in audiometrically normal and impaired subjects," J. Acoust. Soc. Am. **109**, 638–647.

Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). "Modeling otoacoustic emission and hearing threshold fine structures," J. Acoust. Soc. Am. **104**, 1517–1543.

Talmadge, C. L., Tubis, A., Wit, H. P., and Long, G. R. (**1991**). "Are spontaneous otoacoustic emissions generated by self-sustained cochlear oscillators?," J. Acoust. Soc. Am. **89**, 2391–2399.

Wit, H. P., Langevoort, J. C., and Ritsma, R. J. (**1981**). "Frequency spectra of cochlear acoustic emissions (Kemp-echoes)," J. Acoust. Soc. Am. **70**, 437–445.

Zimatore, G., Hatzopoulos, S., Giuliani, A., Martini, A., and Colosimo, A. (**2002**). "Comparison of transient otoacoustic emission responses from neonatal and adult ears," J. Appl. Physiol. **92**, 2521–2528.

Cochlear transducer operating point adaptation

Yuan Zou, Jiefu Zheng, and Tianying Ren
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098

Alfred Nuttall[a)]
Oregon Hearing Research Center, NRC04, Oregon Health & Science University, 3181 SW Sam Jackson Park Road, Portland, Oregon 97239-3098, Biomedical Engineering, Oregon Health & Science University, 20000 NW Walker Rd, Beaverton, Oregon 97006, and Kresge Hearing Research Institute, University of Michigan, 1301 E. Ann Street, Ann Arbor, Michigan 48109-0506

(Received 14 September 2005; revised 11 January 2006; accepted 17 January 2006)

The operating point (OP) of outer hair cell (OHC) mechanotransduction can be defined as any shift away from the center position on the transduction function. It is a dc offset that can be described by percentage of the maximum transduction current or as an equivalent dc pressure in the ear canal. The change of OP can be determined from the changes of the second and third harmonics of the cochlear microphonic (CM) following a calibration of its initial value. We found that the initial OP was dependent on sound level and cochlear sensitivity. From CM generated by a lower sound level at 74 dB SPL to avoid saturation and suppression of basal turn cochlear amplification, the OHC OP was at constant 57% of the maximum transduction current (an ear canal pressure of −0.1 Pa). To perturb the OP, a constant force was applied to the bony shell of the cochlea at the 18 kHz best frequency location using a blunt probe. The force applied over the scala tympani induced an OP change as if the organ of Corti moved toward the scala vestibuli (SV) direction. During an application of the constant force, the second harmonic of the CM partially recovered toward the initial level, which could be described by two time constants. Removing the force induced recovery of the second harmonic to its normal level described by a single time constant. The force applied over the SV caused an opposite result. These data indicate an active mechanism for OHC transduction OP. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2173517]

PACS number(s): 43.64.Kc, 43.64.Bt [BLM] Pages: 2232–2241

I. INTRODUCTION

The cochlear microphonic potential (CM) can be used to characterize mechano-electrical transduction (MET) in outer hair cells (OHCs) (Nieder and Nieder, 1971; Patuzzi *et al.*, 1989; Kirk *et al.*, 1997). The opening probability of MET channels in the stereocilia of hair cells follows a Boltzmann activation function (Holton and Hudspeth, 1986). The relationship between instantaneous pressure in the ear canal $[P(t)]$ and the current through the OHCs (I_{OHC}) follows a similar function. The instantaneous CM potential (V_{CM}) measured between the scala vestibuli (SV) and the round window, is proportional to the summed current through the outer hair cells and the effective resistance of the extra cellular fluid (R_{fluid}). We assume that R_{fluid} is constant. V_{CM} could be given by Eq. (1) (Patuzzi and Moleirinho, 1998; Patuzzi, 2002):

$$V_{\mathrm{CM}} = I_{\mathrm{OHC}} R_{\mathrm{fluid}} \propto V_{\mathrm{sat}} - \frac{2V_{\mathrm{sat}}}{1+\exp[Z(P(t)+P_0)/KT)}, \quad (1)$$

where V_{sat} is the maximal voltage change from OHC current induced by the large deflection of OHC hair bundles; Z is a sensitivity parameter in units of electron volts per Pascal (eV/Pa); K is Boltzmann's constant; T is temperature; $P(t)$ is a sinusoid sound pressure in the ear canal; P_o is an offset parameter to account for the fact that OHC stereocilliary displacement may not be centered for a zero ear canal pressure stimulus, which is used as the indicator of the operating point (OP). Parameters V_{sat}, Z, and P_o are shown in Fig. 1(a). In order to make the OP have a more clear biological meaning, here we also define OP as the percentage of the maximum transduction current (MTC) again assuming a proportional relationship of V_{CM} to I_{OHC}. It can be calculated by Eq. (2),

$$\mathrm{OP} = \frac{\dfrac{2V_{\mathrm{sat}}}{1+\exp(ZP_0/KT)}}{2V_{\mathrm{sat}}} \times 100\,\% = \frac{1}{1+\exp(ZP_0/KT)} \times 100\,\%\,. \quad (2)$$

Waveform symmetry for V_{CM} occurs when P_0 is 0; then OP is 50% of MTC (i.e., 50% of the MET channels open). When P_0 has a positive value, OP will be <50% MTC, which is equivalent to the bias of the stereocilia bundle toward channel closure, as if the entire organ of Corti was biased toward scala tympani (ST). Likewise, when P_0 has a negative value, OP will be >50% MTC, the stereocilia bundle bias is toward the channel's open position, as if the entire organ of Corti were biased toward SV.

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

 0001-4966/2006/119(4)/2232/10/$22.50

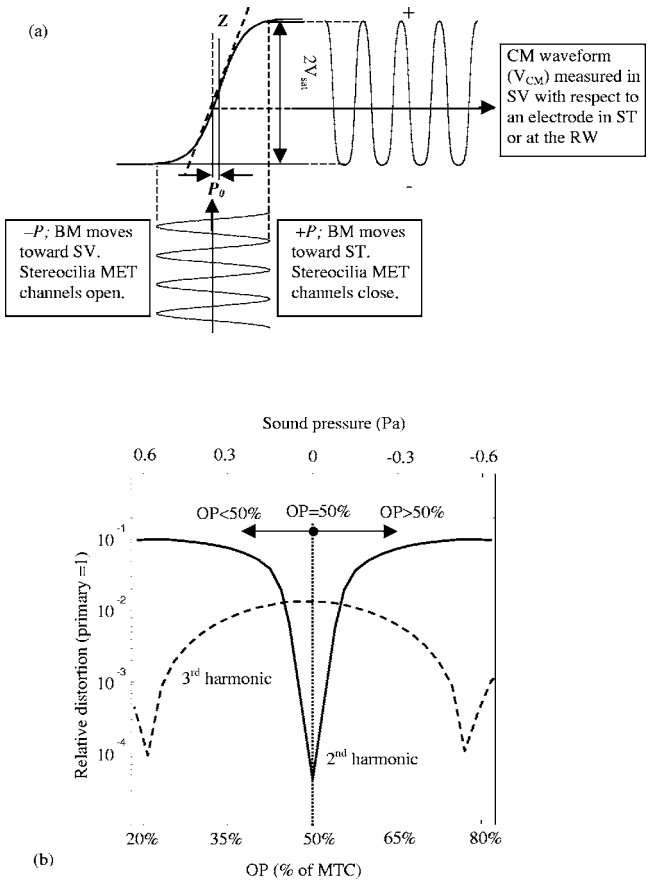


FIG. 1. (a) Graphical illustration of the Boltzmann function parameters; (b) the second and third derivatives of the Boltzmann function.

The Boltzmann parameter estimation technique is accomplished using a low-frequency, high-level sound to saturate the OHC transduction current (Patuzzi and Moleirinho, 1998). The transduction function of basal turn cochlear OHCs can be obtained because, in practice, a high sound level given at low frequency will phase synchronize their responses and generally will not damage the basal turn cochlear OHCs. Travelling wave mechanics strongly attenuates their stimulation. For example, cochlear sensitivity could quickly and totally recover immediately after terminating a low-frequency high-level sound (such as 200 Hz at 110 dB SPL) (Patuzzi and Moleirinho, 1998; Sirjani *et al.*, 2004). However, stimulation with a high-level sound (>90 dB SPL) will cause basal turn cochlear amplification to be suppressed. It has been noted that the OHC OP changes during and following the high level sound stimulation (Frank and Kössl, 1996; Patuzzi and Moleirinho, 1998). To determine the "normal" OHC OP and its stability, we use lower sound stimulation levels.

Boltzmann parameter estimation does not work well without CM waveform saturation. The OP changes for lower sound level probe tones can be analyzed based on the second and the third harmonics of CM (Frank and Kössl, 1996; Kirk and Patuzzi, 1997; Bian *et al.*, 2002; Bian, 2004; Sirjani *et al.*, 2004; Lukashkin and Russell, 2005). The second and the third harmonics of CM are proportional to the second and the third derivatives of the Boltzmann function, respectively [Fig. 1(b)]. The second harmonic distortion is extremely sensitive to small operating point changes and rapidly increases as the OP moves away from the value of 50% of MTC ($P_0=0$). The third harmonic distortion is less sensitive to small operating point changes, but it decreases as OP moves away from the value of 50% MTC ($P_0=0$). By measuring CM distortion, we were able to determine OHC OP change induced by cochlear amplification suppression from high-frequency high-level sounds or from a low-frequency high-sound level probe stimulus. By applying a force to the cochlea using a probe we devised a way to rapidly perturb the OHC OP and measure its time-dependent change.

II. METHODS

A. Animal preparation

Thirty pigmented guinea pigs weighing 250–380 g were used in this study. The Oregon Health and Science University Committee on the Use and Care of Animals approved the experimental protocols. Anesthesia was achieved with ketamine (30 mg/kg, i.m.) and xylazine (5 mg/kg, i.m.) and maintained by half doses of both agents approximately every hour. The animal's head was firmly fixed in a head holder, which was electrically isolated and heated to prevent cochlear thermal cooling. A tracheotomy was performed and a ventilation tube was inserted into the trachea to ensure free breathing. The left ear was surgically prepared to expose the cochlea leaving the medial portion of the external auditory canal to facilitate placement of the acoustic speculum. The bulla was widely opened to expose the cochlea. Then the middle ear muscle tendons were carefully sectioned.

B. Cochlear microphonic (CM) recording and cochlear sensitivity measurement

The 400 Hz sound was generated by a loudspeaker (ER-2, Etymotic). The speaker output was coupled to the external ear canal using tubing that terminated in a plastic speculum. The CM waveform was recorded using a differential electrode pair of two wires. One recording electrode of the differential pair, a Teflon-insulated 2-T platinum-iridium wire (50 μm in diameter), was inserted into the SV at the 18 kHz BF place and sealed. The other electrode, a 3-T Ag wire (75 μm in diameter), was placed at the round window niche. The reference electrode was put at the neck of animal. The signal was amplified 1000 times by a differential preamplifier (Grass instrument Co. Model P15) and a custom-designed ac amplifier. The SV electrode was directed to the positive input of the differential amplifier. After A/D conversion and averaging, the CM waveforms were saved for post-processing.

In order to monitor the hearing sensitivity during the experiment, compound action potential (CAP) thresholds were recorded. Tone bursts (10 ms in duration, 1 ms rise/fall) from 2 to 30 kHz in 2 kHz increments were generated using a 250 kHz, 16 bit D/A converter (Tucker Davis Technologies) and delivered to the ear canal to evoke CAP. The amplified signal from the round window was also displayed on an oscilloscope for CAP threshold assessment.

C. The operating point (OP) position calculation

When giving a 400 Hz 94 dB SPL sound to saturate the OHC transduction current, the absolute OP can be calculated by a Boltzmann parameter estimation technique. By fitting the CM waveform with a Boltzmann function, P_0 is estimated by Eq. (1). And OP could be calculated by Eq. (2) (Patuzzi and Moleirinho, 1998; Patuzzi, 2002).

Boltzmann parameter estimation does not work well without CM waveform saturation. The OP changes for lower sound level probe tones is analyzed based on the second and the third harmonics change of CM (Frank and Kössl, 1996; Kirk and Patuzzi, 1997; Bian *et al.*, 2002; Bian, 2004; Sirjani *et al.*, 2004; Lukashkin and Russell, 2005). By measuring CM distortion, we were able to determine the OHC OP change direction.

We used a ratio in the analysis of CM harmonics, effectively calculating % harmonic distortion. We scaled the harmonics for two reasons. First, we believe the change of amplitude of CM during the force application is partially due to factors separate from any change in the OP, for example, due to the stiffness change of the organ of Corti. Second, the absolute magnitude of the CM and harmonics varied among the animals. The harmonic ratio showed less variation.

D. Basilar membrane velocity measurement

To observe the loud sound effects on the cochlear mechanical responses, the magnitude and phase transfer function of the basilar membrane (BM) transverse velocity was measured at the 18 kHz BF location. To record the responses of the BM, a small opening was made in the first-turn ST bony wall of the cochlea. Gold-coated glass beads (20 μm diameter) were "dropped" onto the BM (at approximately the 18 kHz place) to serve as reflective objects to track the motion of the BM. The laser beam from the Doppler velocimeter (Polytec, OFV 1102) was focused on a bead with the aid of a compound microscope (Nuttall *et al.*, 1991). BM velocities were determined after fast Fourier transform of the Hanning-windowed responses from the velocimeter. We used two guinea pigs for this test.

E. Cochlear perfusion to hydraulically bias the stereociliary bundle position

The perfusion of artificial perilymph (AP, containing in mM: NaCl 125, KCl 3.5, glucose 3.3, HEPES 2.8, $CaCl_2$ 1.3, $NaHCO_3$ 25, NaH_2PO_4 0.5, and $MgCl_2$ 1.1) was used to hydraulically bias the basilar membrane (BM) to the SV direction. This will induce a deflection of OHC stereocilia in the excitatory direction. We used eight guinea pigs for this test. A small hole was made in the cochlear bone at the ST at the 18–30 kHz BF place. A tube connected to a syringe filled with AP was inserted into this hole, then cemented and sealed in place. The AP was injected into the ST by the microsyringe pump at a constant rate of 4 or 16 μl/min. The cochlear apex was opened as the outlet for the AP to exit the cochlea.

F. Constant force application to the bony shell of the cochlea and measurement of cochlear bone deformation

A custom designed force transducer was used to apply a force on the bony shell of the cochlea [Fig. 8(b), later]. Two syringes whose diameters were matched well (one tube could move smoothly in the other tube) compressed an internal spring. A blunt probe at the end of one syringe was placed against the otic capsule. We used a constant force of 17 g, which was produced by the probe and was calibrated by weight. We used 23 guinea pigs for this test.

For the post-mortem experiment, five animals were killed by intracardial injection of 0.1 cm^3 of Euthasol. The data was collected after the animals ceased breathing for 15 min.

An *in vitro* experiment was done in order to determine the cochlear bone deformation caused by the force application. We used five guinea pigs for this test. The temporal bone from the guinea pig was excised and fixed to a tiltable platform. The bulla was opened wide, and the specimen was tilted to place the ST of basal turn horizontal. Force was applied to the bony shell of ST at the 18 kHz BF place. Gold-coated beads were placed on the cochlear shell around the transducer probe. The displacement of individual beads was measured by the interferometer (Polytec OVF 1102).

III. RESULTS

A. The initial operating point (OP) position and its change with sound stimulation level and cochlear sensitivity

Using a low-frequency high-sound stimulation level ($>$90 dB SPL), we could find the absolute OP position from the CM waveform fitted by a Boltzmann function in Eq. (2). When the stimulation level is lower (e.g., 74 dB SPL), as mentioned in Sec. II, the OP position change can be estimated by an analysis of the changes of the second and third harmonics. However, it is critical to know the initial OP position. We always first determined that initial OP position by the application of a brief 10 ms 400 Hz tone at 94 dB SPL, which saturates OHC mechanotransduction producing a distorted CM waveform. In Fig. 1(b), as the OP moves away from the value of 50% MTC (P_0=0), the second distortion increases rapidly, and the third harmonic distortion decreases slightly (Frank and Kössl, 1996; Kirk and Patuzzi, 1997; Bian *et al.*, 2002; Sirjani *et al.*, 2004; Lukashkin and Russell, 2005). When the initial OP has a value $>$50% MTC ($P_0 < 0$), it means that there is a positive bias of the stereocilia laterally as if the organ of Corti were displaced toward SV. With an OP movement further to a higher value, the second harmonic will further increase, and the third harmonic will further decrease. If the initial OP has a value $<$50% MTC ($P_0 > 0$), the stereocilia movement toward the SV will induce a decrease of the second harmonic and an increase of the third harmonic distortion.

Previous studies using a low-frequency high-sound stimulation level found that the value of initial OP (when tone was first applied) varied among the test subjects. The initial OP could be $>$50% MTC ($P_0 < 0$), $<$50% MTC (P_0

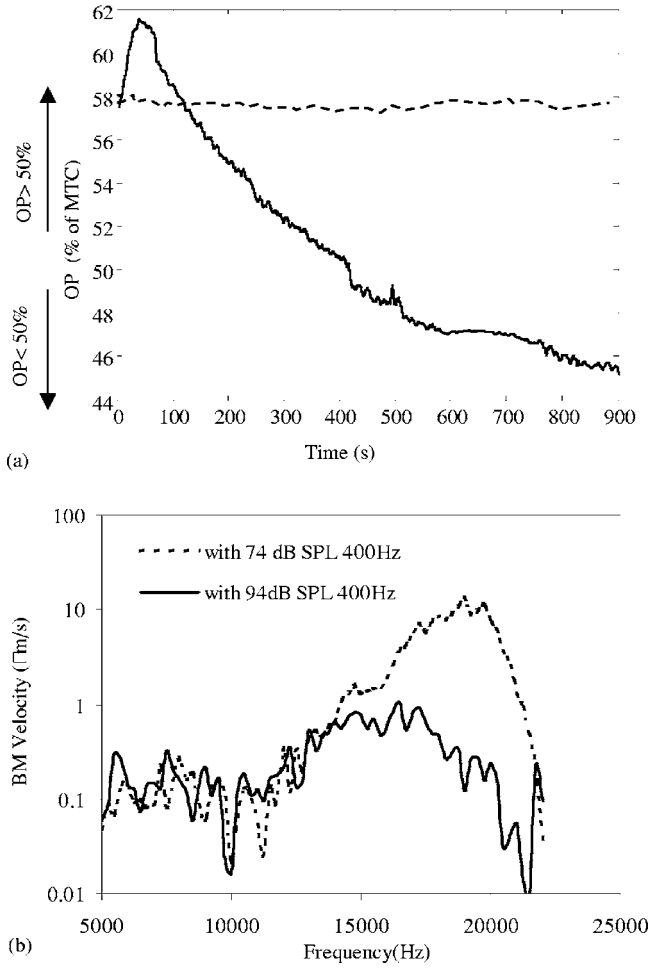


FIG. 2. (a) The derived OP from the Boltzmann fitting of the CM waveform during the first 15 min first application of 400 Hz 94 dB SPL sinusoidal acoustic stimulation. Following a brief increase, the OP moved to a value <50% MTC, consistent with a slow movement of the stereocilia toward channel closure as if the organ of Corti moved toward the ST. (b) The BM mechanical tuning curves measured at the 18 kHz BF place using a laser interferometer. The frequency of a 30 dB SPL tone was varied from 5 to 22 kHz in 200 Hz steps. While a continuous sound of 400 Hz 94 dB SPL (solid line) or 74 dB SPL tone (dotted line) was simultaneously presented. The 400 Hz 94 dB SPL tone suppressed the cochlear amplification. The 74 dB SPL probe sound did not change the tuning curve (the tuning without a low-frequency tone is similar to that shown by a dotted line; data not shown).

>0), and sometimes was very close to 50% MTC ($P_0=0$) (Frank and Kössl, 1996; Patuzzi and Moleirinho, 1998). The reasons for this are that the initial OP is a function of both the stimulation sound level and the cochlear sensitivity.

An example is shown in Fig. 2(a), where a 400 Hz 94 dB SPL sound was presented continuously for 900 s. The initial OP (solid line) was 57% MTC ($P_0=-0.1$ Pa), and underwent a bidirectional shift, ultimately moving to a value <50% MTC ($P_0>0$). However, for the stimulation with a continuous 400 Hz 74 dB SPL sound, the second and the third harmonics of CM did not change in time, indicating that the OP (dashed line) maintained the initial value.

The sound-induced OP changes are linked to basal turn cochlear amplification that becomes saturated by the high-level 400 Hz sound (94 dB SPL) as shown in Fig. 2(b). The BM mechanical tuning curve (BM velocity spectrum for constant ear cannel sound pressure) was measured from a reflective bead placed on the BM at a 18 kHz BF location simultaneously with the application of a probe stimulus of 400 Hz at 94 dB SPL (solid line) or at 74 dB SPL tone (dotted line). The probe stimulus at 94 dB SPL caused a reduction of tuning and a downward shift of the BF.

TABLE I. The relationship between the initial OP value and cochlear sensitivity. The sensitivity is assessed by 18 kHz CAP threshold.

	Sensitive ($n=15$)	20–40 dB loss ($n=10$)	Insensitive ($n=5$)
Initial OP	57% ±3%	48% ±4%	45% ±5%

In order to determine the influence of hearing sensitivity on OP, the CAP thresholds were measured in animals grouped by different sensitivity. Threshold sensitivity was determined for different frequencies (from 2 to 32 kHz) and recorded using differential electrodes located at the 18 kHz BF place. Cochlear sensitivity was categorized in three levels. The first group level was sensitive when CAP threshold (our laboratory the mean CAP threshold at frequency 18 kHz for the normal guinea pig is 27 dB SPL) had <20 dB of change. The second group had intermediate sensitivity, if the CAP had lost 20–40 dB (the 18 kHz CAP threshold is between 50 and 70 dB SPL). The third group level was insensitively defined when the CAP threshold had lost more than 40 dB for frequency $\geq$18 kHz. We observed that when the CAP sensitivity at 18 kHz frequency was reduced by 20–40 dB SPL (even if the CAPs for frequencies lower than 12 kHz did not change), the initial OP changed from 57% ($P_0=-0.1$ Pa) MTC to a value very close to 50% MTC ($P_0=0$). When the cochlear sensitivity was poor, the initial OP was <50% MTC ($P_0>0$) (Table I).

B. Control experiment: OP change with artificial perilymph perfusion

In this experiment, the CM waveform changes during artificial perilymph (AP) perfusion were recorded. AP perfusion through the ST of the first turn to exit at the apex of the cochlea was used to hydraulically bias the organ of Corti toward the SV direction. This will induce OHC stereocilia displacement in the excitatory direction increasing a MET channel current.

First, CM was elicited by 400 Hz 74 dB SPL sound. Then the OP shift in response to a hydrostatic bias was analyzed by measuring the change of the harmonics of the CM. Figure 3 shows an example of this result from a sensitive animal. The initial OP is >50% MTC ($P_0<0$). When perfusion started, it biased the stereocilia further toward SV, which induced an OP movement toward an increased value, increased the second harmonic [Fig. 3(c)], and decreased the third harmonic distortion [Fig. 3(d)]. Following the second harmonic increase at the start of perfusion, the second harmonic remained constant during the perfusion. Therefore the hydrostatic pressure, which would deflect the organ of Corti toward SV, and shifts OP to a greater percentage current value (a greater negative P_0). The hair bundle biased to the excitatory side should increase the transduction current. This

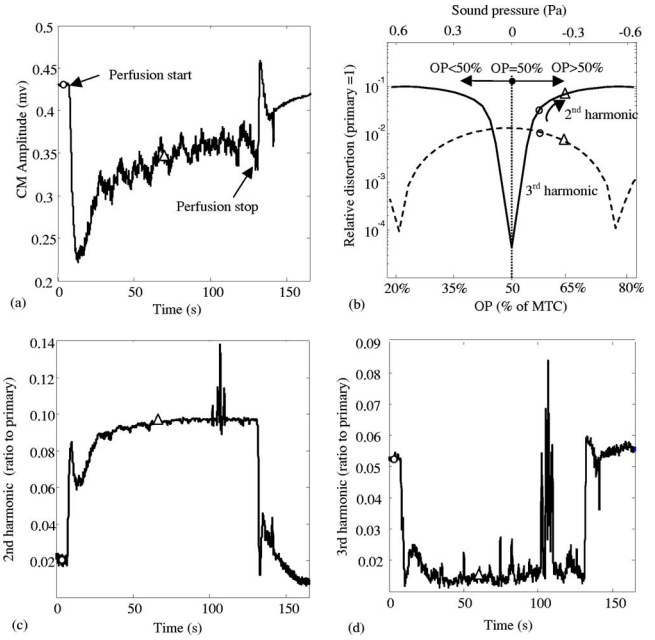


FIG. 3. ST-perfusion-induced changes in the CM (a), OP (b), the second (c) and third harmonics (d). The CM induced by a 74 dB SPL tone was recorded as a function of time before, during, and after ST perfusion. The data of all animals showed a similar result. One typical dataset (animal 05-27) was shown. The CM amplitude (mv) in panel (a), the second harmonic in panel (c), and the third harmonic in panel (d) during perfusion with AP at 4 μl/min are shown. Panel (b) shows a diagram of the OP position shift and the second and third harmonics changes predicted by the Boltzmann model before perfusion (◦) and during perfusion (△).

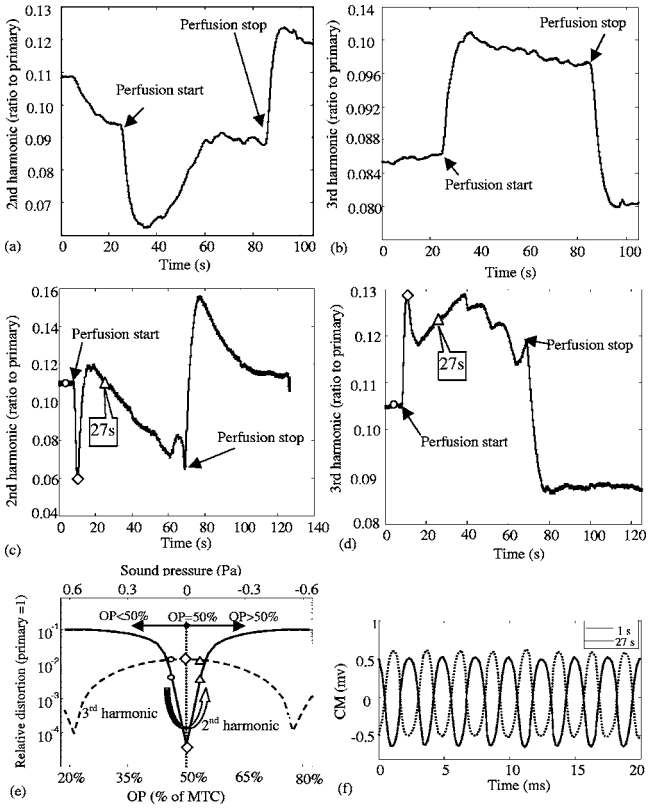


FIG. 4. The CM evoked by a 94 dB SPL sound changes over time with ST perfusion at 4 μl/min (left panel) or 16 μl/min (right panel). Fifteen minutes after applying the 400 Hz 94 dB SPL sound, the initial OP moved in the ST position and became stable (◦). When perfusion started at the rate of 4 μl/min, the stereocilia were biased toward SV, which induced a decrease of the second harmonic [panel (a)] and an increase of the third harmonic [panel (b)]. When the perfusion rate was 16 μl/min, the changes of the second and third harmonics are shown in panels (c) and (d), respectively. In Panel (e), the OP increased and passed through the value of 50% MCT (◇), to arrive at a value >50% MCT (△). The U-shaped arrow emphasized this dynamic movement of OP. The CM changed its asymmetrical shape after perfusion (27 s) compared with the waveform before perfusion (1 s) [panel (f)].

bundle bias, however, does not necessarily result in an increase of the CM magnitude due to the following reasons. The CM amplitude is related to the mechano-transduction gain, the amount of harmonic distortion, and the organ of Corti passive mechanical properties, such as the BM stiffness. The CM mechanotransduction gain (the slope of the curve) is the maximum at the 50% current position (p_0=0). Any OP change away from this position will decrease the gain. Upon stopping the AP perfusion, the second harmonic decreased [Fig. 3(c)], and the third harmonic distortion increased [Fig. 3(d)], restoring the original OP value.

To test the validity of our OP shift interpretation, we forced the OP to pass the value of 50% MTC (P_0=0) during perfusion. To accomplish this, the initial OP must be <50% MTC (P_0>0). Therefore, in this experiment, CM was elicited by 400 Hz 94 dB SPL sound. After 15 min of continuous stimulation by 400 Hz at 94 dB SPL, the initial OP moved and became stable in a value <50% MTC (P_0>0) [as illustrated by the example in Fig. 2(a)]. When perfusion started at a low rate (4 μl/min), the stereocilia biased toward SV. Subsequently, OP moved to a value close to the value of 50% MTC (P_0=0), and the CM waveform became more symmetric (data not shown), and the second harmonic decreased [Fig. 4(a)] while the third harmonic increased [Fig. 4(b)]. When the perfusion rate was 16 μl/min, the dynamic shift of OP was observed to pass through the value of 50% MTC (P_0=0), arriving in a value >50% MTC (P_0<0) [Figs. 4(c)–4(e)]. The second harmonic decreased to zero, and then increased [Figs. 4(c) and 4(e)]. The third harmonic increased to maximum and then decreased [Figs. 4(d) and 4(e)]. If the initial OP is <50% MTC (P_0>0), the effective position of the stereociliary bundles is biased toward inhibition or closure of the transduction channel direction. The higher sound stimulation level would produce a distorted CM waveform with clear saturation for inhibition displacement of the OHC stereocilia, i.e., for BM displacement toward the ST. If the initial OP is >50% MTC (P_0<0), the CM waveform should be more saturated for excitatory direction displacements, i.e., for BM displacement toward the SV. Thus, when the OP dynamically moved through the value of 50% MTC, the CM waveform changed its symmetry. When the time is 1 s, the initial OP was <50% MTC (P_0>0), when ST perfusion started (after 10 s), the OP will be hydraulically biased to a higher value. Therefore, when the bias is large enough, the OP will pass the value of 50% MTC (P_0=0) and arrive at a value >50% MTC (P_0<0). The waveform will be distorted and saturated in the opposite direction. This is what happened between 1 and 27 s [Fig. 4(f)]. The U-shape arrow in Fig. 4(e) illustrates this dynamic movement of OP.

During the perfusion, the change of CM amplitude, the

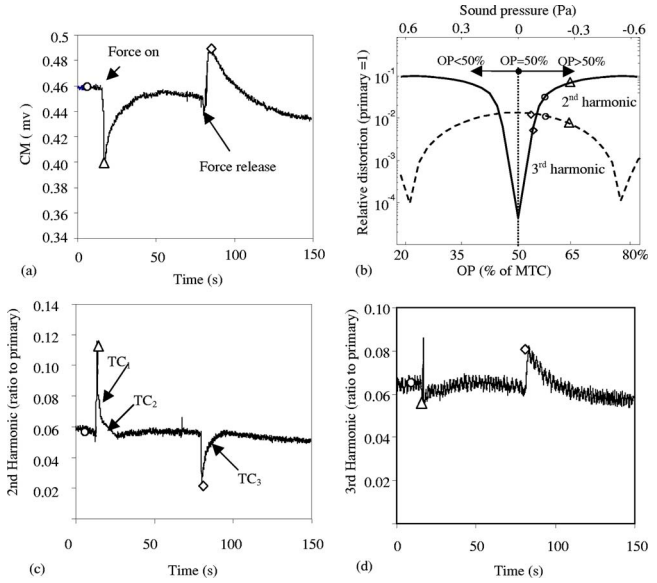


FIG. 5. CM (evoked by a 74 dB SPL sound) change induced by a constant force applied on the cochlea at the ST 18 kHz BF place (animal 04-28). The changes of CM amplitude (mv), the second harmonic, and the third harmonic are shown in panels (a), (c), and (d), respectively. The patterns of the change of the second and third harmonics were the same as those induced by AP perfusion in the ST (Fig. 3). Thus, the stereocilia move toward SV with the force applied to the lateral wall of the ST [panel (b)]. Removing the force caused a cochlear sensitivity-dependent overshoot of CM.

second harmonics, and the third harmonics showed an adaptation-like phenomenon with onset and offset time constants. Because it takes time to change the hydraulic perfusion pressure in the cochlea, artificial perilymph perfusion is not an ideal type of stimulus to perturb the OP and to measure the time constants. To study the adaptation-like phenomena, a more instantaneous pressure application is needed. Therefore, a custom designed force transducer was used to apply a force on the bony shell of the cochlea.

C. Force application to the cochlear bony shell at ST of basal turn

In this experiment, a constant 17 g force was applied on the bony shell of the cochlea by a custom-designed force transducer. The blunt probe was placed against the otic capsule at the ST 18 kHz BF place. Because of the OP changes over time when a low-frequency high-level probe is used, we selected the lower sound level of 74 dB SPL 400 Hz sound as the probe stimulus for OP. The CM waveform change with the force application was recorded for offline analysis of the second and the third harmonics.

When force was applied on an otic capsule, there was a sudden reduction of CM amplitude, which then underwent a slow partial recovery toward the initial level. Offline analysis showed that the second harmonic increased [Fig. 5(c)], and the third harmonic distortion decreased [Fig. 5(d)]. These changes were in the same direction as the second and the third harmonic changes that were caused by artificial perilymph perfusion through the ST in the control experiment. Thus the stereocilia moved toward SV with the force applied to the cochlear bone at the ST. OP moved away from the value of 50% MTC (P_0=0) to a value >50% MTC (P_0 <0), and the CM waveform became more asymmetric and distorted. During the application of a constant force, the cochlea appeared to have an "adaptation-like" mechanism to restore the sensitivity and optimal stereocilia position. The force caused a sudden increase of the second harmonic, and then underwent a slow partial recovery toward the initial level, which could be fitted by a two time constant exponential function to derive a fast time constant (TC_1) and a slow time constant (TC_2). We modeled the onset adaptation with the expression $A(t)=A(1)e^{t/TC_1}+A(2)e^{t/TC_2}$, where $A(t)$ is the value of the state at time t, $A(1)$ and $A(2)$ are the values of two exponential functions, respectively, at a zero time point, and t is the time that has elapsed from the start of the exponential decay. The fitted time constants are shown in Table II. This "adaptation-like" mechanism will be the subject in a later report.

TABLE II. The second harmonic of CM recovery time when the force was applied (TC_1 and TC_2), and released (TC_3) (n=16).

	TC_1 (fast TC)	TC_2 (slowTC)	TC_3 (forcerelease)
For 74 dB SPL	1.8±0.6	9.2±3.5	6.2±0.8
For 94 dB SPL	–	7.3±3	4.5±2
Postmortem	–	—	—

When the force was released, the CM amplitude increased and had an overshoot, which returned to the control level after a much longer time period [Fig. 5(a)]. The released force caused a sudden increase of the second harmonic, then underwent a slow partial recovery toward the initial level [Fig. 5(b)], which could be modeled by a single exponential function with a time constant (TC_3). The third harmonic distortion increased [Fig. 5(c)] upon removal of the force. These data indicate that the stereocilia moved toward the ST direction, i.e., back to the initial optimal position. The origin of the recovery is a dynamic shift in the OHC transduction operating point.

The CM waveform recovery time for the force applied and released is related to the sound level. For a high-level sound, 94 dB SPL, when applying the same force, the CM amplitude decreased less than for the 74 dB SPL sound. This could be because of saturation of OHC current at 94 dB SPL sound. Compared with the case of 94 dB SPL sound, the same deflection of OHC will induce less change in saturated OHC current. When the force was applied, the CM waveform recovery time with 94 dB SPL probe sound was much faster than that with the 74 dB SPL probe sound. Especially, the use of a high sound level, which results in a suppression of basal turn cochlear amplification, eliminates the overshoot (Fig. 6). Postmortem, all the "adaptation-like" phenomena were absent and the harmonic change resembled those from the use of a 94 dB SPL probe tone (Fig. 7).

D. Cochlear bone deformation from application of force on the external surface

Data indicate that the force on the otic capsule will result in the OP change. It is expected that the force can distort cochlear shape, causing deflection of OHC stereocilia. Therefore, in this *in vitro* experiment, the cochlear bone de-

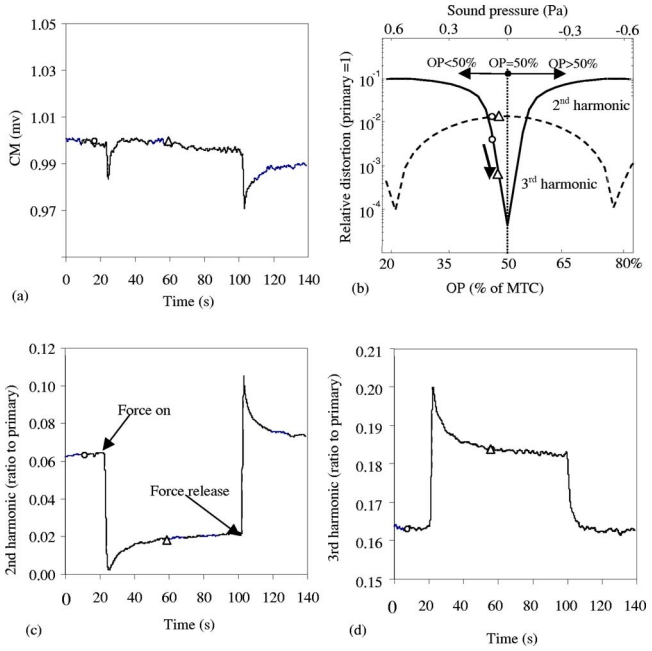


FIG. 6. Changes of CM and its harmonics of a 94 dB SPL sound induced by a constant force applied on the cochlear surface at the ST 18 kHz BF place (animal 05-27). The changes of CM amplitude (mv), second harmonic, and third harmonic are shown in panels (a), (c), and (d), respectively. Panel (b) shows a diagram of the OP position before (◦), during (△), and after (◇) the constant force application when the initial OP <50% MTC. The time constant of the slow partial recovery (TC_2) is 7.3±3 s, which is faster than that at the lower sound stimulation level (9.2±3.5 s). Removing the force caused CM to recovery to its normal level, but the overshoot was not observed.

formation by the force application can be calibrated.

The gold-coated beads were placed around the force application probe, whose positions are recorded as P_1, P_2, P_3, P_4, P_5, and P_6, as its distance away from the probe in Fig. 8(b). The 17 g force-induced displacement of individual

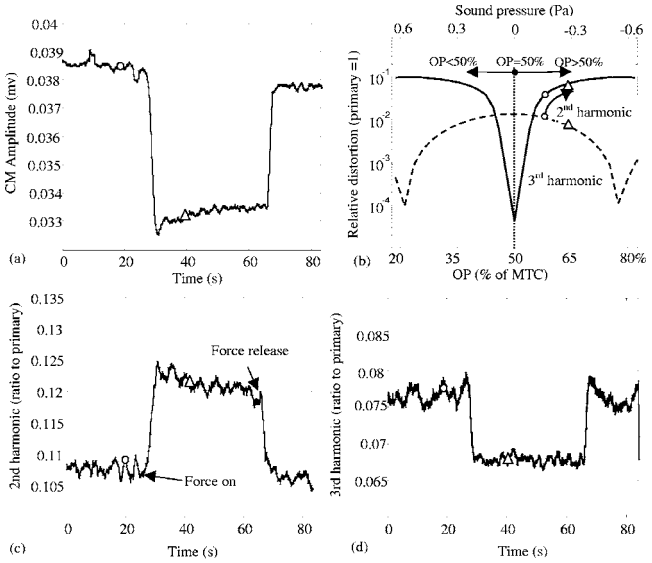


FIG. 7. The postmortem effects on the change of the CM evoked by a 94 dB SPL sound change induced by a constant force applied on the cochlea at the ST 18 kHz BF place. The changes of CM amplitude, shift of OP, the second harmonic, and the third harmonics are shown in panels (a), (b), (c), and (d), respectively. Under the postmortem condition, the initial OP moved to a value <50% MTC, and all "adaptation-like" phenomena were absent.

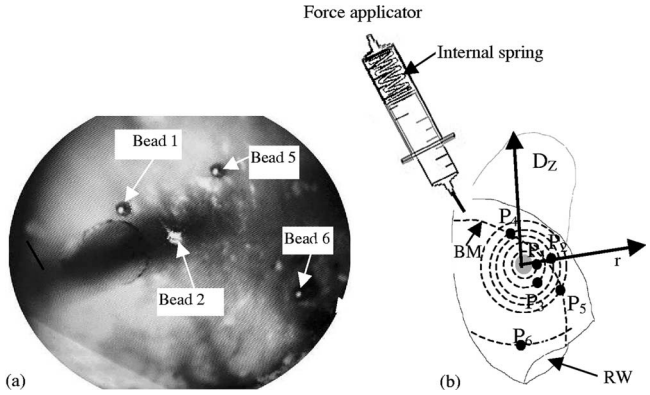


FIG. 8. A picture (a) and diagram (b) show bead locations on the cochlear bony surface, from which the deformation of the cochlear shell was measured using a laser interferometer

beads, $D'_z|_{Pn}$ $(n=1,2,3,4,5,6)$, was measured using an interferometer. The $D'_z|_{P6}$, displacement of beads P_6 whose position is distance away from the place touched by probe, was used as the reference point to calibrate the system movement (i.e., the whole cochlea movement independent of the probe-induced deformation). Therefore, the real displacement of every position is $D_z|_{Pn}$ $(n=1,2,3,4,5)=D'_z|_{Pn}(n=1,2,3,4,5)-D'_z|_{P6}$ (Table III).

Force on the otic capsule resulted in the cochlear shape distortion. Points close to the probe were forced inward by a greater amount than farther from the probe. For example, the bead located in the P_1 position moved inward 0.6 μm by 17 g force on the otic capsule. However, bead P_5 did not move (Table III). These data indicate that the force applied to the cochlear bony shell distorts cochlear shape and results in the deflection of OHC stereocilia, consequently changing the CM.

E. Force application to the cochlear bony shell at the SV of the basal turn

In this experiment, a constant 17 g force was applied on the otic capsule at the SV. Since one of the CM recording electrodes was placed in a hole created at the SV 18 kHz BF place, the blunt probe was pushed against the bony shell at about the 14 kHz BF place. The CM record electrode in SV was sealed on the cochlear surface with dental cement to prevent any leak of perilymph. The CM was produced by a 400 Hz 74 dB SPL sound. The CM waveform change with the force application was recorded and the second and the third harmonics were analyzed offline. The application of a force induced a sudden increase of CM amplitude [Fig. 9(a)]. At the same time, the second harmonic decreased [Fig. 9(b)], and the third harmonic distortion increased [Fig. 9(c)], which

TABLE III. The three-dimensional movement of beads located on cochlear bone at places P_1, P_2, P_3, P_4, P_5, and P_6, with the distances from the probe L (μm).

	P_1	P_2	P_3	P_4	P_5	P_6
L (μm)	100	300	500	750	1200	2000
D_z (μm)	−0.8	−0.56	−0.4	−0.3	−0.2	−0.2
D'_z	−0.6	−0.36	−0.2	−0.1	0	

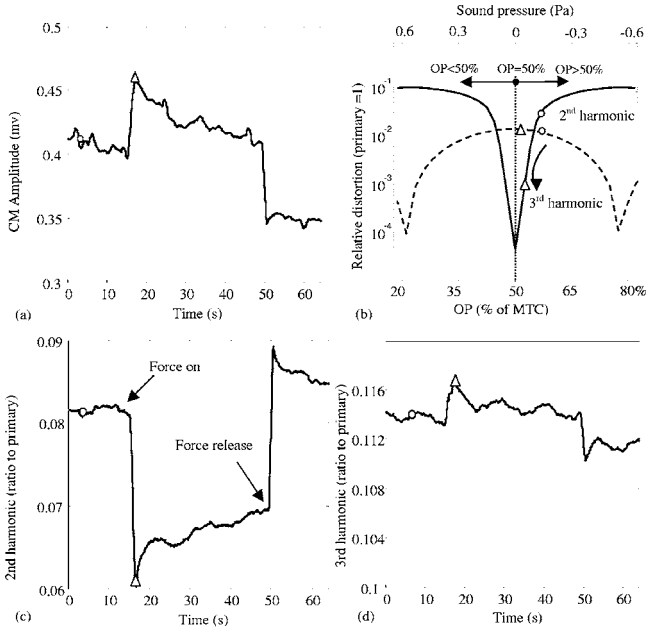


FIG. 9. CM change induced by a force applied on the lateral bony wall of the SV near the SV 14 kHz BF place (animal 09-10). The application of the force induced an increase of CM amplitude (a), a shift of OP toward the value of 50% MTC (b), a decrease of the second harmonic (c), and an increase of the third harmonic (d).

are opposite to the second and the third harmonic changes when AP was injected into the ST or force was applied on cochlea the bony shell at ST. The OP moved toward the value of 50% MTC (P_0=0), and the CM waveform became more symmetric and less distorted.

IV. DISCUSSION

A. The initial OP position

The CM recorded at a round window is thought to be generated by the OHCs near the recording electrode in response to displacement of the BM (Patuzzi *et al.*, 1989). This spatial bias is further enhanced when CM is recorded differentially at the 18 kHz BF by placing one electrode in the SV, and another at the round window. A high sound level given at low frequency will synchronize the phase of these basal turn OHCs but will generally not damage the cochlear sensitivity as travelling wave mechanics strongly attenuates their stimulation. Therefore, all previous studies to estimate the OP change from the fitted MET transfer function, analyze the CM produced by a low-frequency high-sound-level (>100 dB SPL) probe tone. However, such a sound can suppress cochlear sensitivity dramatically. In Fig. 2(b), when a 94 dB SPL sound was presented, the BM vibration velocity at 18 kHz was suppressed by about 32 dB [see Geisler and Nuttall (1997) for a further discussion of such low-side BM velocity suppression]. The previous studies that have used low-frequency high-sound-level probe tones assume that cochlear sensitivity is not influenced by this sound (Frank and Kössl, 1996; Patuzzi and Moleirinho, 1998; Sirjani *et al.*, 2004). However, any time CM shows a saturation distortion, the suppression of cochlear amplification will also occur.

It has been found, *in vivo*, that the OP of MET in OHCs from the basal turn of the mammalian cochlea is close to 50% (p_0=0), and the CM waveform is symmetrical (Russell *et al.*, 1986; Russell and Richardson, 1987; Russell and Kössl, 1991). *In vitro*, the OP of the OHCs is situated close to the channel's closed region of the stereocilia bundle transfer function, i.e., with approximately 10% of the MET channels open (Holton and Hudspeth, 1986). This is similar to the situation of inner hair cells, both *in vitro* and *in vivo* (Russell and Richardson, 1987; Russell and Kössl, 1991). The difference of OP between *in vitro* and *in vivo* experiments is thought to be due to the removal of the tectorial membrane for the *in vitro* preparation (Russell *et al.*, 1986; Russell and Richardson, 1987; Russell and Kössl, 1991). *In vivo*, the mechanical interaction with the tectorial membrane may allow OHC stereocilia to be adjusted in position to assure that MET channels operate symmetrically at a region of maximum gain of their transfer function. In other words, the OP may be optimized so that force generation by the OHCs is most effective (Russell *et al.*, 1986; Russell and Richardson, 1987; Russell and Kössl, 1991). In our study, in the sensitive animal group, we find that the initial OP is about 57% (P_0 =−0.1 Pa), which means the MET channels appear to be slightly displaced from optimal gain. The reason for this offset is not yet known.

B. OP shift depending on the sound stimulation level

It has been observed that the OP will change during and following a high level sound stimulation (Frank and Kössl, 1996; Patuzzi and Moleirinho, 1998). In Fig. 2(a), the OP moves in time from a value >50% MTC ($P_0<0$) into a value <50% MTC ($P_0>0$). Thus, the result obtained in an experiment such as force applied to the cochlea will depend upon the sound level of the probe tone. In Fig. 6, where a high-sound-level probe was used, the changes of the second and third harmonics follow quite a different direction as compared to the harmonics change for the lower level sound stimulation in Fig. 5. Yet in both cases, the force applied on the bony shell at the ST of basal turn will induce an OP change toward a value >50% MTC ($P_0<0$), indicating that the stereocilia are moved in the excitatory direction.

C. Time constant of second harmonic change

The OP changes toward a value <50% MTC ($P_0>0$) (more MET channels are closed) during loud sound. This could be the result of cell depolarization with prolonged stimulation resulting in cell shortening. OHC shortening would draw the reticular lamina surface of the organ of Corti toward ST deflecting stereocilia medially. Fast (prestin-based) motility could be the basis for the adaptation, but the long time constant we see (in seconds) is far slower than the inherent speed of fast motility. The slow depolarization of the OHC could also come from intracellular ionic shifts such as sodium loading (Zeddies and Siegel, 2004), or from potassium accumulation in the Nuel's space of the organ of Corti (Johnstone *et al.*, 1989). OHCs slow motility being calcium and ATP dependent (Zenner, 1986) could also be responsible for this OP change, as well as mechanical processes (such as cell turgor or stiffness). A change of intrac-

ellular cytosolic calcium could influence OP following the activation of second messenger pathways (e.g., by protein kinase G phosphorylation of cytoskeleton or prestin (Szonyi *et al.*, 1999; Sziklai *et al.*, 2001).

In the experiment of force application on the bony shell at ST, a constant force caused a rapid reduction of CM amplitude, which then underwent a slow partial recovery toward the initial level; and upon removing the force on the bony shell caused an overshoot of CM. These recovery and overshoot patterns are also cochlear sensitivity related and have a much faster time constant than a loud sound-induced OP shift. For the high-sound-level (>90 dB SPL) evoked CM, the overshoot with the release of force on the cochlea did not occur. Yet when the force was applied, the second harmonic waveform "recovery time" of high-sound-level (>90 dB SPL) evoked CM was much faster than that of the 74 dB SPL sound. For the insensitive animal, the "adaptation" response was lessened and for the postmortem condition, all these "adaptation-like" phenomena were abolished. The mechanical and/or biochemical pathways, responsible for force-induced OP change, are beyond the scope of this paper.

D. A model of organ of Corti response to otic capsule deformation

If the BM (and organ of Corti) were modeled as a simple elastic structure (a spring model), when the otic capsule is deformed, the BM would have reduced stiffness and a shorter dimension. In this case, both ST and SV position force application would be expected to cause the same relaxation of the BM/organ of Corti and thus the deflection of OHC stereocilia. However, we find that OHC stereocilia bundle deformation is dependent on the direction of force to the otic capsule.

When we applied force on the bony shell on ST and SV, respectively, it apparently induced BM movement toward different directions. This suggests that BM could be simulated by a beam model (Fig. 10). When force applied on the bone at ST, the bony shell deformed. If the BM attachment to the lateral wall is rigid enough, it will rotate the structure, thereby bending the BM toward SV. Similarly, force on the otic capsule at the SV would have the opposite effect.

V. CONCLUSIONS

The operating point movement of OHC stereociliary bundles could be derived from an analysis of the second and third harmonics of the CM waveform. The position of the stereocilia could be inferred to be shifted by sound or by applied force to the otic capsule. Data indicates that the force on the otic capsule can distort cochlear shape and result in the deflection of OHC stereocilia, causing CM to change.

Artificial perilymph perfusion through the ST was used to define the initial position and direction of a change of OP. Hydraulic bias of the BM toward SV will induce the stereocilia deflection toward the excitatory direction. Force applied to the otic capsule at the ST caused the same change in the second and the third harmonics of CM through the ST artificial perilymph perfusion. Therefore, the force applied to the ST shifts the stereocilia toward the excitatory direction. Force applied to the SV caused an opposite effect.

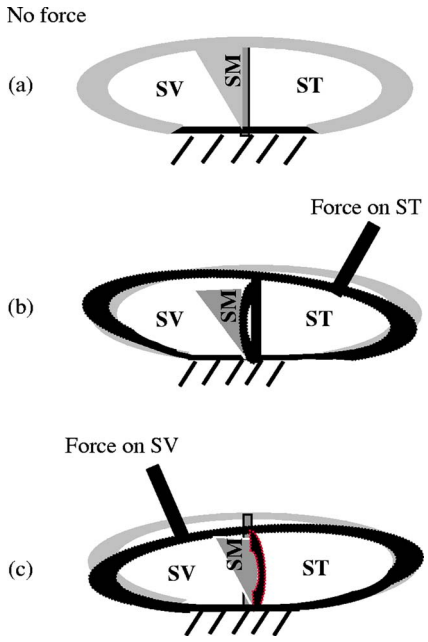


FIG. 10. The beam model of BM bending. The gray part is the cochlear bone position without a force application (a). When force is applied on the bone at ST, the cochlear bone will be deformed, and BM attachment to the lateral wall will bend the BM toward SV (b). Similarly, the force on the otic capsule at the SV will bend the BM toward ST (c). SM: scala media.

The application of perfusion hydraulic pressure, force, and loud sound all reveal a dynamic process of operating point change. The adaptation mechanisms probably differ depending upon the stimulus being used because OP time constants differ. Stereocilia "adapt" back toward control or the initial levels. This adaptation could be an important mechanism for OHC gain optimization or gain reduction to prevent damage.

ACKNOWLEDGMENTS

The study was supported by Grants R01 DC00141, DC00105, DC005983, and DC04554 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health.

Bian, L. (**2004**). "Cochlear compression: effects of low-frequency biasing on quadratic distortion product otoacoustic emission," J. Acoust. Soc. Am. **116**, 3559–3571.

Bian, L., Chertoff, M. E., and Miller, E. (**2002**). "Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions," J. Acoust. Soc. Am. **112**, 198–210.

Frank, G., and Kössl, M. (**1996**). "The acoustic two-tone distortions 2f1-f2 and f2-f1 and their possible relation to changes in the operating point of the cochlear amplifier," Hear. Res. **98**, 104–115.

Geisler, C. D., and Nuttall, A. L. (**1997**). "Two-tone suppression of basilar membrane vibrations in the base of the guinea pig cochlea using "low-side" suppressors," J. Acoust. Soc. Am. **102**, 430–440.

Holton, T., and Hudspeth, A. J. (**1986**). "The transduction channel of hair cells from the bullfrog characterized by noise analysis," J. Physiol. (London) **375**, 195–227.

Johnstone, B. M., Patuzzi, R., Syka, J., and Sykova, E. (**1989**). "Stimulus-related potassium changes in the organ of Corti of guinea-pig," J. Physiol. (London) **408**, 77–92.

Kirk, D. L., Moleirinho, A., and Patuzzi, R. B. (**1997**). "Microphonic and DPOAE measurements suggest a micromechanical mechanism for the

'bounce' phenomenon following low-frequency tones," Hear. Res. **112**, 69–86.
Kirk, D. L., and Patuzzi, R. (**1997**). "Transient changes in cochlear potentials and DPOAEs after low-frequency tones: the 'two-minute bounce' revisited," Hear. Res. **112**, 49–68.
Lukashkin, A. N., and Russell, I. J. (**2005**). "Dependence of the DPOAE amplitude pattern on acoustical biasing of the cochlear partition," Hear. Res. **203**, 45–53.
Nieder, P., and Nieder, I. (**1971**). "Determination of microphonic generator transfer characteristics from modulation data," J. Acoust. Soc. Am. **49**, 478–493.
Nuttall, A. L., Dolan, D. F., and Avinash, G. (**1991**). "Laser Doppler velocimetry of basilar membrane vibration," Hear. Res. **51**, 203–213.
Patuzzi, R. B. (**2002**). "Non-linear aspects of outer hair cell transduction and the temporary threshold shifts after acoustic trauma," Audiol. Neuro-Otol. **7**, 17–20.
Patuzzi, R. B., and Moleirinho, A. (**1998**). "Automatic monitoring of mechano-electrical transduction in the guinea pig cochlea," Hear. Res. **125**, 1–16.
Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (**1989**). "The origin of the low-frequency microphonic in the first cochlear turn of guinea-pig," Hear. Res. **39**, 177–188.
Russell, I. J., and Kössl, M. (**1991**). "The voltage responses of hair cells in the basal turn of the guinea-pig cochlea," J. Physiol. (London) **435**, 493–511.
Russell, I. J., and Richardson, G. P. (**1987**). "The morphology and physiology of hair cells in organotypic cultures of the mouse cochlea," Hear. Res. **31**, 9–24.
Russell, I. J., Cody, A. R., and Richardson, G. P. (**1986**). "The responses of inner and outer hair cells in the basal turn of the guinea-pig cochlea and in the mouse cochlea grown *in vitro*," Hear. Res. **22**, 199–216.
Sirjani, D., Salt, A. N., Gill, R. M., and Hale, S. A. (**2004**). "The influence of transducer operating point on distortion generation in the cochlea," J. Acoust. Soc. Am. **115**, 1219–1229.
Sziklai, I., Szonyi, M., and Dallos, P. (**2001**). "Phosphorylation mediates the influence of acetylcholine upon outer hair cell electromotility," Acta Oto-Laryngol. **121**, 153–156.
Szonyi, M., He, D. Z., Ribari, O., Sziklai, I., and Dallos, P. (**1999**). "Cyclic GMP and outer hair cell electromotility," Hear. Res. **137**, 29–42.
Zeddies, D. G., and Siegel, J. H. (**2004**). "A biophysical model of an inner hair cell," J. Acoust. Soc. Am. **116**, 426–441.
Zenner, H. P. (**1986**). "Motile responses in outer hair cells," Hear. Res. **22**, 83–90.

Frequency- and level-dependent changes in auditory brainstem responses (ABRs) in developing mice

Lei Song, JoAnn McGee, and Edward J. Walsh[a)]
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131
and Department of Biomedical Sciences, Creighton University School of Medicine, 2500 California Plaza, Omaha, Nebraska 68178

(Received 15 September 2005; revised 2 February 2006; accepted 6 February 2006)

The development of the auditory brainstem response was studied to quantitatively assess its dependence on stimulus frequency and level. Responses were not observed to stimuli ⩾16 kHz on P12, however, the full range of responsive frequencies included in the study was observed by P14. Response thresholds were high on P12, exceeding 100 dB SPL for all stimuli tested. The rate of threshold development increased progressively for stimulus frequencies between ~2 and 10 kHz, with the most rapid changes occurring at frequencies >10 kHz. Adultlike thresholds were observed by P18. Response latencies and interpeak intervals matured rapidly over the course of the second and third postnatal weeks and did not achieve adultlike characteristics until after P18. Latencies of higher-order peaks were progressively and sequentially delayed relative to wave I. Wave I amplitudes developed nonmonotonically, growing during the first 24 days and stabilizing at adult values by ~P36. Slopes of wave I amplitude-and latency-level curves were significantly steeper than those of adults during the neonatal period and the outcome of input-output analyses, as well as frequency-specific maturational profiles, support developmental models in which function initially matures in the mid-frequency range and proceeds, simultaneously, in both apical and basal directions. [DOI: 10.1121/1.2180533]

PACS number(s): 43.64.Pg, 43.64.Qh, 43.64.Nf, 43.64.Ri [BLM] Pages: 2242–2257

I. INTRODUCTION

Mice have become popular subjects of auditory research over the course of the past decade or so, a situation that is at least partially the consequence of a successful effort to map the murine genome and to develop molecular tools to assess the functional relevance of signaling pathways operating within specific tissue compartments and regulating the function of specific cell types. As anticipated, efforts to determine the genetic and molecular basis of cochlear development have revealed a complex system of regulation involving a relatively large number of maturational factors (Torres and Giraldez, 1998; Muller and Littlewood-Evans, 2001; Chen and Corey, 2002; Zine, 2003). In that light, over 80 loci associated with nonsyndromic forms of deafness have been mapped in humans and at least 40 genes associated with deafness have been identified (Van Camp and Smith, 2005). The resulting body of knowledge establishes a platform from which the molecular basis of auditory pathology, as well as a more refined understanding of hearing mechanisms generally, can be understood (for review see Petit *et al.*, 2001; Steel and Kros, 2001; McGee and Walsh, 2004) and used to advance treatments in this area (Duan *et al.*, 2002; Holley, 2002; Li *et al.*, 2003; Izumikawa *et al.*, 2005). In that context, one purpose of this study was to comprehensively characterize auditory function in BALB/c mice using a noninvasive measure, the auditory brainstem response (ABR), as a prelude to a study of the neurological and otological consequences of genetically determined congenital hypothyroidism in the *Tshr*^{hyt} strain of mutant mice that are maintained on a BALB/c background. The primary purpose of the study was to comprehensively characterize the development of auditory function in the frequency domain and thereby revisit the question of how the space-frequency map is laid down during development.

The final stages of morphological and electrochemical inner ear development in altricial mammals have been studied extensively and the resulting model is one of dynamic and rapid change during which adultlike peripheral auditory function is acquired over the course of approximately 1 to 2 weeks [see Walsh and Romand (1992) for review]. Changes in cochlear morphology that occur during this period are far reaching and include the opening of the tunnel of Corti and space of Nüel (Pujol and Hilding, 1973; Kraus and Aulbach-Kraus, 1981; Sato *et al.*, 1999; Souter *et al.*, 1997), the transformation of the epithelium of the inner spiral sulcus (Kölliker's organ) (Pujol and Hilding, 1973; Kraus and Aulbach-Kraus, 1981; Roth and Bruns, 1992; Souter *et al.*, 1997; Sato *et al.*, 1999), the reduction of the basilar membrane's (BM) tympanic cover layer, the thickening of the BM (Kraus and Aulbach-Kraus, 1981; Roth and Bruns, 1992; Souter *et al.*, 1997; Sato *et al.*, 1999), the disappearance of marginal pillars (Kraus and Aulbach-Kraus, 1981; Roth and Bruns, 1992; Souter *et al.*, 1997), and the differentiation of both outer and inner hair cells (Pujol and Marty, 1970) as well as surrounding support cells (Souter *et al.*, 1997). Along with these relatively clear anatomical changes, the stria vascularis acquires maturity in conjunction with other components that together comprise the electroanatomy of the mammal cochlea, and the endocochlear potential develops

a)Electronic mail: walsh@boystown.org

 0001-4966/2006/119(4)/2242/16/$22.50

accordingly (Bosher and Warren, 1971; Fernandez and Hinojosa, 1974; Woolf *et al.*, 1986; Rybak *et al.*, 1992; McGuirt *et al.*, 1995; Sadanaga and Morimitsu, 1995).

Among altricial mammals studied thus far, the organ of Corti rapidly acquires the filter properties observed in adult animals after development enters its final stages. Although the commonly accepted, longstanding view of this process is that development begins in the basal turn of the cochlea and proceeds in a baso-apical direction, more recent studies suggest that the process actually begins at a mid-basal location and proceeds in both directions [see Walsh and Romand (1992) for a review]. During this period, the basilar membrane is sculpted from a continuous sheet of acellular, fibroelastic tissue producing the tonotopic organization observed in adult mammals (Schweitzer *et al.*, 1996; Kuhn and Vater, 1997). Data presented here support the view that development begins at a point in the basal turn, possibly a mid-basal location, and progresses in an apico-basal direction in addition to the more widely recognized baso-apical direction.

As with the brain generally, the auditory pathway also continues developing during this period. Auditory nerve fibers and central tracts alike undergo myelination (Romand *et al.*, 1976; Walsh and McGee, 1986) and synaptogenesis and neurite extension occurs generally (Pujol and Abonnenc, 1977; Pujol *et al.*, 1979; Lenoir *et al.*, 1980; Pujol *et al.*, 1980; Simmons *et al.*, 1996). It is against this background that mature auditory function emerges in mammals.

II. MATERIAL AND METHODS

A. Animals

A total of 64 mice made up of 43 mice from the BALB/c strain and 21 heterozygotes of the *Tshr*^{hyt} strain that is maintained on a BALB/c background, drawn from 26 litters, were used in this study. Although this mouse strain (*Tshr*^{hyt}) carries a point mutation in the gene that encodes the thyrotropin receptor and homozygous mutant animals exhibit autosomal recessive hypothyroidism, heterozygotes are euthyroid and exhibit normal otological characteristics (Sprenkle *et al.*, 2001) that are indistinguishable from those of the progenitor strain (unpublished observations). Consequently, heterozygotes were selected as subjects in the present investigation.

BALB/c and *Tshr*^{hyt} mice were originally obtained from the Jackson Laboratory (Bar Harbor, ME) and breeding pairs were established in-house to generate all experimental subjects. Pregnant females were separated from males and housed individually in separate cages. Litters were weaned between postnatal day 21 (P21) and P28, with the day of birth designated P0.

Animals were randomly selected from litters (approximately 49% male and 51% female) and were studied between P12 and P90. The data sampling strategy employed in this study was based on previous findings suggesting that developmental changes occur rapidly between P12 and P15 and gradually thereafter (Sprenkle *et al.*, 2001). Consequently, data were acquired daily between P12 and P15 and intermittently between P18 and P90. Based on *posthoc* observations, data acquired on P24–P25, P36–P37, P48–P49, and P60–P90 were subsequently pooled for further analyses, unless otherwise noted. Individuals comprising the P12, P13, and P14 groups consisted of wild type BALB/c mice exclusively, while *Tshr*^{hyt} heterozygotes were included in older age groups (P15–P90). Individual animals were studied longitudinally when possible and each age group included animals derived from at least five different litters, except for groups P12 and P36 and P37 which were composed of animals from three and four litters, respectively. All procedures used for this study were approved by the BTNRH Animal Care and Use Committee.

B. Recording and stimulation procedures

ABRs were recorded in animals anesthetized with chloral hydrate (480 mg/kg IP) and supplemental doses (120 mg/kg) were administered as needed. Initial doses were decreased to either 360 or 240 mg/kg and supplemental doses to 90 mg/kg in the case of young pups. Body temperature was thermostatically regulated and maintained at 38.5 °C throughout recording sessions. All recordings were conducted in a double-walled sound-attenuating chamber (Industrial Acoustics Corp.).

Platinum subdermal needle electrodes (Grass Instruments) were positioned at the vertex (active, non-inverting), the infra-auricular mastoid region (reference, inverting), and in the neck region (ground). Differentially recorded scalp potentials were amplified 100 000×, bandpass filtered between 0.03 and 10 kHz (Grass Model P511K), and digitized (Cambridge Electronics Design 1401 *plus*) using a 20-kHz sampling rate over a 15-ms epoch. A total of 200 trials were averaged for each waveform and two waveforms were obtained for each stimulus condition.

Symmetrically shaped tone bursts were 3 ms long (1-ms raised cosine on/off ramps and 1-ms plateau). Click stimuli were 96 μs in duration and both clicks and tone bursts were generated digitally using a clock rate of 125 kHz. All acoustic stimuli were delivered free field via a high impedance piezoelectric tweeter (Radio Shack) positioned 10 cm from the vertex. Stimulus levels were calibrated using a 0.5-in. Brüel and Kjær microphone (Model 4134) positioned at the approximate location of the subject's head during recording sessions and are reported in decibels sound pressure level (dB SPL: referenced to 20 μPa). Stimuli of alternating polarity were delivered at a rate of approximately 10/s.

ABR thresholds were determined for clicks and for tone bursts in half octave steps ranging from 32 to 2.0 kHz. The effects of level on response amplitudes, latencies, and interpeak intervals were determined by decreasing stimulus intensity from suprathreshold values (>50 dB above threshold when possible) to threshold in 10-dB steps. Near threshold, levels were adjusted in 5-dB steps. A maximum stimulus level of 90 dB SPL was used for older animals.

C. Data acquisition and analysis

Digitized response waveforms were displayed in real time during recording sessions and trails with voltages exceeding 70 μV were rejected and the trial repeated. Re-

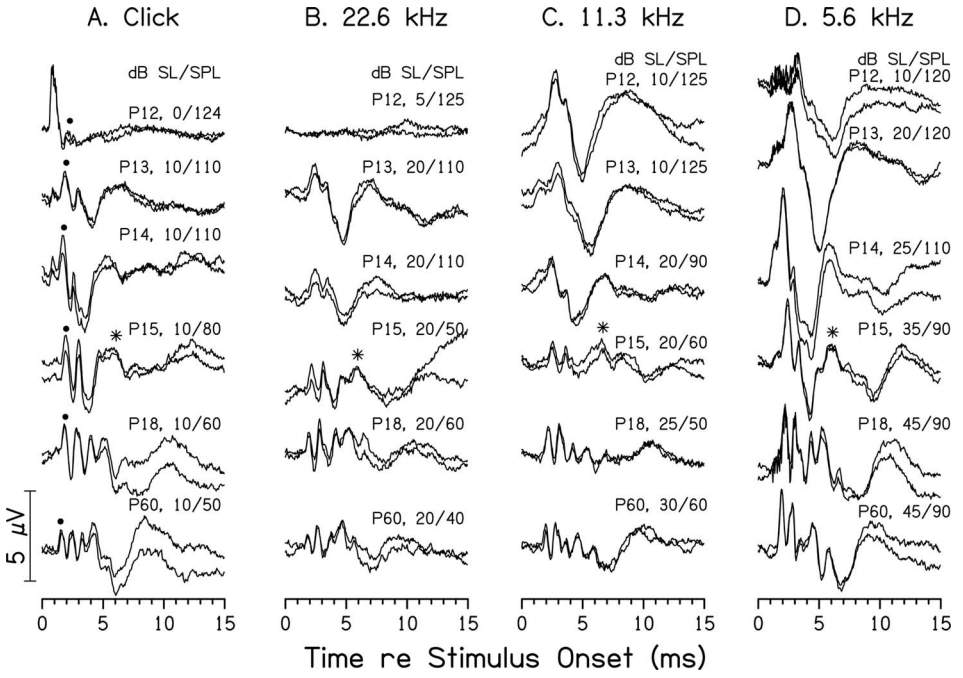


FIG. 1. Representative examples of ABR waveforms obtained in response to the stimuli indicated at the top of each column illustrate the evolution of waveform morphology during the period of functional development (i.e., P12–P18 and P60). Two overlapping replicates are shown at each age. Stimulus levels in dB are designated in terms of both sensation level (SL) and sound pressure level (SPL). For animals older than P14, 90 dB SPL was the highest stimulus exposure level. Asterisks associated with P15 waveforms identify wave IV, a response feature that can be difficult to identify in isolated response waveforms as it emerges from the broader wave III/IV typically observed in younger mice. Wave I is identified by filled circles for all ages in panel (a); the peak at P12 preceding wave I is the summating potential.

sponse waveforms were stored in computer files upon completion of a run for use in off-line analyses. Threshold was defined as the lowest stimulus level eliciting a replicable response. Values representing response maxima (peaks) and minima (valleys) were identified manually using cursors, and x,y coordinates for each value were automatically stored in computer files. Peak latencies were determined relative to the onset of the stimulus and included air conduction time, which was estimated to be approximately 0.295 ms, and peak amplitudes were computed using a triangulation procedure as in Walsh *et al.* (1986c). Custom software was used for data acquisition, measurements of peak amplitudes, latencies and interpeak intervals, and subsequent analyses, including curve-fitting procedures.

Best-fitting decaying exponentials of the form $y=a+be^{-cx}$ were used to characterize the time course of threshold and latency development. The independent variable, x, represents age in postnatal days, and the dependent variable, y, represents either threshold in dB SPL or wave latency in ms, and e^{-cx} is the exponential function, $\exp(-cx)$. The value of "a" denotes asymptote and is used to estimate adult values and the implied value of the dependent variable at birth (y intercept) is given by the sum of "a" and "b." The overall rate of development is indicated by the value of "c," with the reciprocal of "c" being the time constant (τ) of the function. Generally, time constant estimates were adjusted by adding the age of the earliest response to computed values so that developmental rates were presented in terms that matched the actual developmental time course. In addition, rates of threshold improvement during the rapid phase of development (P12–P15) were determined for each stimulus condition using least-squares linear regression analyses. The significance of each regression (i.e., that the regression slope differed from zero) was tested using an ANOVA approach, and comparisons of the rates of threshold improvement were made across frequency. The proportion of threshold variance accounted for by age was determined for both exponential and linear regressions. Similarly, the proportion of latency variance that was accounted for by age was determined for each wave and stimulus condition.

Slopes of latency-intensity curves were analyzed by (1) normalizing each latency-level curve to the latency value observed at the highest level (i.e., subtracting the latency at the highest level from all other values and adding a value of 1 to the difference), (2) converting normalized latencies to logarithmic values, and (3) performing a least-squares linear regression of $\log_{10}$-latency versus stimulus level (in dB) curves; results are expressed in log μs/dB units. Slopes of amplitude-level curves were derived from a least squares linear regression analysis of responses that fell in the near-threshold response range, and for older animals a second slope estimate was computed for responses evoked by higher level stimuli (approximately 70 dB SPL and higher) [see Fig. 14(c)].

Latencies, interpeak intervals, and amplitudes were compared across stimulus level and age using a general linear model (GLM) approach with repeated measures across level and limited to the level ranges and age groups specified in the legend for each relevant graph. Similarly, slopes of input-output (latency and amplitude versus level) analyses were compared across frequency and age using a GLM with repeated measures across frequency and limited to the frequency range represented by the age groups tested. Bonferroni *posthoc* tests were used for multiple comparisons to determine the age that the dependent variable acquired adult values. Results were considered statistically significant when $P<0.05$, unless otherwise specified.

III. RESULTS

A. General findings

ABR waveforms recorded from adult BALB/c mice (see responses obtained on P60 in Fig. 1) were similar to responses recorded from other mammals in previous studies

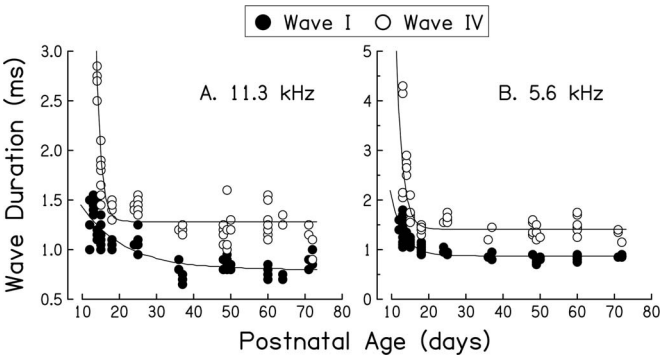


FIG. 2. The duration of ABR waves I (●) and IV (○) elicited by 11.3-kHz tone pips [panel (a)] and 5.6-kHz tone pips [panel (b)] decreased during the period of active, rapid development. Duration was calculated by subtracting peak minima representing the leading and trailing "edges" of both waves I and IV (i.e., subtracting the latencies corresponding to the "valleys" preceding and following each wave peak). Lines represent exponential fits to the data. Values for R^2 ranged from 0.69 to 0.80.

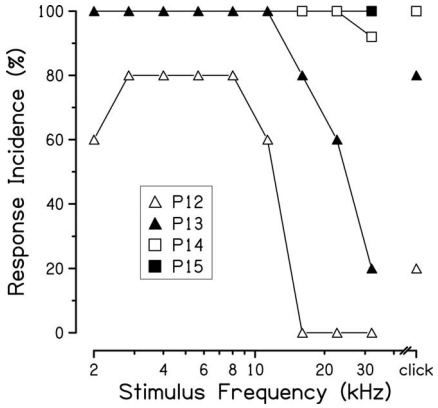


FIG. 3. The percentage of individuals that responded at the maximum output of the sound system for all stimulus frequencies tested is shown with age as parameter. At P12, the majority of mice responded to stimuli between 2 and 11.3 kHz but not to higher frequencies. Developmental changes occurred generally between P12 and P13 for stimuli ≤11.3 kHz and between P13 and P15 at higher stimulus frequencies.

(cats: Jewett and Romano, 1972, Walsh *et al.*, 1986a), (gerbils: Smith and Kraus, 1987, McFadden *et al.*, 1996), (rats: Jewett and Romano, 1972, Iwasa and Potsic, 1982, Blatchley *et al.*, 1987), (monkeys: Doyle *et al.*, 1983), (rabbits: Pettigrew and Morey, 1987), (hamsters: Schweitzer, 1987), (humans: Stein *et al.*, 1981), consisting of a series of highly replicable voltage deflections that occurred within 10 ms following acoustic stimulation (Fig. 1). Specifically, five positive-going, sharp voltage peaks each lasting ~1.0 ms (Fig. 2) were occasionally followed by either a sixth broad positive-going peak that lasted ~5 ms or two additional peaks, waves VI and VII lasting ~2.5–3 ms each. Response peak duration was estimated by computing the difference between amplitude minima ("valleys") preceding and following each positive-going peak comprising the waveform.

Waveforms recorded from immature animals differed significantly from those of adults (see response waveforms recorded from P12 and P13 animals in particular as part of Fig. 1). Early on, just following the "onset" of function, response amplitudes were small and the number of discriminable positive-going peaks was less than that observed in adults. Responses were occasionally dominated by the summating potential on P12 (see Fig. 1, response to clicks on P12), although the number of animals that responded to acoustic stimulation was small (e.g., only one animal out of five responded to the maximum sound level available in the case of clicks). Although the first and second waveform peaks observed in 12- and 13-day-olds appeared to correspond with the first and second peaks observed in adults, the third peak appeared as a compound wave corresponding to adult waves III and IV. Neither wave V nor later-occurring waves were observed during the earliest stage of functional development.

As with adults, the duration of wave I was unaffected by stimulus intensity (data not shown), regardless of age, and average wave duration decreased from ~1.6 ms in the case of responses to 11.3 kHz on P13 to ~0.8 ms in adults. In the case of wave IV (wave complex III/IV in very young animals), wave duration varied somewhat with near-threshold stimuli, however, the more notable change occurred with age such that average duration decreased from ~2.9 ms to just over 1.3 ms in adults for 11.3 kHz. Similar changes were observed in the duration of waveform peaks I and IV (or III/IV in very young animals) in response to 5.6 kHz between the age that responses were first observed and P15. However, in the latter case, values decreased over a wider range for both peaks; i.e., the width of peak I decreased from just under 1.8 ms to ~0.9 ms and peak IV duration decreased from ~4.3 to ~1.4 ms (Fig. 2).

B. Development of response incidence and sensitivity

On the 12th postnatal day four of the five animals studied responded to 125 dB SPL tone bursts between 2.8 and 8.0 kHz and three responded to simulation at 2 and 11.3 kHz; no responses were observed at or above 16 kHz. By P13, all animals tested ($n=10$) responded to tone bursts at and below 11.3 kHz, but only 80%, 60%, and 20% of that group responded to 16-, 22.6-, and 32-kHz tone bursts, respectively. By P14, 11 of 12 animals responded to all stimulus conditions and one animal failed to respond to 32-kHz tone bursts presented at 117 dB SPL, the maximum stimulus level available at that frequency. All animals responded to all test stimuli by P15. These findings are represented graphically in Fig. 3.

Average response thresholds of responsive animals were uniformly high on P12 as shown in Fig. 4, with estimates exceeding 100 dB SPL for all stimulus conditions. Changes in the overall form of threshold versus frequency curves during the active phase of functional development is best described as an increase in the upper cutoff frequency coupled with a dramatic decrease in threshold, particularly for responses to higher frequency tone bursts. At 11.3 kHz, for example, thresholds improved by 18 dB between P12 and P13, by 26 dB between P13 and P14, and by ~33 dB between P14 and P15. Thresholds at 11.3 kHz were within 18 dB of average adult estimates by P15. As illustrated in Fig. 4, thresholds improved ~95 dB between P12 and the age of maturity in the case of 11.3 kHz and threshold im-

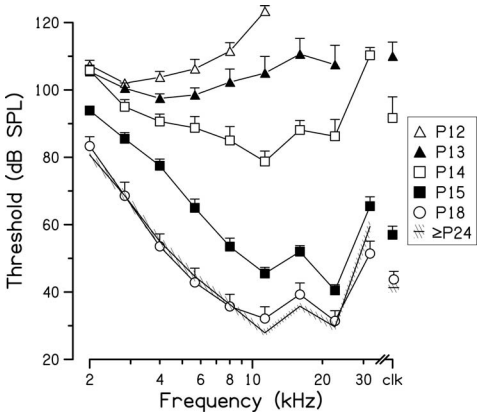


FIG. 4. Average threshold versus frequency curves reveal the progressive increase in sensitivity observed between P12 and P18. The responsive bandwidth also increased between P12 and P14. Thresholds to click stimuli are also shown. Standard errors of the mean are indicated for each group. Thresholds obtained from P24 through P90 were pooled and served as adult values. Thresholds at all ages below P18 were significantly different ($P<0.001$) from adults ($\geq$P24), except for those obtained at P15 in response to 32 kHz. Thresholds acquired at P18 did not differ significantly from those of adults.

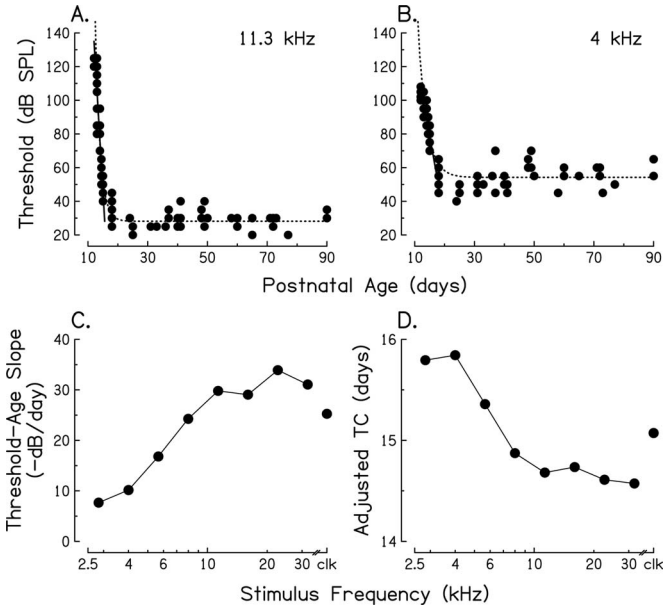


FIG. 5. (a) and (b) Examples of ABR thresholds plotted as a function of postnatal age are shown for responses to 11.3 kHz [panel (a)] and 4 kHz [panel (b)]. Dotted lines represent decaying exponentials of the form $y=a+be^{-cx}$ and are fitted to values between birth and 90 postnatal days. Parameters for the fit are $a=28.1$, $b=178383$, and $c=0.59$ for the data shown in panel (a) and 54.2, 4432, and 0.35, respectively, for the data shown in panel (b). (c) Slopes of least-squares linear regressions fitted to threshold-age values between birth and P15 (shown as solid lines in panels (a) and (b) are plotted as a function of stimulus frequency and the slope computed for responses to click stimuli is also shown. Values for r^2 ranged between 0.59 (2.8 kHz) to 0.88 (16 kHz) and the average was 0.78. Slopes and y intercepts computed for each stimulus condition were statistically significant ($P<0.001$). (d) Time course of threshold development estimated from exponential fits indicated in the form of adjusted time constants [time constant (TC) plus age of first response to airborne sound] is plotted as a function of stimulus frequency and for click stimuli. R^2 values computed from the fit ranged between 0.65 (32 kHz) and 0.92 (16 kHz) and the average was 0.85.

provement progressively decreased at frequencies lower and higher than 11.3 kHz, such that the overall gains in sensitivity observed during maturation were ~25 and ~55 dB for 2 and 32 kHz, the lowest and highest frequencies studied, respectively.

Although the overall character of threshold versus age curves is relatively well represented by an exponential expression, during the first days following the onset of function sensitivity remained relatively constant for frequencies $\leq$2.8 kHz. That observation notwithstanding, the overall rate of threshold development was determined for all stimulus conditions studied by plotting threshold estimates as a function of age and computing the time constant of the best-fitting exponential. Representative threshold versus age plots, along with exponential fits, are shown in Figs. 5(a) and 5(b) for responses to 11.3- and 4-kHz tone bursts, respectively. Maturation rate was clearly frequency dependent, and based on adjusted time constant estimates derived from exponential fits to threshold-age data, adultlike sensitivity to high-frequency stimuli was achieved at a rate faster than in the case of lower frequency stimuli [Fig. 5(d)]. Specifically, sensitivity to tone bursts greater than ~10 kHz matured at nearly the same rate while the rate of maturation gradually decreased for stimulus conditions below 10 kHz.

The rates of threshold change during the period of rapid development (P12–P15) were also estimated by analyzing threshold versus age curves using a least-squares linear regression approach [Fig. 5(c)]. Regression parameters, slope, and Y-intercept values were statistically significant ($P<0.001$) for all stimuli. The rate of threshold improvement was directly related to stimulus frequency below ~11.3 kHz, and rates for high-frequency tone bursts (11.3–32 kHz) were uniformly steep (~30 dB/day). Rates of threshold improvement for middle and high frequencies ($\geq$5.6 kHz) were significantly greater than those to lower frequencies ($P<0.05$). In the case of click stimuli, thresholds decreased by approximately 25 dB/day. Age accounted for a significant portion of the variance in threshold ranging from 88% at 16 kHz to 70% at 4 kHz, and at 2.8 kHz age accounted for 59% of the variance.

C. Development of response latency

As with other mammals thus far studied, the latencies of ABR peaks were initially prolonged in BALB/c mice and acquired adultlike values in an orderly manner over the course of the first 20–30 postnatal days. However, the latency of later occurring waves, and wave IV in particular, continued to decrease after P30 in the case of responses to lower frequency stimuli (Fig. 6). From a qualitative perspective, a rapid developmental phase lasting approximately 2 to 3 days after responses to airborne sounds were first observed (P12–P15) was followed by a gradual maturational phase that brought wave I response latency into the adult range by the end of the third postnatal week.

Although latency maturation could be separated into an early phase of rapid development and a subsequent slow phase during which latency values gradually achieved maturity, the time course over which response latencies developed was reasonably approximated by computing the time constant (τ) of best fitting exponentials representing latency-age plots, as in Figs. 6(a)–6(c). Using adjusted time constants to

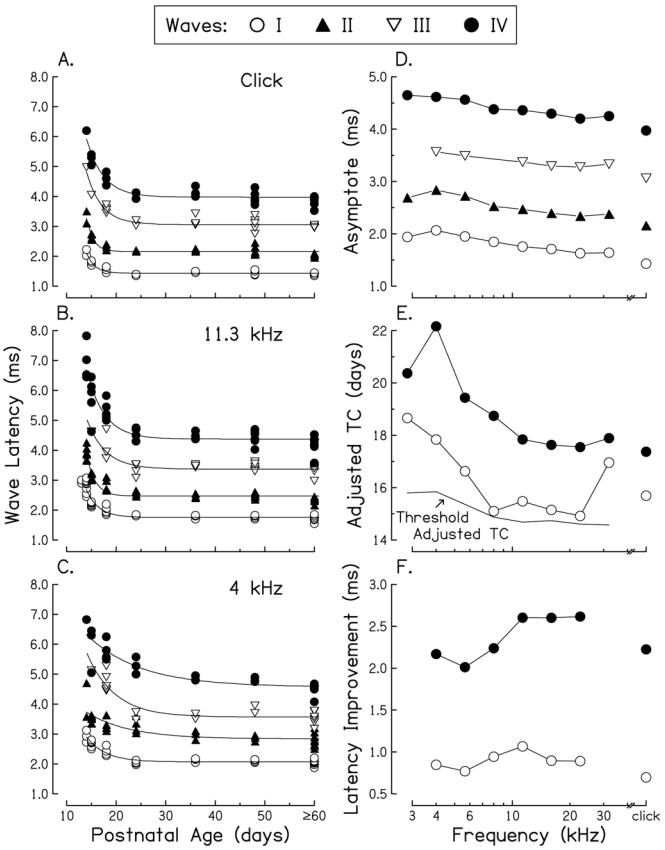


FIG. 6. (a)–(c) ABR wave latencies obtained in response to clicks (a), 11.3-kHz (b), and 4-kHz (c) tone bursts presented at 90 dB SPL are plotted as a function of postnatal age. Solid lines represent decaying exponentials fitted to ABR waves I–IV using a least-squares procedure. Values for R^2 ranged from 0.88 to 0.89, 0.76 to 0.90, and 0.63 to 0.89 for the fits shown in panels (a)–(c), respectively. (d) Asymptotic values derived from exponential fits shown in (a)–(c) and at other stimulus frequencies studied are shown for each wave. (e) The rate of latency maturation was determined by comparing time constants derived from exponential fits and adjusted by adding the postnatal age at which responses were first observed at 90 dB SPL. The rate of threshold maturation versus stimulus frequency is replotted from Fig. 5(d) for comparison. (f) The overall range over which latencies improved during development was determined by subtracting latency estimates measured in response to 90 dB SPL on P14 from latency estimates observed in adults.

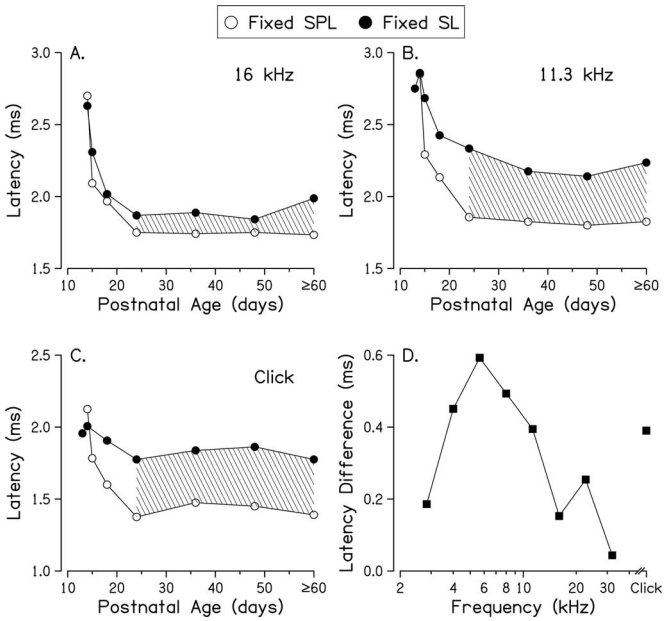


FIG. 7. (a)–(c) Mean latencies of wave I recorded using stimuli presented at fixed sensation levels (SL; filled symbols) and at fixed sound pressure levels (SPL; open symbols) are plotted as a function of age for 16-kHz (a), 11.3-kHz (b), and click (c) stimuli. Hatched areas represent latency improvement differences that could not be accounted for based on threshold elevation, a value referred to as *latency residue* in this study. (d) *Latency residue* observed in response to stimuli presented at a constant SPL was computed by subtracting asymptotic latency values associated with responses to 90 dB SPL stimuli from those associated with the appropriate SL condition and is plotted as a function of frequency.

assess and compare the overall rates of latency maturation for peripheral (wave I) and mesencephalic (wave IV) components of the ABR, it is clear that the endpoint of maturation was delayed by 2 to 4 days, depending on stimulus frequency, for response generators located in the vicinity of the midbrain (wave IV) relative to the periphery (wave I) [Fig. 6(e)]. At 32 kHz, the highest frequency studied, the rate of latency maturation was nearly the same for waves I and IV. Although wave I response latencies representing frequencies in the ~8–22.6-kHz band matured at about the same rate as threshold, development was notably delayed for higher and lower frequencies studied here, as shown in Fig. 6(e). The time course over which wave IV latencies matured was significantly extended for most stimuli studied. As shown in Fig. 6(f), the range over which latency values decreased during maturation was larger in the case of later occurring peaks than that observed in the case of earlier occurring waves. Specifically, wave IV latencies declined between ~2 and 2.6 ms, and wave I values declined between ~0.8 and 1.1 ms. When combined with asymptote values shown in Fig. 6(d), a clear and orderly latency maturation profile similar to that reported for other mammals is evident in the case of BALB/c mice.

D. Influence of threshold on latency maturation

Because response latency is inversely related to stimulus intensity, it is reasonable to assume that elevated thresholds contribute to the latency prolongation observed during development by reducing effective stimulus levels (i.e., sensation levels), expressed in this analysis as dB *re* ABR threshold. The relative contribution that elevated thresholds make to response latency during periods of immaturity can be determined by comparing response latencies under constant sound pressure level (SPL) and constant sensation level (SL) conditions. If SPL and SL derived latency-age curves are identical, one would reasonably conclude that threshold is not a factor affecting latency prolongation observed during development. On the other hand, if differences between SPL and SL determined curves are observed, it would be reasonable to conclude that elevated threshold is at least one factor affecting response latency. To estimate the contribution of threshold, latencies representing constant SPL conditions must be the same as response latencies representing constant SL conditions at some arbitrarily defined point in development; i.e., latencies associated with SPL and SL responses must be normalized. Although it may be argued that the analysis should be based on equivalent adult responses, insufficient acoustic power was available to generate matching

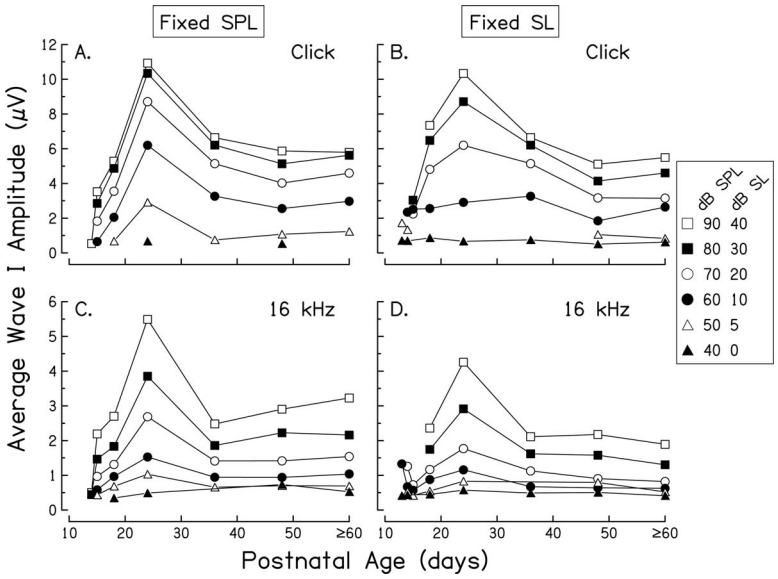


FIG. 8. Average ABR wave I amplitudes elicited by clicks [(a) and (b)] and 16-kHz tone bursts [(c) and (d)] increased during the active, rapid phase of development then decreased to adult levels near the end of the first postnatal month. Developmental changes in amplitudes are shown for stimuli presented at fixed sound pressure levels [SPL; panels (a) and (c)] and at fixed sensation levels [SL; panels (b) and (d)]. Amplitudes measured at P24 differed significantly from those acquired at older ages for the following stimulus conditions: click stimuli at 70–90 dB SPL and 30–40 dB SL; 16 kHz at 50–90 dB SPL and 10–40 dB SL ($P<0.05$).

SL conditions in adults. Consequently, analyses were based on matching conditions in which both the SL and the SPL stimuli produced approximately equal response latencies on the first postnatal day that responses were measured.

In Fig. 7, latency versus age curves representing these conditions for three stimuli are shown. These plots illustrate three significant points: (1) there are clear differences in latency maturation profiles for SPL and SL conditions, (2) differences are frequency dependent, and (3) the influence of threshold progressively diminishes during the final phase of cochlear differentiation, i.e., between P15 and P24.

Differences between SL and SPL curves are referred to in this study as *latency residue*, and that value can be taken as an estimate of the relative contribution made by threshold on response latency. Estimates of *latency residue* resulting from the SPL versus SL conditions were made by subtracting the average asymptotic response latency measured in response to 90 dB SPL stimuli from that determined using the appropriate SL condition and are represented by the hatched areas shown in Figs. 7(a)–7(c). Results of that exercise are shown across the responsive frequency range in Fig. 7(d). The resulting difference curve has the appearance of a bandpass filter with a center frequency of ~5.6 kHz and low- and high-frequency skirts that drop off at rates of approximately 0.4 and 0.2 ms/oct, respectively. Based on this analysis, it appears that elevated threshold accounted for very little, if any, of the latency prolongation observed during development for 32 kHz. A similar conclusion can be drawn regarding differences observed under low- (2.8 kHz) and other high-frequency stimulus conditions (16 and 22.6 kHz). The largest latency residue was observed in the mid-frequency band (~4 to 11.3 kHz). These findings suggest that between 28% and 62% of the total latency associated with responses recorded from immature animals can be accounted for by threshold-related factors such as the state of the endocochlear potential, the physical characteristics (e.g., stiffness, mass and damping power) of mechanical systems underlying transduction, as well as changes in ion channel kinetics and the extent of myelination.

E. Development of response amplitude

Although highly variable from day to day, it is clear that average wave I response amplitudes increased between P13 and P24 except, perhaps, in the case of near threshold conditions, as shown for both SPL and SL cases in Fig. 8. Average amplitude values subsequently declined, achieving adultlike values by ~P36. The rate of response amplitude growth was determined for the linear growth phase by calculating the slope of a best-fitting line to wave I amplitude versus age plots between P12 and P15 for each stimulus frequency studied, as shown in a family of curves in Fig. 9. Growth rate turned out to be directly correlated with stimulus frequency for the highest level condition studied (90 dB SPL), with values gradually increasing from <0.1 μV/day in the case of the lowest frequencies studied to ~0.45 μV/day at 11.3 kHz and even more gradually increasing to achieve values of >0.5 μV/day at 32 kHz. Ex-

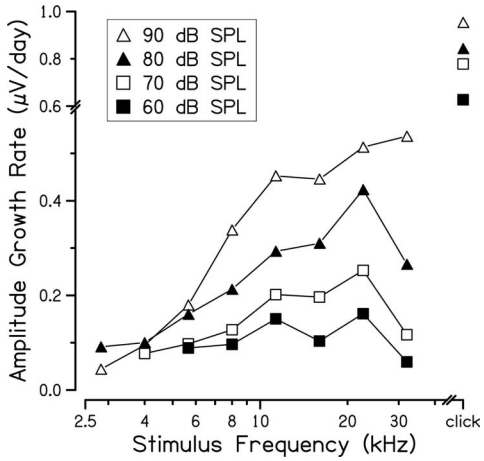


FIG. 9. The rate of wave I amplitude growth during development was computed from the linear growth phase (P12–P24) based on linear regression analysis and plotted as a function of stimulus frequency. Amplitude growth rates were determined using fixed sound pressure levels indicated in the symbol key. Growth rates are shown for click stimuli also. Note the change in the scale of the ordinate.

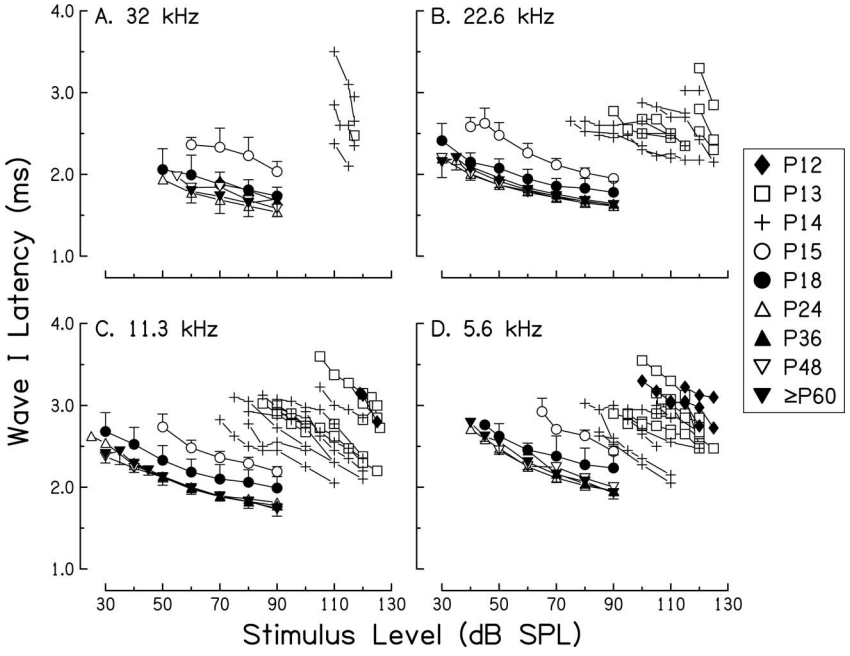


FIG. 10. Wave I latencies are plotted as a function of stimulus level for responses to 32 kHz (a), 22.6 kHz (b), 11.3 kHz (c), and 5.6 kHz (d) at the specified ages. Latency-level curves shown for P14 and younger are from individual animals and those shown from animals at older ages are averaged across individuals to increase the clarity of the figure. Standard deviations are shown for responses at P15, P18, and ≥P60. Based on statistical tests performed for groups ≥P15 in response to stimuli between 60 and 90 dB SPL [panel (a)], 50 and 90 dB SPL [panels (b) and (c)], and 70 and 90 dB SPL [panel (d)], latencies remained significantly elevated ($P < 0.05$) compared to adult values (≥P24) at P15 for all frequencies tested, and although still somewhat elevated at P18, latencies differed significantly from adults for the following conditions: 22 kHz at 50, 70, and 80 dB SPL; 11.3 kHz at 50, 60, 80, and 90 dB SPL.

cept for 32 kHz, the form of amplitude growth versus frequency curves representing lower stimulus levels were similar, although maximum values were predictably smaller than in the 90 dB SPL case.

F. Slopes of latency-intensity functions

The slopes of latency-intensity curves observed on P12, P13, and P14 were generally steeper than those observed in adults (Fig. 10). This observation was particularly evident in responses to high-frequency stimulation and was most notable in the case of 32 kHz in which slopes as steep as $-42 \log_{10}(\mu s)/\text{dB}$ were observed on P14, the earliest that responses could be elicited at this frequency. In contrast, the slopes of latency-level curves representing adults were in the range of $-5 \log_{10}(\mu s)/\text{dB}$. Differences between the slopes of latency-level curves representing neonates and adults were generally smaller than those observed in the 32-kHz case when responses to lower frequency stimuli were considered at comparable ages [Figs. 10(c) and 10(d)]. For example, average slope estimates at 5.6 kHz were approximately $-10 \log_{10}(\mu s)/\text{dB}$ on P13 and P14, twice as steep as those observed in adults.

While the 32-kHz case stands as a clear example of extreme immaturity in young BALB/c mice, the developmental trend associated with responses to 22.6 kHz is more typical of patterns observed in this investigation. On P13, one day after animals were first responsive to this frequency, latency-level curves segregated into two groups. In one group latency-level curves were relatively shallow, even adultlike, whereas slopes were considerably steeper than those observed in adults in the other group. It is also notable that latency-level curves exhibiting relatively shallow, adultlike slopes were generally associated with lower thresholds than those with steep slopes. On P14, although absolute latency estimates were more variable than in adults, latency-level curve slopes appeared qualitatively adultlike. The same observation was made in the case of 11.3 kHz, although slopes were more uniform on P13 than for 22.6 kHz; i.e., latency-level curves did not segregate into two distinct groups on the basis of slope and sensitivity.

This trend is illustrated in a family of curves depicting latency-level slopes plotted as a function of stimulus frequency with age as parameter in Fig. 11(a) [see representative examples of latency-level slope versus age plots in Fig. 11(b)]. Slopes of latency input-output (I/O) curves derived

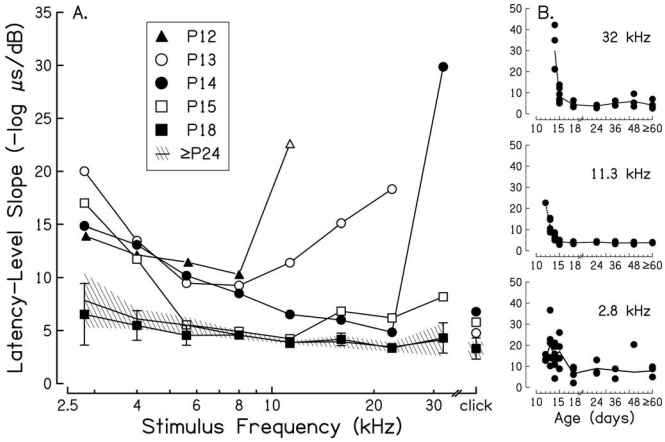


FIG. 11. (a) Changes in the slope of wave I latency versus level functions estimated throughout the period of rapid development mature in a frequency-specific manner, with mature relationships occurring earliest for mid-frequency stimuli, followed by high-frequency stimuli, and last for low-frequency stimuli. Slope estimates were based on linear regression analysis of normalized latency-level curves acquired from individuals. The hatched area above and below the solid line represents ±1 standard deviation of the mean slopes determined for the adult group (≥P24). Standard deviations are also shown for responses obtained at P18. The unfilled triangle represents a single measurement. Slopes of latency-level curves were significantly different ($P < 0.05$) from those of adults (≥P24) at P12 for 2.8–5.6 kHz; at P13 for 2.8–8 kHz; at P14 for 2.8–11.3 kHz, 32 kHz, and for click stimuli; and at P15 for 16–22.6 kHz. Slope estimates acquired at P18 did not differ significantly from those of adults. (b) Examples of slopes of latency-level curves from which average slope estimates were used in panel (a) are plotted as a function of age in the right column for three stimulus frequencies: 32, 11.3, and 2.8 kHz.

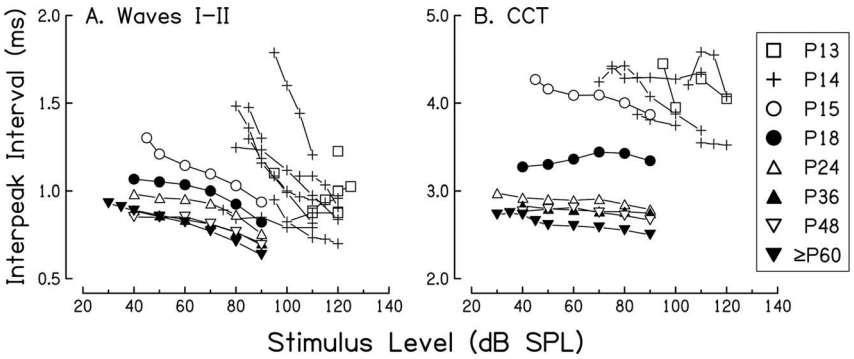


FIG. 12. Interpeak intervals between waves I and II (a) and central conduction times [CCT; panel (b)] are plotted as a function of stimulus level for responses to 11.3 kHz at the specified ages. CCT was estimated as the interval between waves I and IV. Individual curves are shown for animals at P13 and P14, and average curves are shown for animals at P15 and older. Based on statistical tests performed for animals ≥P15 in response to stimuli between 50 and 90 dB SPL, the wave I-II interval remained significantly elevated ($P<0.05$) compared to adult values (≥P36) for P15, P18, and P24 at 60–80 dB SPL, and for P15 at 50 and 90 dB SPL. CCTs were significantly elevated ($P<0.05$) compared to adult values (≥P60) for P15, P18, and P24 at 50–80 dB SPL and for P15 at 90 dB SPL.

from responses observed on P12, P13, and P14 and representing stimulus frequencies below 10 kHz were indistinguishable from one another and, as a group, they were approximately twice as steep as adult curves. Although insignificant from a statistical point of view, it is noteworthy that a solitary individual responded to intense tone bursts at more than one level at 11.3 kHz on P12 and the slope of the associated I/O curve (open triangle) was approximately six times steeper than in adults and two times steeper than the slopes of I/O curves representing lower stimulus frequencies (<8 kHz). This is especially interesting in light of the observation that the average latency-level slopes representing the 11.3–22.6-kHz band were also notably steeper than adult values just one day later, on P13, as was the average slope of 32-kHz I/O curves on P14, observations clearly suggesting that the primary developmental events occurring at this point in time are associated with high-frequency portions of the cochlear spiral. By P15, the slopes of I/O curves representing 5.6–11.3 kHz achieved maturity. Although latency-level slopes representing the lowest frequencies studied on P15 (2.8 and 4.0 kHz) were not statistically distinguishable from adult values, trends shown in Fig. 11 suggest that development may be delayed in this low-frequency band (2.8–4 kHz) relative to the mid-frequency band (5.6–11.3 kHz). The last maturational stage occurred between P15 and P18, generally, as latency-level slopes derived from responses to the highest frequency tone bursts studied acquired adultlike values.

It is interesting to note that although latency-level slopes were adultlike by P18, as were response thresholds for all stimulus conditions studied, absolute latencies remained prolonged regardless of level at this age (cf. Fig. 10), a finding strongly suggesting that the source of latency prolongation in immature animals is partially determined by factors other than threshold, including auditory nerve fiber (ANF) myelination, synapse maturation, and the formation of ion channels that affect the time course of the depolarization and repolarization of both ANFs and sensory cells.

G. Interpeak interval and intensity relationships

As shown in Fig. 12, developmental changes in interpeak interval (IPI) values were similar to those observed in the case of response latencies. However, unlike latency, on P13 measurement variability made it difficult, if not impossible, to reliably estimate input-output trends, particularly in the case of the interval between waves I and II. The overall tendency for interpeak-interval versus level functions to be steep in immature animals was more evident in data acquired on P14. By P15 IPI-level curves were similar to those observed in adults for both the interval between waves I and II and central conduction time (CCT, defined as the interval between waves I and IV). Maturation appears complete by P36.

H. Response amplitude and intensity relationships

The immaturity of the auditory system on P12, P13, and P14 and its relatively abrupt maturation by P15 is evident in an examination of a representative collection of wave I response amplitude versus stimulus level curves shown in Fig. 13. The extended immaturity of responses to 32-kHz tone bursts relative to other stimulus types is again evident in the case of response amplitude. As late as P14, only low-amplitude responses to high-level stimulation were observed and the range over which amplitudes increased with level was small relative to that observed in adults, in some cases reflecting the maximum output of the system.

Amplitude-level curves representing animals from P18 on were adultlike in form, consisting of a gradually sloping near-threshold segment and a steeper higher-level segment, a relationship that is referred to in this analysis as "complex." Between P12 and P15 a second class of amplitude versus level curves was observed. Curves falling into this category were composed entirely of high-level, near-threshold responses that were best represented with a single line segment and are referred to here as "simple." The slopes of simple input-output (I/O) curves were generally similar to the slopes of the high-level segment observed in complex, adultlike curves. It is interesting to note that amplitude versus level curves derived from responses to stimuli below 4 kHz were generally simple in form regardless of age.

To characterize the development of amplitude-level relationships, the slopes of near threshold I/O curve segments

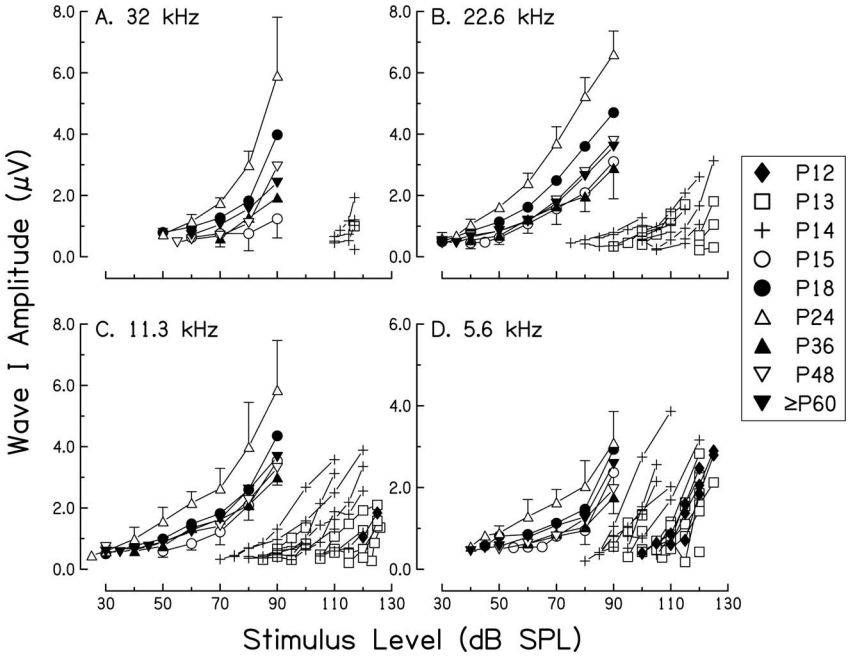


FIG. 13. Wave I amplitude values are plotted as a function of stimulus level for responses to 32 kHz (a), 22.6 kHz (b), 11.3 kHz (c), and 5.6 kHz (d) at the specified ages. Curves from individuals are shown for animals at P14 and younger, and average curves are shown for animals at P15 and older. Error bars are plotted for the responses at P15 and P24 and represent the standard deviation. Based on statistical tests performed for animals ⩾P15 in response to stimuli between 60 and 90 dB SPL [panel (a)], 40 and 90 dB SPL [panel (b) and (c)], and 50 and 90 dB SPL [panel (d)], amplitudes measured at P24 were elevated significantly ($P<0.05$) compared to adult values (⩾P36) for the following conditions: 32 kHz at 60–90 dB SPL, 22.6 kHz at 40–90 dB SPL, 11.3 kHz at 50–90 dB SPL, and 5.6 kHz at 50–70 dB SPL.

["slope 1" as depicted in Fig. 14(c)] were tracked during the period of development and compared with measurements made in adults, i.e., animals at P24 and older [see Fig. 14(b) for representative examples of I/O slope versus age plots]. Average slope values representing each age studied were analyzed as a function of stimulus frequency as shown in Fig. 14(a). As expected based on the casual observation of amplitude-level curves from individual animals, slopes were very steep on P12, even exceeding values observed for high-level I/O segments among adults (the upper cross-hatched area representing "slope 2" in Fig. 14) under some stimulus conditions. On P13, although I/O curves were shallower than

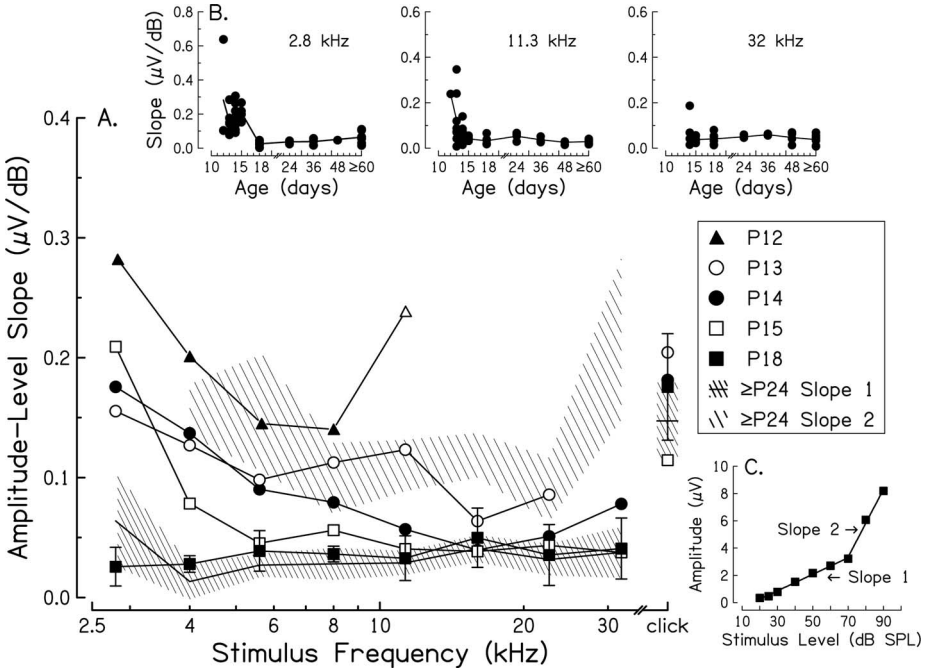


FIG. 14. (a) Changes in the slope of wave I amplitude versus level functions estimated throughout the period of rapid development occur in a frequency-specific manner, with mature relationships occurring earlier for high-frequency stimuli and last for low-frequency stimuli. Slope estimates were based on linear regression analysis of the near threshold portion of amplitude-level curves. Hatched areas above and below the solid line represent ±1 standard deviation of the mean slopes (slope 1) determined for the adult group (⩾P24). For comparison, the slope of the upper leg of the amplitude-level curve (slope 2) was computed for adults (⩾P24) and is also shown (±1 sd). Error bars are plotted for responses at P18 and represent standard deviations. The open triangle (P12) represents a single measurement. Slopes of amplitude-level curves were significantly different ($P<0.05$) from those of adults (⩾P24) at P13 and P14 for 2.8–8 kHz and at P15 for 2.8–4 kHz. Slope estimates acquired at P18 did not differ significantly from those of adults. (b) Examples of amplitude-level slope versus age plots from which average slope estimates were used in this analysis are shown at the top for three stimulus frequencies: 2.8, 11.3, and 32 kHz. (c) An example of an adult amplitude-level curve indicating the lower (near threshold, slope 1) and upper (slope 2) legs is shown in the lower right panel, along with a line fitted to each segment.

on P12, they remained as steep, or nearly as steep as slope 2 values observed in adults. On P14, slope reductions were generally restricted to responses in the 8–22.6-kHz band, and by P15 input-output relations were generally adultlike, with the exception of responses elicited by tone bursts ≤4 kHz. I/O curves representing that portion of the audiogram remained decidedly steep relative to adult counterparts but acquired adultlike stature by P18.

IV. DISCUSSION

A. General comments

The development of auditory function among vertebrates follows a relatively clear pattern that reflects the final stage of cochlear differentiation during which peripheral auditory function is acquired and matures as the elements of both passive and active transduction undergo significant transformation (Rubel, 1978; Walsh and Romand, 1992). As a consequence, the system itself is transformed from an immature state characterized by high thresholds, limited audiometric range, and an inability to detect frequency-specific acoustic events, to the mature state that is 100 000 times more sensitive than it is at the time of birth and is capable of resolving frequency differences as small as a single Hz. Based on findings reported here it appears that the BALB/c mouse conforms to this generally accepted rule of maturation.

Understanding the maturational forces that shape the development of function in mammals is complicated by the fact that the sensory organ, including nonsensory components and connections with the brain, undergoes simultaneous and continuous transformation during a relatively brief period in developmental time. Parsing the changes that contribute to the development of specific functional attributes is consequently difficult and is considered in the discussion that follows.

B. Sensitivity

Initially high thresholds to acoustic stimulation improve rapidly and achieve near-maturity over the course of several days in all altricial mammals thus far studied. In domestic cats, for example, responses to airborne sounds are rarely observed prior to the end of the first postnatal week regardless of stimulus frequency and mature responses are not observed until approximately the end of the third postnatal week of life, following a period of rapid then relatively slow maturation (Pujol and Marty, 1970; Carlier *et al.*, 1975; Romand, 1979; Rubel, 1984; Dolan *et al.*, 1985; Walsh *et al.*, 1986a; Walsh and McGee, 1990; Walsh and Romand, 1992). In the case of gerbils, ABRs are first observed around P12 and response thresholds mature by P18, again depending on stimulus frequency and again following sequential rapid and gradual developmental periods (Woolf and Ryan, 1984; Smith and Kraus, 1987; McGuirt *et al.*, 1995; McFadden *et al.*, 1996). The same basic developmental rule applies to rats (Jewett and Romano, 1972; Iwasa and Potsic, 1982; Elatchley *et al.*, 1987) and other mammals thus far studied (Stein *et al.*, 1981; Doyle *et al.*, 1983; Pettigrew and Morey, 1987; Schweitzer, 1987), including other strains of mice (Shnerson and Pujol, 1981; Steel and Harvey, 1992). Based on these comparisons, the time line over which BALB/c mice acquire acoustic sensitivity is predictably similar to what has been observed in other strains of mice and other mammals generally.

C. Influence of the external ear

It is almost certainly the case that multiple factors contribute to the development of acoustic sensitivity in mammals. In the case of the immature external ear, for example, although it does not appear to play an integral role in the development of transduction *per se*, it has been shown to attenuate acoustic input and consequently obscure otherwise clear responses to high-level, low-frequency tone bursts in neonatal cats and gerbils (Villablanca and Olmstead, 1979; Walsh *et al.*, 1986a; Woolf and Ryan, 1988; McFadden *et al.*, 1996). While studies of the external ear's influence on development have not been reported for mice, it is reasonably safe to assume that the immature structure attenuates the transfer of acoustic energy to the middle ear and thereby diminishes acoustic sensitivity in much the same way it does in other mammals. Because the external ear canal appears, at least on qualitative grounds, to be patent in BALB/c mice by P12 (Sprenkle *et al.*, 2001), the influence of external ear immaturity is most likely limited in scope and restricted to the earliest period of functional development.

D. Influence of the middle ear

With the notable exceptions of tympanic membrane area, the volume of the auditory bulla, and ossicular mass, most middle ear structures are adult or nearly so by the time the first responses to airborne sounds are observed in altricial mammals studied thus far (Stephens, 1972; Huangfu and Saunders, 1983; Thomas and Walsh, 1990; Cohen *et al.*, 1993). Even though relatively mature anatomically, the ability of the middle ear to transfer acoustic energy to the cochlea remains immature beyond birth, extending well into the second postnatal month in some species like the hamster and the rat (Relkin *et al.*, 1979; Relkin and Saunders, 1980; Doan *et al.*, 1994, 1996). In other cases, represented most commonly by the gerbil, certain aspects of middle ear function appear to mature as early as the end of the second postnatal week (Cohen *et al.*, 1993), although maturity is achieved most frequently during the third postnatal week (Woolf and Ryan, 1984; Overstreet and Ruggero, 2002). Middle ear maturation in the rat is especially interesting in that responses to both high- (32 kHz) and low- (1 kHz) frequency stimulation are notably delayed (i.e., 50 and 68 postnatal days, respectively) relative to mid-frequency responses that achieve maturity during the third and fourth postnatal weeks (Doan *et al.*, 1996).

Based on findings made by Doan *et al.* (1994), middle ear development in BALB/c mice follows the same basic developmental pattern described in other mammals and, as such, middle ear immaturities appear to affect sensitivity to airborne sounds throughout the period of development. Estimates of the magnitude of the immature middle ear's influence on response threshold have not been studied extensively

and are variable from lab to lab, but range from ~10 to 25 dB (Woolf and Ryan, 1988; Cohen *et al.*, 1993; Overstreet and Ruggero, 2002). Increased tympanic membrane compliance, as well as decreased middle ear frictional resistance and inertia of the tympanic membrane and ossicular chain, along with the enlargement of bullar volume, may all have a role in the development of middle ear function (Relkin *et al.*, 1979; Relkin and Saunders, 1980; Doan *et al.*, 1994, 1996; Overstreet and Ruggero, 2002). Although it is apparent that the middle ear plays a role during development, the need to more thoroughly characterize the process in BALB/c mice and other mammals is nonetheless clear.

E. Influence of the endocochlear potential

While changes in the functional status of the middle ear occur throughout development of the auditory periphery and determine the effective bandwidth and level of acoustic energy reaching the cochlea, thereby affecting acoustic sensitivity, changes associated with intracochlear differentiation that occur during the same period are equally, if not more noteworthy. One well-recognized determinant of acoustic sensitivity is the endocochlear potential (EP). Given the relationship between the EP and sensitivity in adult animals, it has been widely noted that the EP grows monotonically during development, generally following the form of a sigmoid, and the rise in EP observed among altricial mammals has long been recognized as a major factor affecting the maturation of acoustic sensitivity. As long ago as 1971, Bosher and Warren (1971) suggested that the rise in EP during development played a key role in the acquisition of acoustic sensitivity and these authors, along with Fernandez and Hinojosa (1974), further suggested that the increase in the electrical resistance of cell membranes bounding the endolymphatic space was essential, and perhaps sufficient, to affect the maturation of the EP. In support of this view, expression of the inwardly rectifying K^+ channel, Kir4.1, in stria vascularis marginal cells increases and is correlated with the rise of the EP during development (Hibino *et al.*, 1997). In addition, Souter and Forge (1998) provide direct support for the suggestion raised by Kikuchi and Hilding (1966) and Fernandez and Hinojosa (1974) that the development of EP is at least partially dependent on an increase in the electrical resistance of Claudius and Boettcher cell membranes that form tight junctions among cells forming the boundary of the endolymphatic space. Kuijpers (1974) also noted that $Na-{}^+K^+$-adenosine triphosphatase activity increased in the stria vascularis during the period that the EP acquires maturity, suggesting the possibility that multiple, overlapping developmental events combine to produce adultlike cochlear electroanatomy.

It is difficult to assess the day by day contribution that EP maturation makes in the development of acoustic sensitivity among various mouse strains for at least two reasons. First, differences in the maturational rate whereby the EP achieves maturity are substantial. For example, in the case of the IRC strain, Sadanaga and Morimitsu (1995) tracked EP development and found that the positive potential matured along a sigmoidal trajectory, as in other mammals, and achieved maturity on P20. If the EP followed the same maturational profile in BALB/c mice, EP development would account for much of the process whereby auditory thresholds are achieved developmentally, a highly unlikely possibility. On the other hand, if the time course of EP development in BALB/c mice is the same as that shown for $+/dn$ mice (Bock and Steel, 1983), EP maturation would account for only a portion of the nearly 100 dB over which thresholds improved (cf. Fig. 4). The second and most significant difficulty is associated with the assumption that an EP change of 1 mV results in a change of 1 dB in sensitivity (Sewell, 1984) in developing animals. This assumption is almost certainly invalid given the immature character of the organ of Corti during much of the period that EP is on the rise. Clearly, therefore, the importance of determining the time line over which BALB/c mice acquire mature EPs and the relative contribution that EP makes to threshold values (i.e., how does the voltage-to-dB conversion factor change as the organ of Corti matures) is underlined given the complicated nature of this relationship.

F. Influence of passive versus active cochlear mechanics

It is generally understood that the magnitude, frequency specificity, and sensitivity of basilar membrane vibration is governed by both passive and active transduction events [for a review see Robles and Ruggero (2001)]. The passive mechanical properties of the basilar membrane support the propagation of a broadly tuned "traveling wave" that was described and characterized by von Békésy (1960). In addition to passive transduction, an energy-consuming active process associated with electromechanical transduction is the source of a compressive nonlinearity that is dependent on the functional integrity of outer hair cells (OHCs) that contract and expand during depolarization and hyperpolarization (Rhode, 1971; Brownell *et al.*, 1985; Robles *et al.*, 1986; Ashmore, 1987). Changes in the length of OHCs are thought to produce intracochlear forces that feed back onto the vibrating organ of Corti and amplify its vibration magnitude, producing sharp tuning and high sensitivity to airborne sounds as a consequence.

Although it is clear that the development of both passive and active aspects of electromechanical transduction contribute to the functional maturation of the auditory periphery, it is less clear which process dominates maturational events at discrete developmental points in time. It is generally held that the process of transduction is essentially linear during the earliest stages of functional development, suggesting that the maturation of passive transduction elements dominate the early phase, and that the compressive nonlinearity observed in mature mammals is acquired later in development, producing the commonly noted triad of functional correlates, sharp tuning, high sensitivity, and broad dynamic range. Direct support for this view can be found in Fitzakerley *et al.* (1994a, b, c) and Tubach *et al.* (1996), who saw no evidence of two-tone suppression or cubic distortion product generation, respectively, in the discharge patterns of auditory nerve fibers recorded from neonatal cats. Their findings are consistent with the outcomes of distortion product otoacoustic

emission (DPOAE) studies (Lenoir and Puel, 1987; Henley *et al.*, 1990; Norton *et al.*, 1991; Mills and Rubel, 1996), as well as the observation that auditory nerve fiber tuning curves from neonatal animals are tipless and very broadly tuned (Dolan *et al.*, 1985; Walsh and McGee, 1987; Echteler *et al.*, 1989; Walsh and McGee, 1990; Muller, 1996; Tubach *et al.*, 1996). These findings collectively support a "sequential model" in which the maturation of passive influences dominate early stages of functional development and active processes emerge later in postnatal life.

Recently, however, Overstreet *et al.* (2002) suggested an alternate view. Based on basilar membrane measurements, Overstreet and colleagues suggested that it is the development of passive aspects of transduction that limit the expression of active mechanics in the extreme basal turn of the cochlear spiral, pointing to a shift in the best frequency of the recording site during development and the apparent inability of the extreme base to support high-frequency traveling waves as supporting evidence. The findings of He *et al.* (1994) address this question directly and suggest that OHCs harvested from locations along the length of the cochlear spiral of neonatal gerbils and studied *in vitro* are electromotile by P12 and achieve adultlike electromotility by approximately the 13th or 14th postnatal day, an age that precedes the age that adultlike sensitivity is reported here. Leaving open the possibility that transduction in the extreme basal turn of the cochlea is governed by a different set of properties than those operating in more apical regions, the findings of He and colleagues suggest that OHC motility, *per se*, and by inference the development of active mechanics, is not a rate-limiting developmental event, generally.

Interestingly, one can turn to developmental changes that occur in the slopes of latency versus level curves representing hypothyroid mice for additional insight into the answer to this complex question. Initially steep wave I latency versus intensity curves are known to flatten out during development in hypothyroid mice as they do in normal mice, approaching, but never completely achieving, normal adult values (Song *et al.*, 2002). Because active transduction (i.e., cochlear amplification) fails to develop normally in hypothyroid mice (Song *et al.*, 2006), one can reasonably argue that the maturation of latency-level curve slopes is a reflection of the maturation of passive transduction mechanics. Because latency-level curves acquire mature-appearing slopes on approximately the 18th postnatal day, roughly the same time that thresholds achieve maturity, findings from this investigation support a model in which passive and active aspects of mechanoelectrical transduction develop simultaneously.

G. Response latency

As with mammals generally (Jewett and Romano, 1972; Pujol and Hilding, 1973; Walsh *et al.*, 1986b; Schweitzer, 1987; Smith and Kraus, 1987; Blatchley *et al.*, 1987; Eggermont and Salamy, 1988; Walsh *et al.*, 1992; Griffiths *et al.*, 1996), response latencies decline during the period of rapid development in BALB/c mice and it is reasonable to suggest that prolonged latencies associated with immature animals reflect the combined immaturity of neural, as well as passive and active processes associated with cochlear transduction.

Although it is impossible to isolate and characterize each of the factors that contribute to prolonged response times in immature animals on the basis of evoked potential findings, it is clear based on findings reported here that elevated thresholds account for what appears to be the majority of the wave I latency prolongation in excess of that observed in adults under some stimulus conditions studied (i.e., the mid-frequency range) (cf. Fig. 7). Acknowledging the contribution that elevated thresholds make in the prolongation of response latency, at least three additional sources must be considered in those cases in which adult response latency is determined by multiple factors: myelination, synapse development, and the maturation of membrane conductance channels that determine the time constant of receptor potentials and action potentials.

By comparing cochlear microphonic response latencies with the latency of ABR wave I in immature cats, Walsh and McGee (1991) showed that cohlear immaturities account for as much as 70% of the prolonged response times observed in immature animals. In this study, the contributions of developmental factors affecting threshold and nonthreshold mechanisms underlying latency prolongation were assessed by determining latency differences under SPL and SL conditions. The difference, referred to as latency residue, is taken as an indicator of the relative contribution of threshold influences on latency prolongation. As can be seen in Fig. 7, latency residue was considerable under some stimulus conditions, most notably those in the mid-frequency band, indicating that threshold contributions to absolute response latency were significant under those stimulation conditions. At lower and higher frequencies, non-threshold-related factors dominated the contribution to developmental changes in latencies. For example, at 16 kHz less than 0.2 ms of the developmental change in latency could be credited to threshold, while non-threshold-related factors contributed to the remainder, approximately 80% of the overall latency reduction.

While the relative contribution of all factors affecting latency prolongation cannot be estimated at the present time, Kros *et al.* (1998) have shown that inner hair cells express a large, fast potassium conductance during the postnatal period, the consequence of which is to shorten membrane time constants and decrease response latencies. The influences of other molecular and mechanical changes that occur during development remain unknown at this time.

H. Response amplitude

As shown with other mammals, the amplitude of ABR waves representing the BALB/c mouse rise to a peak, maximum value at about the time that adultlike sensitivity is achieved, only to drop thereafter and stabilize at adult values (Iwasa and Potsic, 1982; Walsh *et al.*, 1986c). The importance of the maturation of the endocochlear potential, and thereby response sensitivity, as a factor influencing the maturation of response amplitude is suggested by the similarity of their respective developmental time courses. In addition to the role that EP maturation plays as a factor affecting re-

sponse amplitude development, the opening of the external ear canal, improved transmission through the middle ear, and myelination most likely play roles as well. Although the developmental "hump" frequently observed early in the postnatal life of altricial mammals (Iwasa and Potsic, 1982; Walsh *et al.*, 1986c) is poorly understood generally, in the case of ABR amplitudes it may reflect overall brain growth, ossification of the skull, and changes in other elements of electrical resistance that lie between generators and the surface of the head, as might be expected as a consequence of the myelination of central pathways.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of Deafness and Other Communication Disorders. (NIH) Grant Nos. DC00982, DC04566, and DC04662. Portions of this work were presented at the 25th meeting of the Association for Research in Otolaryngology (Song *et al.*, 2002). The authors would also like to thank two anonymous reviewers for their thoughtful and comprehensive review of the manuscript.

Ashmore, J. F. (**1987**). "A fast motile response in guinea-pig outer hair cells: the cellular basis of the cochlear amplifier," J. Physiol. (London) **388**, 323–347.

Blatchley, B. J., Cooper, W. A., and Coleman, J. R. (**1987**). "Development of auditory brainstem response to tone pip stimuli in the rat," Brain Res. **429**, 75–84.

Bock, G. R., and Steel, K. P. (**1983**). "Inner ear pathology in the deafness mutant mouse," Acta Oto-Laryngol. **96**, 39–47.

Bosher, S. K., and Warren, R. L. (**1971**). "A study of the electrochemistry and osmotic relationships of the cochlear fluids in the neonatal rat at the time of the development of the endocochlear potential," J. Physiol. (London) **212**, 739–761.

Brownell, W. E., Bader, C. R., Bertrand, D., and de Ribaupierre, Y. (**1985**). "Evoked mechanical responses of isolated cochlear outer hair cells," Science **227**, 194–196.

Carlier, E., Abonnenc, M., and Pujol, R. (**1975**). "Maturation of single auditory-nerve fibers responses to tone-burst stimulation in kittens," J. Physiol. Paris **70**, 129–138.

Chen, Z. Y., and Corey, D. P. (**2002**). "Understanding inner ear development with gene expression profiling," J. Neurobiol. **53**, 276–285.

Cohen, Y. E., Doan, D. E., Rubin, D. M., and Saunders, J. C. (**1993**). "Middle-ear development. V: Development of umbo sensitivity in the gerbil," Am. J. Otolaryngol. **14**, 191–198.

Doan, D. E., Cohen, Y. E., and Saunders, J. C. (**1994**). "Middle-ear development. IV. Umbo motion in neonatal mice," J. Comp. Physiol. **174**, 103–110.

Doan, D. E., Igic, P. G., and Saunders, J. C. (**1996**). "Middle-ear development. VII: Umbo velocity in the neonatal rat," J. Acoust. Soc. Am. **99**, 1566–1572.

Dolan, D. F., Teas, D. C., and Walton, J. P. (**1985**). "Postnatal development of physiological responses in auditory nerve fibers," J. Acoust. Soc. Am. **78**, 544–554.

Doyle, W. J., Saad, M. M., and Fria, T. J. (**1983**). "Maturation of the auditory brain stem response in rhesus monkeys (*Macaca mulatta*)," Electroencephalogr. Clin. Neurophysiol. **56**, 210–223.

Duan, M. L., Ulfendahl, M., Laurell, G., Counter, A. S., Pyykko, I., Borg, E., and Rosenhall, U. (**2002**). "Protection and treatment of sensorineural hearing disorders caused by exogeneous factors: experimental findings and potential clinical application," Hear. Res. **169**, 169–178.

Echteler, S. M., Arjmand, E., and Dallos, P. (**1989**). "Developmental alterations in the frequency map of the mammalian cochlea," Nature (London) **341**, 147–149.

Eggermont, J. J., and Salamy, A. (**1988**). "Development of ABR parameters in a preterm and a term born population," Ear Hear. **9**, 283–289.

Fernandez, C., and Hinojosa, R. (**1974**). "Postnatal development of endocochlear potential and stria vascularis in the cat," Acta Oto-Laryngol. **78**, 173–186.

Fitzakerley, J. L., McGee, J., and Walsh, E. J. (**1994a**). "Responses of peripheral auditory neurons to two-tone stimuli during development. I. Correlation with frequency-selectivity," Hear. Res. **77**, 135–149.

Fitzakerley, J. L., McGee, J., and Walsh, E. J. (**1994b**). "Responses of peripheral auditory neurons to two-tone stimuli during development. II. Factors related to neural responsiveness," Hear. Res. **77**, 150–161.

Fitzakerley, J. L., McGee, J., and Walsh, E. J. (**1994c**). "Responses of peripheral auditory neurons to two-tone stimuli during development. III. Rate facilitation," Hear. Res. **77**, 162–167.

Griffiths, S. K., Pierson, L. L., Gerhardt, K. J., Abrams, R. M., and Peters, A. J. (**1996**). "Auditory brainstem response in sheep. Part II: Postnatal development," Dev. Psychobiol. **29**, 53–68.

He, D. Z., Evans, B. N., and Dallos, P. (**1994**). "First appearance and development of electromotility in neonatal gerbil outer hair cells," Hear. Res. **78**, 77–90.

Henley, C. M., III, Owings, M. H., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (**1990**). "Postnatal development of 2f1-f2 otoacoustic emissions in pigmented rat," Hear. Res. **43**, 141–148.

Hibino, H., Horio, Y., Inanobe, A., Doi, K., Ito, M., Yamada, M., Gotow, T., Uchiyama, Y., Kawamura, M., Kubo, T., and Kurachi, Y. (**1997**). "An ATP-dependent inwardly rectifying potassium channel, KAB-2 (Kir4.1), in cochlear stria vascularis of inner ear: Its specific subcellular localization and correlation with the formation of endocochlear potential," J. Neurosci. **17**, 4711–4721.

Holley, M. C. (**2002**). "Application of new biological approaches to stimulate sensory repair and protection," Br. Med. Bull. **63**, 157–169.

Huangfu, M., and Saunders, J. C. (**1983**). "Auditory development in the mouse: structural maturation of the middle ear," J. Morphol. **176**, 249–259.

Iwasa, H., and Potsic, W. P. (**1982**). "Maturational change of early, middle, and late components of the auditory evoked responses in rats," Otolaryngol.-Head Neck Surg. **90**, 95–102.

Izumikawa, M., Minoda, R., Kawamoto, K., Abrashkin, K. A., Swiderski, D. L., Dolan, D. F., Brough, D. E., and Raphael, Y. (**2005**). "Auditory hair cell replacement and hearing improvement by Atohl gene therapy in deaf mammals," Nat. Med. **11**, 271–276.

Jewett, D. L., and Romano, M. N. (**1972**). "Neonatal development of auditory system potentials averaged from the scalp of rat and cat," Brain Res. **36**, 101–115.

Kikuchi, K., and Hilding, D. A. (**1966**). "The development of the stria vascularis in the mouse," Acta Oto-Laryngol. **62**, 277–291.

Kraus, H. J., and Aulbach-Kraus, K. (**1981**). "Morphological changes in the cochlea of the mouse after the onset of hearing," Hear. Res. **4**, 89–102.

Kros, C. J., Ruppersberg, J. P., and Rusch, A. (**1998**). "Expression of a potassium current in inner hair cells during development of hearing in mice," Nature (London) **394**, 281–284.

Kuhn, B., and Vater, M. (**1997**). "The postnatal development of F-actin in tension fibroblasts of the spiral ligament of the gerbil cochlea," Hear. Res. **108**, 180–190.

Kuijpers, W. (**1974**). "Na–K-ATPase activity in the cochlea of the rat during development," Acta Oto-Laryngol. **78**, 341–344.

Lenoir, M., and Puel, J. L. (**1987**). "Development of 2f1-f2 otoacoustic emissions in the rat," Hear. Res. **29**, 265–271.

Lenoir, M., Shnerson, A., and Pujol, R. (**1980**). "Cochlear receptor development in the rat with emphasis on synaptogenesis," Anat. Embryol. (Berl) **160**, 253–262.

Li, H. W., Roblin, G., Liu, H., and Heller, S. (**2003**). "Generation of hair cells by stepwise differentiation of embryonic stem cells," Proc. Natl. Acad. Sci. U.S.A. **100**, 13495–13500.

McFadden, S. L., Walsh, E. J., and McGee, J. (**1996**). "Onset and development of auditory brainstem responses in the Mongolian gerbil (*Meriones unguiculatus*)," Hear. Res. **100**, 68–79.

McGee, J., and Walsh, E. J. (**2004**). "Cochlear transduction and the molecular basis of peripheral auditory pathology," *Cummings Otolaryngology: Head and Neck Surgery*, edited by C. Cummings, P. W. Flint, L. A. Harker, B. H. Haughey, M. A. Richardson, K. T. Robbins, D. E. Schuller, and J. R. Thomas (Elsevier, Philadelphia), pp. 3402–3465.

McGuirt, J. P., Schmiedt, R. A., and Schulte, B. A. (**1995**). "Development of cochlear potentials in the neonatal gerbil," Hear. Res. **84**, 52–60.

Mills, D. M., and Rubel, E. W. (**1996**). "Development of the cochlear amplifier," J. Acoust. Soc. Am. **100**, 428–441.

Muller, M. (**1996**). "The cochlear place-frequency map of the adult and developing mongolian gerbil," Hear. Res. **94**, 148–156.

Muller, U., and Littlewood-Evans, A. (**2001**). "Mechanisms that regulate mechanosensory hair cell differentiation," Trends Cell Biol. **11**, 334–342.
Norton, S. J., Bargones, J. Y., and Rubel, E. W. (**1991**). "Development of otoacoustic emissions in gerbil: Evidence for micromechanical changes underlying development of the place code," Hear. Res. **51**, 73–92.
Overstreet, E. H., III, and Ruggero, M. A. (**2002**). "Development of wide-band middle ear transmission in the Mongolian gerbil," J. Acoust. Soc. Am. **111**, 261–270.
Overstreet, E. H., III, Temchin, A. N., and Ruggero, M. A. (**2002**). "Passive basilar membrane vibrations in gerbil neonates: mechanical bases of cochlear maturation," J. Physiol. (London) **545**, 279–288.
Petit, C., Levilliers, J., and Hardelin, J. P. (**2001**). "Molecular genetics of hearing loss," Annu. Rev. Genet. **35**, 589–646.
Pettigrew, A. G., and Morey, A. L. (**1987**). "Changes in the brainstem auditory evoked response of the rabbit during the first postnatal month," Brain Res. **430**, 267–276.
Pujol, R., and Abonnenc, M. (**1977**). "Receptor maturation and synaptogenesis in the golden hamster cochlea," Arch. Oto-Rhino-Laryngol. **217**, 1–12.
Pujol, R., and Hilding, D. (**1973**). "Anatomy and physiology of the onset of auditory function," Acta Oto-Laryngol. **76**, 1–10.
Pujol, R., and Marty, R. (**1970**). "Postnatal maturation in the cochlea of the cat," J. Comp. Neurol. **139**, 115–118.
Pujol, R., Carlier, E., and Devigne, C. (**1979**). "Significance of presynaptic formations in early stages of cochlear synaptogenesis," Neurosci. Lett. **15**, 97–102.
Pujol, R., Carlier, E., and Lenoir, M. (**1980**). "Ontogenetic approach to inner and outer hair cell function," Hear. Res. **2**, 423–430.
Relkin, E. M., and Saunders, J. C. (**1980**). "Displacement of the malleus in neonatal golden hamsters," Acta Oto-Laryngol. **90**, 6–15.
Relkin, E. M., Saunders, J. C., and Konkle, D. F. (**1979**). "The development of middle-ear admittance in the hamster," J. Acoust. Soc. Am. **66**, 133–139.
Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mossbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.
Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.
Robles, L., Ruggero, M. A., and Rich, N. C. (**1986**). "Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases," J. Acoust. Soc. Am. **80**, 1364–1374.
Romand, R. (**1979**). "Development of auditory-nerve activity in kittens," Brain Res. **173**, 554–556.
Romand, R., Sans, A., Romand, M. R., and Marty, R. (**1976**). "The structural maturation of the stato-acoustic nerve in the cat," J. Comp. Neurol. **170**, 1–15.
Roth, B., and Bruns, V. (**1992**). "Postnatal development of the rat organ of Corti. I. General morphology, basilar membrane, tectorial membrane and border cells," Anat. Embryol. (Berl) **185**, 559–569.
Rubel, E. W. (**1978**). "Ontogeny of structure and function in the vertebrate auditory system," *Development of Sensory Physiology*, edited by M. Jacobson (Springer-Verlag, New York), pp. 135–238.
Rubel, E. W. (**1984**). "Ontogeny of auditory system function," Annu. Rev. Physiol. **46**, 213–229.
Rybak, L. P., Whitworth, C., and Scott, V. (**1992**). "Development of endocochlear potential and compound action potential in the rat," Hear. Res. **59**, 189–194.
Sadanaga, M., and Morimitsu, T. (**1995**). "Development of endocochlear potential and its negative component in mouse cochlea," Hear. Res. **89**, 155–161.
Sato, M., Leake, P. A., and Hradek, G. T. (**1999**). "Postnatal development of the organ of Corti in cats: a light microscopic morphometric study," Hear. Res. **127**, 1–13.
Schweitzer, L. (**1987**). "Development of brainstem auditory evoked responses in the hamster," Hear. Res. **25**, 249–255.
Schweitzer, L., Lutz, C., Hobbs, M., and Weaver, S. P. (**1996**). "Anatomical correlates of the passive properties underlying the developmental shift in the frequency map of the mammalian cochlea," Hear. Res. **97**, 84–94.
Sewell, W. F. (**1984**). "The effects of furosemide on the endocochlear potential and auditory-nerve fiber tuning curves in cats," Hear. Res. **14**, 305–314.
Shnerson, A., and Pujol, R. (**1981**). "Age-related changes in the C57BL/6J mouse cochlea. I. Physiological findings," Brain Res. **254**, 65–75.
Simmons, D. D., Mansdorf, N. B., and Kim, J. H. (**1996**). "Olivocochlear innervation of inner and outer hair cells during postnatal maturation: evidence for a waiting period," J. Comp. Neurol. **370**, 551–562.
Smith, D. I., and Kraus, N. (**1987**). "Postnatal development of the auditory brainstem response (ABR) in the unanesthetized gerbil," Hear. Res. **27**, 157–164.
Song, L., McGee, J., and Walsh, E. J. (**2002**). "Development of auditory brainstem responses (ABRs) in normal and hypothyroid (*Tshr* mutant) mice," Assoc. Res. Otolaryngol. Abstr. **25**, 167.
Song, L., McGee, J., and Walsh, E. J. (**2006**). "Effects of profound congenital hypothyroidism on the development of auditory function in $Tshr^{hyt}$ mutant mice," Brain Res. (submitted).
Souter, M., and Forge, A. (**1998**). "Intercellular junctional maturation in the stria vascularis: possible association with onset and rise of endocochlear potential," Hear. Res. **119**, 81–95.
Souter, M., Nevill, G., and Forge, A. (**1997**). "Postnatal maturation of the organ of Corti in gerbils: morphology and physiological responses," J. Comp. Neurol. **386**, 635–651.
Sprenkle, P. M., McGee, J., Bertoni, J. M., and Walsh, E. J. (**2001**). "Development of auditory brainstem responses (ABRs) in *Tshr* mutant mice derived from euthyroid and hypothyroid dams," J. Assoc. Res. Otolaryngol. **2**, 330–347.
Steel, K. P., and Harvey, D. (**1992**). "Development of auditory function in mutant mice," *Development of Auditory and Vestibular Systems 2* (Elsevier, Amsterdam), Chap. 7, pp. 221–242.
Steel, K. P., and Kros, C. J. (**2001**). "A genetic approach to understanding auditory function," Nat. Genet. **27**, 143–149.
Stein, L. K., Ozdamar, O., and Schnabel, M. (**1981**). "Auditory brain-stem responses (ABR) with suspected deaf-blind children," Ear Hear. **2**, 30–40.
Stephens, C. B. (**1972**). "Development of the middle and inner ear in the golden hamster (*Mesocricetus auratus*)," Acta Oto-Laryngol., Suppl. **296**, 1–51.
Thomas, J. P., and Walsh, E. J. (**1990**). "Postnatal development of the middle ear: Areal ratios in kittens," Otolaryngol.-Head Neck Surg. **103**, 427–435.
Torres, M., and Giraldez, F. (**1998**). "The development of the vertebrate inner ear," Mech. Dev. **71**, 5–21.
Tubach, M., McGee, J., and Walsh, E. J. (**1996**). "Distortion generated by the ear: Its emergence and evolution during development," Laryngoscope **106**, 822–830.
Van Camp, G., and Smith, R. (2005). "Hereditary hearing loss homepage," URL: http://webhost.ua.ac.be/hhh/.
Villablanca, J. R., and Olmstead, C. E. (**1979**). "Neurological development of kittens," Dev. Psychobiol. **12**, 101–127.
von Békésy, G. (**1960**) *Experiments in Hearing* (McGraw-Hill, New York).
Walsh, E. J., and McGee, J. (**1986**). "The development of function in the auditory periphery," *Neurobiology of Hearing: The Cochlea*, edited by R. A. Altschuler, R. Bobbin, and D. Hoffman (Raven, New York), Chap. 14, pp. 247–269.
Walsh, E. J., and McGee, J. (**1987**). "Postnatal development of auditory nerve and cochlear nucleus neuronal responses in kittens," Hear. Res. **28**, 97–116.
Walsh, E. J., and McGee, J. (**1990**). "Frequency selectivity in the auditory periphery: similarities between damaged and developing ears," Am. J. Otolaryngol. **11**, 23–32.
Walsh, E. J., and McGee, J. (**1991**). "Mechanical and neural contributions to prolonged response latencies in immature kittens," Soc. Neurosci. Abstr. **17**, 1212.
Walsh, E. J., and Romand, R. (**1992**). "Functional development of the cochlea and the cochlear nerve," in *Development of Auditory and Vestibular Systems 2* (Elsevier, Amsterdam), Chap. 6, pp. 161–219.
Walsh, E. J., Gorga, M., and McGee, J. (**1992**). "Comparisons of the development of auditory brainstem response latencies between cats and humans," Hear. Res. **60**, 53–63.
Walsh, E. J., McGee, J., and Javel, E. (**1986a**). "Development of auditory-evoked potentials in the cat. I. Onset of response and development of sensitivity," J. Acoust. Soc. Am. **79**, 712–724.

Walsh, E. J., McGee, J., and Javel, E. (**1986b**). "Development of auditory-evoked potentials in the cat. II. Wave latencies," J. Acoust. Soc. Am. **79**, 725–744.

Walsh, E. J., McGee, J., and Javel, E. (**1986c**). "Development of auditory-evoked potentials in the cat. III. Wave amplitudes," J. Acoust. Soc. Am. **79**, 745–754.

Woolf, N. K., and Ryan, A. F. (**1984**). "The development of auditory function in the cochlea of the mongolian gerbil," Hear. Res. **13**, 277–283.

Woolf, N. K., and Ryan, A. F. (**1988**). "Contributions of the middle ear to the development of function in the cochlea," Hear. Res. **35**, 131–142.

Woolf, N. K., Ryan, A. F., and Harris, J. P. (**1986**). "Development of mammalian endocochlear potential: normal ontogeny and effects of anoxia," Am. J. Physiol. **250**, R493–R498.

Zine, A. (**2003**). "Molecular mechanisms that regulate auditory hair-cell differentiation in the mammalian cochlea," Mol. Neurobiol. **27**, 223–238.

Binaural detection with narrowband and wideband reproducible noise maskers. III. Monaural and diotic detection and model results

Sean A. Davidson
Department of Biomedical and Chemical Engineering, Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244

Robert H. Gilkey
Department of Psychology, Wright State University, Dayton, Ohio, 45435 and Human Effectiveness Directorate, Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio 45433

H. Steven Colburn
Boston University Hearing Research Center, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215

Laurel H. Carney[a)]
Department of Biomedical and Chemical Engineering, Department of Electrical Engineering and Computer Science, and The Institute for Sensory Research, 621 Skytop Road, Syracuse University, Syracuse, New York 13244

(Received 5 December 2004; revised 13 January 2006; accepted 23 January 2006)

A single-interval, yes-no, tone-in-noise detection experiment was conducted to measure the proportion of "tone present" responses to each of 25 reproducible noise-alone and tone-plus-noise waveforms under narrowband (100 Hz), wideband (2900 Hz), monotic, and diotic stimulus conditions. Proportions of "tone present" responses (estimates of the probabilities of hits and false alarms) were correlated across masker bandwidths and across monotic and diotic conditions. Two categories of models were considered; one based on stimulus energy or neural counts, and another based on temporal structure of the stimulus envelope or neural patterns. Both categories gave significant correlation between decision variables and data. A model based on a weighted combination of energy in multiple critical bands performed best, predicting up to 90% of the variance in the reproducible-noise data. However, since energy-based models are unable to successfully explain detection under a roving-level paradigm without substantial modification, it is argued that other variations of detection models must be considered for future study. Temporal models are resistant to changes in threshold under roving-level conditions, but explained at most only 67% of the variance in the reproducible-noise data.
 [DOI: 10.1121/1.2177583]

PACS number(s): 43.66.Dc, 43.66.Ba [AK] Pages: 2258–2275

I. INTRODUCTION

The simple task of detecting a pure tone in the presence of a noise masker has been studied for more than half a century (e.g., Fletcher, 1940; Hawkins and Stevens, 1950; Jeffress, 1968; Patterson, 1976; Kidd *et al.*, 1989; Richards *et al.*, 1992). Yet, despite extensive study, a definitive explanation of the underlying mechanisms has not emerged. To a first approximation, detection performance can be predicted on the basis of differences in energy statistics between the noise-alone and tone-plus-noise stimuli in a narrow band of frequencies close to the tone frequency. This "critical-band" model can predict the bulk of the masking data and also forms the foundation for much psychoacoustic theory and research. However, the critical-band model is clearly wrong in detail. For example, single-channel energy-based models cannot explain results using the roving-level paradigm (e.g., Kidd *et al.*, 1989). Physiologically, it is known that cochlear tuning is level dependent (Rhode, 1971), and that interactions among frequency bands are inherent (e.g., suppression). Although the rate of auditory-nerve discharge varies monotonically with stimulus energy, the temporal pattern of discharge also has the potential to code differences between the tone-plus-noise and noise-alone stimuli. Finally, intersubject performance differences suggest that both cognitive and peripheral differences influence detection. That is, individual subjects may apply different detection strategies to the response of the auditory nerve in order to generate detection judgments.

This study is part of a series focused on psychophysical and physiological aspects of the coding of tones in noise (Evilsizer *et al.*, 2002; Zheng *et al.*, 2002). The tools of human psychophysics, animal behavior and, in ongoing studies, physiological recordings, are being applied to gain a better understanding of the cues that listeners use to detect tones in noise and the neural mechanisms that make the use of

[a)]Author to whom correspondence should be addressed.

these cues possible. The goal of this paper is to systematically measure and model tone-in-noise detection. Although our focus here was on "monaural" processing, monotic and diotic performance were directly compared and the findings were considered in the context of other relevant binaural results. We use reproducible noise data to evaluate the predictions of a variety of detection models, including the critical-band model. Modeling results under wideband and narrowband stimulus conditions were compared. These models were also compared to the data from eight subjects [four from the current study and four from the study of Evilsizer *et al.* (2002)].

A. Reproducible noise

We employed reproducible noise as a masking stimulus because it allowed for a more detailed comparison between subject responses and model predictions. In a classical tone-in-noise detection experiment, the masker waveform is generated, independently and *without replacement* on each trial, by a random or pseudorandom process, such that the same waveform is never presented twice. The performance of models and subjects is "averaged" across masker waveforms and then compared. Averaging the data in this way discards the information inherent in the trial-by-trial fluctuations in the statistics of the noise and in the subjects' responses to those fluctuations. In contrast, a tone-in-reproducible-noise detection experiment utilizes masker waveforms that are randomly selected on each trial *with replacement* from a small set (typically 10 to a few hundred) of noise waveforms. The performance of subjects and models can then be compared on a waveform-by-waveform basis. Note that the word "waveform" refers to any tone-plus-noise or noise-alone stimulus waveform. The phrase "masker waveform" refers only to the noise-alone waveform before the addition of a tone. Because subjects' responses show substantial and reliable variation across waveforms, the data from this type of reproducible-noise experiment provide a more demanding test for models. Indeed, models that make the same average (across waveforms) responses as the subjects can still fail to predict their responses to individual waveforms (e.g., Gilkey *et al.*, 1985).

Although different in experimental methodology, the reproducible-noise detection task in this study is analogous to more traditional tone-in-noise masking experiments in which the noise is created randomly and without replacement on each trial (e.g., Dolan, 1968). When the number of reproducible masker waveforms exceeds 10, subjects do not "learn" individual masker waveforms (Pfafflin, 1968). Here, 25 masker waveforms were used in each listening condition. Thus, as in studies where each masker waveform is created randomly and used without replacement, subjects in this study did not learn individual masker waveforms. In contrast, if only a single or a few masker waveforms are used (e.g., Langhans and Kohlrausch, 1992) subjects are most likely functioning differently because they have the ability to learn the individual waveforms.

B. Masker bandwidth

The critical-band model suggests that energy outside of the auditory filter centered at the tone frequency will not influence detection performance. However, a variety of evidence indicates that a more broadband process affects masked detection. Research on auditory physiology (e.g., Ruggero, 1973; Kiang and Moxon, 1974; Schalk and Sachs, 1980; and Costalupes *et al.*, 1984) has long suggested the presence of suppressive regions in the response of auditory-nerve fibers outside of the normal excitatory band. Similar effects have also been routinely observed in psychoacoustic data (e.g., Shannon, 1976). Other psychoacoustic findings suggest the presence of broader-bandwidth interactions (some extending more than 2 octaves above and below the tone frequency), both in cases when such interactions are advantageous [e.g., in profile analysis (Green, 1988) and comodulation masking release (Hall *et al.*, 1984)] and in cases when such interactions are disadvantageous (e.g., Neff and Callaghan, 1988). Reproducible-noise studies provide evidence for across-critical-band comparison in tone-in-noise detection (e.g., Ahumada and Lovell, 1971; Gilkey and Robinson, 1986) and for differences between wideband- and narrowband-masker conditions, at least under dichotic conditions (Evilsizer *et al.*, 2002).

C. N_0S_0 to N_mS_m comparison

Experimental results were directly compared under diotic (N_0S_0) and monaural (N_mS_m) conditions. Most models of binaural processing assume that detection patterns collected under N_mS_m and N_0S_0 conditions are "equivalent" (i.e., there is no masking-level difference between the N_mS_m and N_0S_0 configurations). However, despite the fact that comparable average performance is observed in these two conditions, it is possible that different strategies or different cell populations are used under each of the two conditions, and that detection statistics for individual waveforms may differ between the two conditions. Although reproducible noise has not previously been used to compare these conditions directly, such a comparison can reveal similarities and differences that may be obscured when the data are averaged across waveforms in a typical masking experiment. We directly compared the data for each subject under these two conditions.

D. Models of tone-in-noise detection

The major focus of this effort is to evaluate models of tone-in-noise detection. The set of models examined were broadly sampled from the range of ideas that have been applied to tone-in-noise detection and correspond to a diverse set of underlying physiological mechanisms. The specific models were selected to satisfy three criteria: they can be readily implemented and applied to reproducible-noise data, they have been successfully applied to tone-in-noise detection, and they have broad neurophysiological or psychoacoustic relevance. The critical-band model (Fletcher, 1940) is directly evaluated. A variation of the critical-band model, referred to as the multiple-detector model, is also evaluated. This model uses a linear combination of the energy at the

output of multiple critical-band filters as a decision variable (e.g., Ahumada and Lovell, 1971; Gilkey and Meyer, 1987). Energy-related models based on the average discharge rate of model auditory-nerve fibers (Heinz *et al.*, 2001b) were also considered.

Although energy is perhaps the most obvious decision statistic to compute from the narrowband output of an auditory filter, the addition of a tone to noise also changes the temporal properties of the filter output. Richards (1992), for example, has examined how the shape of the narrowband envelope differs between tone-plus-noise and noise-alone stimuli. We extend her analysis to the current conditions and refer to it as the envelope-slope model. Finally, we examine the predictions of a single-cell version of the phase-opponency model (Carney *et al.*, 2002), which uses coincidence detection of auditory-nerve discharges from fibers tuned to different frequencies to reveal similarities in discharge timing across these fibers associated with the presence of the tone. This model has been shown to describe significant features of the tone-in-noise masking data, including the minimal effects on subject performance of randomizing (or roving) overall stimulus level within a trial and across intervals for a 2-interval task (Carney *et al.*, 2002; Kidd, 1987; Kidd *et al.*, 1989). These effects cannot be captured by simple energy-based models, whose decision variables are based entirely on overall stimulus energy. The within-trial rove confounds the energy-based models, which select the interval with more energy as containing the tone, regardless of whether the tone was present, increasing thresholds by about 25% of the rove range (Green, 1984).

Overall, this study takes a broad view of data and preexisting models[1] for monotic and diotic tone-in-reproducible-noise masking data. The ability of each model to explain results for a reproducible-noise task as well as each model's ability to predict roving-level data are considered.

II. PSYCHOPHYSICAL EXPERIMENT

A. Methods

Experimental procedures were matched to those of Evilsizer *et al.* (2002) and Gilkey *et al.* (1985). N_0S_0 (diotic) and N_mS_m (monaural, left ear) interaural configurations were tested. All listening was completed in a double-walled sound attenuating booth (Acoustic Systems, Austin, TX). Four subjects participated in this study and ranged from 19 to 25 years in age. Each subject had audiometrically normal hearing. None of the subjects had prior experience with tone-in-noise detection experiments. Subjects occasionally commented on the various listening tasks, but were not solicited to do so. Subjects were debriefed as to the use of reproducible stimuli at the end of the final testing session.

1. Stimuli

Stimuli were generated and controlled by MATLAB software (Mathworks, Natick, MA). All stimuli were presented via a TDT System 3 (Tucker Davis Technologies, Gainesville, FL) RP2 programmable D/A converter and TDH-39 headphones (Telephonics Corp., Farmington, NY). Reproducible masker waveforms used by Evilsizer *et al.* (2002) were down-sampled from 50 kHz to 48.125 kHz with the MATLAB "resample" function to be compatible with a TDT System 3 sample rate. Each of the 25 wideband maskers was created from a broadband, Gaussian white noise by zeroing frequency components outside 100–3000 Hz. Narrowband maskers were created from the wideband maskers by zeroing frequency components outside 450–550 Hz. This process ensured that the spectral content of the narrowband maskers was identical to that of the wideband maskers within a range of 450–550 Hz. The masker spectrum level was 40 dB SPL. Tones and maskers were mixed in software and had 300 -ms durations including 10-ms cosine-squared on/off ramps. Tones were always added in the sine phase.

2. Training

The final testing procedure was a single-interval task involving large numbers of trials near threshold. In this study, we refer to threshold as the E_S/N_0 value (in dB) resulting in a $d'=1$. The modeling procedures assumed that the subjects were performing at the same level throughout these trials. Therefore, an extensive subject-training paradigm was employed to establish stable performance and decision criteria, and to minimize effects of learning. There were three separate training tasks. These tasks progressed in difficulty and similarity to the final testing procedure. The first was a two-interval, two-alternative forced-choice, tracking task with feedback. This was followed by a one-interval, fixed-level task with feedback and then by a one-interval, fixed-level task without feedback. For all tasks, subjects were given an unlimited amount of time to respond after the final observation interval. Random maskers were employed during all training tasks to prevent learning the individual reproducible noise waveforms.

In the first training procedure, two-down, one-up tracking with feedback was used to quickly estimate a tone level where $d'=0.77$ (the 70.7% correct point on the psychometric function as described by Levitt, 1971). Subjects were asked to use a computer mouse to click one of two large buttons corresponding to "Interval One" or "Interval Two." The buttons were presented following the second observation interval on an external computer screen visible through the window of the sound-attenuating booth. Immediately following each response, the word "correct" or "incorrect" was displayed on the computer screen for 700 ms. Each track had a duration of 100 trials. Each trial contained a pair of randomly generated maskers that were frozen across intervals. Intervals were separated by 500 ms of silence. The step size used in the adaptive track was 4 dB for the first two reversals and 2 dB thereafter. Threshold estimates were calculated by averaging the tone levels at reversals in the track, excluding the first 4 or 5 reversals such that the number of remaining reversals was even. Subjects completed 10–15 runs for each configuration and bandwidth before moving on to single-interval tasks.

In the second training procedure, a one-interval fixed-level task with feedback was used to familiarize subjects with a one-interval task and ensure thresholds were stable. Subjects were asked to click on one of two large buttons

corresponding to "Tone" or "No Tone." Feedback was presented by displaying either "Correct" or "Incorrect" on the monitor for 700 ms after the response button was clicked. Each block had 100 trials with tone levels +3, +1 or −1 dB with respect to the tone level determined by the two-interval, two-alternative forced-choice task. Two 100-trial blocks were completed for each of the three tone levels for each interaural configuration and bandwidth. If a subject's threshold changed, this sequence was repeated, after adjusting the tone level in 1-dB steps, until a tone level was determined that resulted in a stable d' approximately equal to unity for each bandwidth and stimulus configuration.

After a stable tone level for each condition was determined, subjects completed the third training procedure, a one-interval training task without feedback using random noise. This task was used to determine whether d' could be expected to remain near unity when trial-by-trial feedback was eliminated during the testing phase. In the rare case that the value of d' changed such that it was no longer near unity, the level was adjusted with 1-dB resolution until d' returned to near unity.

3. Testing

The testing procedure was identical to the final training procedure except that reproducible maskers were used during testing. When present, the tone was always at the level determined during training without feedback. Final analyses were performed on complete data sets at a single tone level for each subject, with a d' near unity for the duration of the testing period.

Each testing set consisted of four blocks of 100 trials. Before each set, 20 practice trials with feedback were presented using a tone level 2 dB above the testing level. These practice trials were included to help the subjects maintain an effective and consistent detection strategy during the course of the experiment, given that feedback was *never* presented while using reproducible maskers. Random maskers were used during the practice trials to prevent learning the reproducible stimuli. Using the level determined during the training procedure, 12 sets were run without feedback for each bandwidth and configuration. During testing, 25 reproducible masker waveforms were used. Each masker waveform was presented twice with a tone and twice without a tone in each block. Overall, 96 presentations of each tone-plus-noise and 96 presentations of each noise-alone stimulus were made (a total of 48 blocks) in each of the four testing conditions (narrowband and wideband; monotic and diotic).

Bias (β, MacMillan and Creelman, 1991) was computed across all waveforms in a particular listening condition. If β departed more than 15% from 1.0 (i.e., equally likely to respond "tone" and "no tone") on a given run, subjects were instructed to "try to make an equal number of 'Tone' and 'No Tone' responses." All trials were included in the study regardless of the resulting bias measure. No attempt was made to account for or correct data with bias deviating from unity, as this was not the primary focus of the study.

Hit rates, or the conditional probabilities of responding yes given a particular tone-plus-noise waveform [$P(Y|T+N)$], and false-alarm rates, or the conditional probabilities of responding yes given a particular noise-alone waveform [$P(Y|N)$], were computed in each bandwidth and stimulus configuration for individual tone-plus-noise and noise-alone waveforms, respectively. A third set of probabilities of responding yes given a particular stimulus waveform [$P(Y|W)$] was also created for each listening condition. That is, $P(Y|W)$ includes the 25 $P(Y|Y+N)$ and the 25 $P(Y|N)$ values. The set of 50 $P(Y|W)$ values is referred to as a detection pattern (see Fig. 1). Each value of $P(Y|W)$ represents the probability of responding "yes" or "tone present" to an individual waveform (either tone-plus-noise or noise-alone) and can be thought of as the likelihood of the tone being perceived in that particular waveform.

Comparisons between detection patterns were quantified in terms of the square of the correlation coefficient, or the coefficient of determination (r^2). Statistically significant ($p < 0.05$) r^2 values occurred above the critical value of $r^2 = 0.08$ for comparisons involving the set of 50 $P(Y|W)$ values, or above the critical value of $r^2 = 0.17$ for comparisons involving either the set of 25 $P(Y|T+N)$ or the set of 25 $P(Y|N)$ values. These critical values were established using a two-tailed t-test (Bruning and Kintz, 1968). The coefficient of determination allows for comparisons between experimental results and the regression analysis presented in the modeling section of this paper.

B. Results and discussion

Subject performance was characterized on two levels. Overall performance for the ensemble of waveforms was characterized by averaging across noise-alone waveforms and across tone-plus-noise waveforms, as well as across repeated presentations of the waveforms. These ensemble-level results provide traditional measures of performance (e.g., d' and β) that represent the subjects ability to perform the tasks, show consistent responding among subjects, and allow direct comparison to other studies. In addition to these traditional ensemble-level results, detection patterns were computed for each subject under each condition by averaging across repeated presentations of the individual waveforms, but not across waveforms. Note that an "average subject" was created by averaging the $P(Y|W)$ values across the four subjects in this study. Detection patterns of the individual and average subjects were then used for the following empirical comparisons: $P(Y|T+N)$ and $P(Y|N)$ values were compared between N_mS_m and N_0S_0 stimuli to test the hypothesis that detection patterns were the same between stimulus configurations, for both narrowband and wideband stimuli. $P(Y|T+N)$ and $P(Y|N)$ values were compared between narrowband and wideband stimuli to test the hypothesis that stimulus information outside a critical-band influences detection patterns in both N_mS_m and N_0S_0 conditions. Finally, detection patterns were compared across subjects to reveal the possible uses of different/similar strategies within each stimulus configuration and bandwidth.

1. Reliability of the data

The ensemble-level results are shown in the left side of Table I. Data are presented for each subject and the average

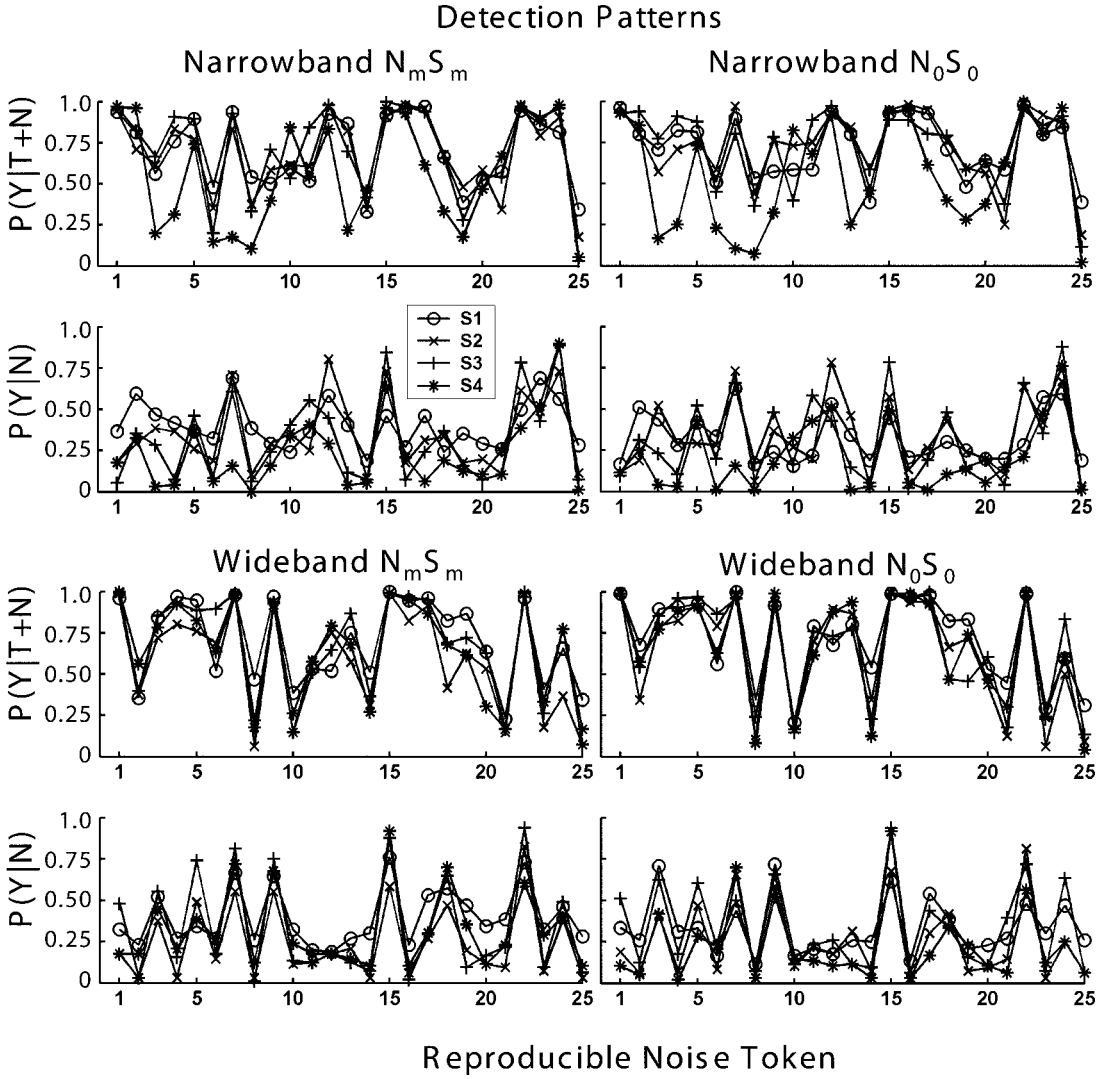


FIG. 1. Detection performance across reproducible maskers for the monaural and diotic stimulus conditions for narrowband maskers (top) and wideband maskers (bottom). Note that the horizontal axis is not a continuous variable; lines connect symbols corresponding to individual subjects to facilitate intersubject comparisons.

subject within each stimulus configuration and bandwidth (NB: narrowband, 450–550 Hz; WB: wideband 100–3000 Hz). Stimulus E_S/N_0 levels were within 3 dB for all subjects in all conditions tested. All d' values were within 20% of unity, thus subjects were tested at threshold. Bias results indicated that S4 had a tendency to report "tone absent" more often than "tone present" in all but the wideband N_mS_m stimulus configuration. *S*2 showed similar bias for the two wideband conditions. Other subjects showed little or no bias. Overall, subjects' performance levels were similar and near unity d' and bias.

The r^2 values between detection patterns constructed from the first and last 48 presentations of each stimulus and across the 25 individual tone-plus-noise and noise-alone waveforms in each bandwidth and stimulus condition were computed using $P(Y|T+N)$, $P(Y|N)$, and $P(Y|W)$ values (Table I). These first-half, last-half r^2 values indicate within-subject response consistency and serve as a reference for intersubject comparisons, as well as a reference for between-bandwidth and between-stimulus configuration comparisons. Subjects tended to respond more consistently on trials with the tone present than on trials with the tone absent. However, all coefficients were significant, and 52 of the 60 values were above 0.80, indicating that subjects were stable throughout the experiment.

2. Differences across waveforms

Results for each subject are presented in Table I and in Fig. 1 for each stimulus configuration. Figure 1 shows both monotic (left panels) and diotic (right panels) detection patterns. The hit-rates [$P(Y|T+N)$, upper panels] are based on trials containing the tone, and the false-alarm rates [$P(Y|N)$, lower panels] are based on trials without the tone.

Consistent with the results of other reproducible-masker studies, subjects' hit rates and false-alarm rates depended on the individual masker waveform (Evilsizer *et al.*, 2002; Isabelle and Colburn, 1991; Siegel and Colburn, 1989). The results of a χ^2 analysis with 24 degrees of freedom, using the procedure developed by Siegel and Colburn (1989), are shown in Table I. The χ^2 statistic was used to test the hypotheses that variation in $P(Y|T+N)$ and $P(Y|N)$ across

TABLE I. Psychophysical results across waveforms. E_S/N_0 is the ratio of tone energy to noise spectrum level (in dB). E_S/N_0, d', and β are shown for each combination of subject, interaural configuration, and noise bandwidth. NB and WB correspond to masker bandwidth: 100 Hz and 2900 Hz, respectively. The χ^2-values are given for performance across reproducible waveforms for both $P(Y|T+N)$ and $P(Y|N)$. The number of presentations of each masker for both $P(Y|T+N)$ and $P(Y|N)$ is given by n. The coefficient of determination, r^2, for the first half of trials vs the last half of trials is presented for both $P(Y|T+N)$ and $P(Y|N)$, individually, as well as across the entire ensemble of waveforms $[P(Y|W)]$. All χ^2 and r^2 values are significant to $p<0.001$.

						$P(Y\|T+N)$			$P(Y\|N)$			$P(Y\|W)$
Interaural configuration	BW	S	E_S/N_0	d'	β	χ^2	n	r^2	χ^2	n	r^2	r^2
N_mS_m	NB	S1	11.8	0.80	0.90	510.9	96	0.94	184.6	96	0.49	0.86
		S2	10.8	0.92	0.99	597.2	96	0.77	475.7	96	0.72	0.84
		S3	10.8	1.06	0.99	866.0	96	0.94	718.9	96	0.81	0.90
		S4	10.8	0.92	1.45	1003.4	96	0.92	602.2	96	0.86	0.91
		$S_{avg(4)}$	11.1	0.93	1.08	2379.6	384	0.96	1391.3	384	0.91	0.96
	WB	S1	11.8	0.80	0.91	699.5	96	0.92	278.5	96	0.72	0.88
		S2	10.8	0.96	1.40	926.3	96	0.94	599.3	96	0.83	0.92
		S3	10.8	0.90	1.00	1017.1	96	0.96	952.5	96	0.90	0.95
		S4	9.8	0.86	1.06	856.1	96	0.84	543.0	96	0.81	0.88
		$S_{avg(4)}$	10.8	0.88	1.09	3240.9	384	0.98	2057.7	384	0.96	0.98
N_0S_0	NB	S1	12.8	1.08	0.94	400.5	96	0.79	233.3	96	0.64	0.87
		S2	10.8	1.09	0.94	593.5	96	0.79	498.6	96	0.72	0.86
		S3	10.8	1.12	0.96	608.4	96	0.86	666.4	96	0.84	0.92
		S4	10.8	1.01	1.58	984.7	96	0.94	610.0	96	0.92	0.95
		$S_{avg(4)}$	11.3	1.07	1.11	1907.3	384	0.94	1427.3	384	0.87	0.95
	WB	S1	11.8	1.03	0.94	728.9	96	0.88	300.0	96	0.81	0.92
		S2	10.8	1.08	1.25	1138.0	96	0.92	647.8	96	0.88	0.94
		S3	9.8	0.91	1.03	1066.6	96	0.94	671.9	96	0.88	0.94
		S4	10.8	1.17	1.37	1079.8	96	0.92	686.6	96	0.86	0.93
		$S_{avg(4)}$	10.8	1.04	1.15	3776.1	384	0.97	1903.0	384	0.98	0.98

maskers was not due to chance (i.e., that the detection patterns were waveform dependent). All χ^2 statistics greatly exceeded the $p<0.001$ significance level ($\chi^2_{crit}=51.18$), and thus the null hypotheses of only random variations in detection patterns were rejected. Note that a larger χ^2 statistic indicates a more reliable detection pattern.

3. Comparisons between stimulus configurations

This experiment tested the hypothesis that psychophysical detection patterns are the same for the N_mS_m and N_0S_0 stimulus configurations. Detection patterns were significantly ($p<0.001$) correlated between the N_mS_m and N_0S_0 stimulus configurations for both narrowband and wideband maskers (see Table II). In Fig. 1, the similarity between N_mS_m and N_0S_0 detection patterns for individual subjects is also visible.

In this study, tone levels producing unity d' for individual subjects were within about 1 dB for the N_mS_m and N_0S_0 conditions. These results were consistent with the failure to find $N_mS_m-N_0S_0$ MLDs in several other experiments (Sever and Small, 1979; Egan *et al.*, 1969; Egan, 1965; Hirsh and Burgeat, 1958; Blodgett *et al.*, 1958) that used free-running noise and similar masker levels as in this experiment. However, at lower masker levels, some have observed an $N_mS_m-N_0S_0$ MLD.[2] In a study that used a *single* frozen noise masker, Langhans and Kohlrausch (1992) compared monaural and diotic thresholds and found an $N_mS_m-N_0S_0$ MLD. It is not surprising that the results in these two studies are different, since subjects could use the learned properties of the single noise waveform in their study. As previously explained, our study was designed to provide results that are comparable to performance in traditional experiments with random noise.

TABLE II. Correlations between detection patterns for N_mS_m and N_0S_0 stimulus configurations. Narrowband and wideband r^2 values are presented for each subject. These r^2 values were calculated from responses across the 25 tone-plus-noise waveforms $[r^2_{P(Y|T+N)}]$, 25 noise-alone waveforms $[r^2_{P(Y|N)}]$, as well as across the ensemble of all 25 tone-plus-noise and 25 noise-alone waveforms $[r^2_{P(Y|W)}]$.[a]

Bandwidth	Subject	$r^2_{P(T+N)}$	$r^2_{P(Y\|N)}$	$r^2_{P(Y\|W)}$
NB	S1	0.94	0.70	0.90
	S2	0.86	0.88	0.92
	S3	0.81	0.86	0.90
	S4	0.96	0.86	0.95
	$S_{avg(4)}$	0.89	0.90	0.91
WB	S1	0.79	0.64	0.83
	S2	0.92	0.92	0.95
	S3	0.90	0.82	0.89
	S4	0.88	0.90	0.92
	$S_{avg(4)}$	0.95	0.91	0.95

[a]All values $p<0.001$.

TABLE III. Correlations between detection patterns for narrow (100 Hz) and wide (2900 Hz) stimulus bandwidths. Individual-subject r^2 values are presented within either the N_mS_m or N_0S_0 stimulus configuration and were calculated from responses across the 25 tone-plus-noise waveforms [$r^2_{P(Y|T+N)}$], 25 noise-alone waveforms [$r^2_{P(Y|N)}$], as well as across the ensemble all 25 tone-plus-noise and 25 noise-alone waveforms [$r^2_{P(Y|W)}$].

Stimulus configuration	Subject	$r^2_{P(Y\|T+N)}$	$r^2_{P(Y\|T)}$	$r^2_{P(Y\|W)}$
N_mS_m	S1	0.21[a]	0.03	0.37[b]
	S2	0.41[b]	0.31[b]	0.54[b]
	S3	0.29[a]	0.38[b]	0.49[b]
	S4	0.07	0.11	0.23[b]
	$S_{avg(4)}$	0.28[a]	0.29[a]	0.48[b]
N_0S_0	S1	0.36[b]	0.09	0.56[b]
	S2	0.41[b]	0.41[b]	0.59[b]
	S3	0.50[b]	0.46[b]	0.61[b]
	S4	0.06	0.07	0.24[b]
	$S_{avg(4)}$	044[b]	0.42[b]	0.64[b]

[a] $p<0.05$.
[b] $p<0.01$.

4. Comparisons between bandwidths

Between-bandwidth comparisons were performed to determine whether stimulus information outside the critical-band centered at the tone frequency influenced detection patterns. Recall that narrowband maskers were created by removing all components outside the band centered at 500 Hz; thus, narrowband and wideband stimuli should be identical within the 500-Hz critical band. Between-bandwidth r^2 values are presented in Table III and the underlying correlations are also visible in the detection patterns in Fig. 1. Although between-bandwidth r^2 values were significant for most subjects in both stimulus configurations (recall that for $n=25$, $r^2>0.17$ is significant, $p<0.05$), values were significantly lower than those observed between the first 48 and last 48 presentations of each masker (first-half, last-half r^2, Table I). A test for significant difference between two nonindependent correlations (Bruning and Kintz, 1968) was performed between the first-half, last-half correlations and cross-bandwidth correlations to determine if the two sets of correlations differed significantly. Results showed significant differences ($p<0.01$) for 37 of the 64 tests (4 subjects $\times 2$ bandwidths $\times 2$ stimulus configurations $\times 2$ tone cases, with and without tone $\times 2$ halves). Of the 27 that failed at the $p<0.01$ level, 15 involved statistically insignificant ($p<0.05$) cross-bandwidth correlations of 0.34 or less. Given that the between-bandwidth variation in detection patterns was statistically larger than the within-bandwidth variation in detection patterns (between the first half and the last half of trials recorded), these results indicate that detection performance was influenced by stimulus information outside one critical-band for both monotic and diotic conditions. It is notable that the noise added outside the critical band did not affect threshold (i.e., E/N_0 where d' was near unity), but it did affect detection patterns.

Narrowband maskers had the same magnitude and phase as wideband maskers for components in the 100-Hz region centered on the stimulus frequency. Consequently, stimulus information in the wideband masker falling inside one critical-band was approximately (ignoring frequency components present in the auditory filter skirts) the same as that of the narrowband masker, while information falling in adjacent critical-bands was unique to the wideband masker.

The effect of the noise added outside the critical band presents a challenge for models based on energy in a single critical band. These models cannot correctly predict changes in detection patterns unless energy differences in the filter skirts are responsible for the differences between the two conditions. Another hypothesis is that listeners use different detection strategies for different noise bandwidths, causing reduced correlations (Evilsizer *et al.*, 2002). In this case, a model based on a single critical band may still be used, but such models would require different decision variables for the different bandwidths.

5. Comparisons among subjects

Comparisons were performed between detection patterns, $P(Y|W)$, for each subject-subject pair to reveal intersubject consistency and the possible uses of different/similar strategies within each stimulus configuration and bandwidth. The means and ranges of intersubject r^2-values are provided in each experimental condition in Table IV.

All possible intersubject correlations were computed and were found to be significant ($p<0.05$) for all monaural stimuli. Within the monaural stimulus condition, higher intersubject correlations were observed for wideband stimuli relative to narrowband stimuli. S4 was the least correlated to other subjects in the monaural stimulus condition (see Fig. 1) and also had the lowest correlation across stimulus bandwidths.

Intersubject correlations were lower for narrowband conditions relative to wideband conditions in both the N_mS_m

TABLE IV. Mean, minimum, and maximum intersubject r^2 values [$P(Y|W)$, calculated from responses across the ensemble of all 25 tone-plus-noise and 25 noise-alone waveforms] for the subjects in the current study and the subjects from Evilsizer *et al.* (2002).

Interaural configuration	Bandwidth	mean[$r^2_{[P(Y\|W)]}$]	min[$r^2_{[P(Y\|W)]}$]	max[$r^2_{[P(Y\|W)]}$]
N_mS_m	NB	0.67	0.48	0.84
	WB	0.85	0.81	0.89
N_0S_0	NB	0.62	0.31	0.81
	WB	0.72	0.51	0.91

TABLE V. Proportions of predictable variance estimated by four different methods (see text for details) for N_mS_m and N_0S_0 stimulus configurations and NB (100 Hz) and WB (2900 Hz) conditions.

Interaural configuration	BW	Statistical method			
		V_p	$\sigma_p^2/\sigma_{\text{Tot}}^2$	$\sigma_{\text{ext}}^2/\sigma_{\text{Tot}}^2$	$\sigma_{\text{Waveform}}^2/\sigma_{\text{Tot}}^2$
N_mS_m	NB	0.990	0.994	0.996	0.939
	WB	0.995	0.995	0.998	0.975
N_0S_0	NB	0.988	0.997	0.997	0.941
	WB	0.996	0.998	0.998	0.977

and N_0S_0 listening configurations in this study and also in both the N_0S_0 and N_0S_π listening configurations in Evilsizer *et al.* (2002). Subjects in the current study reported more difficulty with the narrowband task, which may have led to increased variability in detection strategies and reduced intersubject correlations for narrowband conditions. Diotic intersubject correlations were computed for the four subjects in this work (S1–S4), as well as the four subjects in the Evilsizer *et al.* (2002) study (S5–S8). No noticeable decrease in intersubject correlation occurred when comparing across studies (i.e., when comparing between S1–S4 and S5–S8), although S4 and S7 did show the lowest intersubject correlations [0.67 and 0.52 for 100-Hz N_0S_0 $P(Y|T+N)$ and $P(Y|N)$, respectively]. As in the monotic case, diotic trials using wideband maskers had higher intersubject correlations than those using narrowband maskers. However, low intersubject correlations were rare, and all but 6 of the 112 diotic intersubject correlations were significant at the $p<0.05$ level. Overall, these results indicate that detection patterns were consistent across subjects, except for S4 and S7.

6. Internal noise

In previous studies of detection using reproducible noise, investigators have taken advantage of the fact that the amount of variability in the responses across a set of waveforms is available (i.e., the detection pattern) to estimate the relative amount of internal variability. The ratio of internal to external noise would be large if the external noise sample had a small effect on performance and small if performance was highly variable across noise waveforms. This measure is also useful for comparison to other psychophysical results. To estimate the internal-to-external noise ratio, the internal variance (σ_{int}^2) is assumed to be equal to 1 for all masker waveforms, and the external variance (σ_{ext}^2) is estimated from the sample variance of the z scores across the set of waveforms; the ratio ($\sigma_{\text{int}}^2/\sigma_{\text{ext}}^2$) is then the reciprocal of the square root of the external variance (Siegel and Colburn, 1989). Internal-to-external noise ratios (in standard deviations) ranged from approximately 1.0 to 4.0 across the four subjects in our study[3] (similar to values reported in Spiegel and Green, 1981; Siegel and Colburn, 1989; and Isabelle and Colburn, 1991). Monotic and diotic results were similar for each subject, indicating that the amount of internal noise did not substantially change between the monotic and diotic tasks.

7. Predictable variance

In Sec. III, various models are used to predict the changes in $P(Y|W)$ observed across the combined set of 25 noise-alone and 25 tone-plus-noise waveforms. As discussed in the previous sections, the data are quite reliable, suggesting that the models have the potential to predict a substantial portion of the variability in $P(Y|W)$. Table V presents four estimates of the portion of predictable variance for the average subject under each of the four stimulus conditions.

The first column shows estimates of the proportion of predictable variance based on the first-half, last-half correlations for the results described in Table I. These estimates were obtained using the equation presented by Ahumada and Lovell (1971),

$$V_p = \frac{r_{12}}{r_{12} + 0.5(1 - r_{12})}, \tag{1}$$

where V_p is the proportion of predictable variance, and r_{12} is the first-half, last-half correlation. [Note because the first-half, last-half correlations from the Evilsizer *et al.* (2002) experiment were not readily available; these estimates are based on only the four subjects in this experiment, even under the N_0S_0 condition.]

The second column of Table V presents estimates of the proportion of predictable variance calculated from the assumption that the only source of error is that associated with estimating a proportion based on Bernoulli trials, such that

$$\frac{\sigma_p^2}{\sigma_{\text{Tot}}^2} = 1 - \left[\frac{E\left[\dfrac{P(Y|W)[1 - P(Y|W)]}{n_{\text{trials}}} \right]}{\sigma_{\text{Tot}}^2} \right], \tag{2}$$

where σ_p^2 is the predictable variance, σ_{Tot}^2 is the total variance in the $P(Y|W)$ values for the average subject, $P(Y|W)$ is the probability of a "tone present" response for each tone-plus-noise and each noise-alone waveform, n_{trials} is the number of trials used to estimate each $P(Y|W)$, and the expected value is across the combined set of 50 waveforms (25 tone-plus-noise and 25 noise-alone waveforms).

The third column presents estimates of the proportion of predictable variance based on the internal-to-external noise ratio ($\sigma_{\text{int}}^2/\sigma_{\text{ext}}^2$) discussed in Sec. II B 6. Assuming that the internal variance estimate is reduced by the number of trials,

$$\frac{\sigma^2_{\text{ext}}}{\sigma^2_{\text{Tot}}} = \frac{\sigma^2_{\text{ext}}}{\left[\sigma^2_{\text{ext}} + \frac{\sigma^2_{\text{int}}}{n_{\text{trials}}}\right]} = \frac{1}{\left[1 + \frac{\left(\frac{\sigma^2_{\text{int}}}{\sigma^2_{\text{ext}}}\right)}{n_{\text{trials}}}\right]}, \quad (3)$$

where the internal-to-external noise ratio ($\sigma^2_{\text{int}}/\sigma^2_{\text{ext}}$) is the reciprocal of the variance in the z scores of $P(Y|W)$ across the waveforms (i.e., predictable variance), σ^2_{Tot} is the estimated total variance across the z scores of $P(Y|W)$ for each waveform for the average subject, and n_{trials} is the number of trials used to estimate each z score of $P(Y|W)$.

The fourth column presents estimates of the proportion of predictable variance based on partitioning the variance from separate Waveform × Subject Analyses of Variance computed for each condition. A randomized-blocks design was used, such that each cell had a single value, $P(Y|W)$ for that waveform and that subject. Adapting equations from Kirk (1995, pp. 258 and 267) for the Random-Effects Model, we have

$$\text{MS}_{\text{Waveform}} = n_{\text{subject}} \cdot \sigma^2_{\text{Tot}} = \sigma^2_{\text{error}} + \sigma^2_{(\text{Waveform}\times\text{Subject})} + n_{\text{subject}} \cdot \sigma^2_{\text{Waveform}}$$

and

$$\text{MS}_{\text{Residual}} = \sigma^2_{\text{error}} + \sigma^2_{(\text{Waveform}\times\text{Subject})}.$$

Therefore,

$$\frac{\sigma^2_{\text{Waveform}}}{\sigma^2_{\text{Tot}}} = \frac{\text{MS}_{\text{Waveform}} - \text{MS}_{\text{Residual}}}{\text{MS}_{\text{Waveform}}}, \quad (4)$$

where $\sigma^2_{\text{Waveform}}$ is the variance associated with the effect of waveform (i.e., predictable variance), $\text{MS}_{\text{Waveform}}$ is the mean square for the effect of Waveform, $\text{MS}_{\text{Residual}}$ is the mean square for the residual error based on the Subject by Waveform interaction, n_{subject} is the number of subjects, and σ^2_{Tot} is the estimated total variance in the $P(Y|W)$ values for the average subject.

A similar analysis can be applied to estimate the proportion of predictable variance in the responses of individual subjects as opposed to the proportion of predictable variance in the average subject, as presented here. The results of such an analysis are reported in the Appendix.

All four estimates suggest that over 93% of the variance is potentially predictable in all four conditions, with the estimates of the proportion of predictable variance ranging from 0.939 to 0.998 across methods and conditions.

III. MODELS FOR MASKED DETECTION

A. General approach

Two general categories of models were considered in this study. The general structure of each model is illustrated in Fig. 2: six were based on energy or average firing rates: [CB: critical band; MD: multiple detector; and four AN-count-based models (L: low spontaneous-rate; H: high spontaneous rate; L_{SS}: steady-state, low spontaneous-rate; and H_{SS}: steady-state, high spontaneouse-rate)], and two were

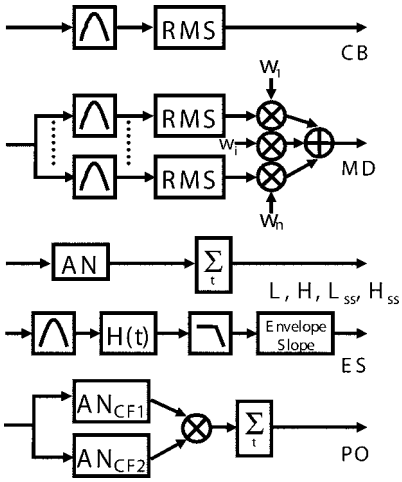


FIG. 2. Summary of model structures. The model inputs were pressure waveforms as a function of time. Each model's output was a decision variable corresponding to a given input waveform. Models, listed top to bottom are: CB, critical band; MD, multiple detector; low (L) and high (H) spontaneous-rate and steady-state versions of low (L_{SS}) and high (H_{SS}) spontaneous-rate auditory-nerve fibers; envelope slope (ES); and phase opponency (PO). $H(t)$ denotes the absolute value of the complex analytic function.

based on temporal variation: [an envelope-based model (ES: envelope slope), and the phase-opponency model (PO)]. The first category combines energy-based models with models based on neural counts of AN fibers, which are a first approximation of a physiologically-based energy estimate. The temporal models considered here were based on either the envelopes of narrowband-filtered stimuli (ES) or on fine-structure, as assessed using a physiological coincidence-detection-based model for detection (PO). The specific function and structure of each model will be described in detail in the following sections. All model simulations used a 50 -kHz sampling rate.

These models used either stimulus-based or physiologically-based decision variables that were computed from specific noise-alone and tone-plus-noise waveforms. Model outputs were compared to the individual data of each subject and also to the average subject [S_{avg}; including subjects from the Evilsizer *et al.* (2002) study for the N_0S_0 condition]. Although the models tested here were monaural, modeling results were compared to both monotic and diotic psychophysical data because subjects' detection patterns were highly correlated across the N_mS_m and N_0S_0 stimulus configurations. Model predictions were quantified using the following logistical regression procedure (Gilkey and Robinson, 1986). Model decision variables were fit to hit and false-alarm rates for individual subjects as well as to those for the average subject. An ogive function with two parameters was used for each case,

$$P'(Y|W) = \frac{1}{1 + e^{[[-(\log_{10}(x)-\mu)]/\theta]}}, \quad (5)$$

where $P'(Y|W)$ is the probability of a model "yes" response, x is the model decision variable, μ is the threshold parameter, and θ is the slope parameter (Gilkey and Robinson, 1986; MacMillan and Creelman, 1991). A Nelder-Mead sim-

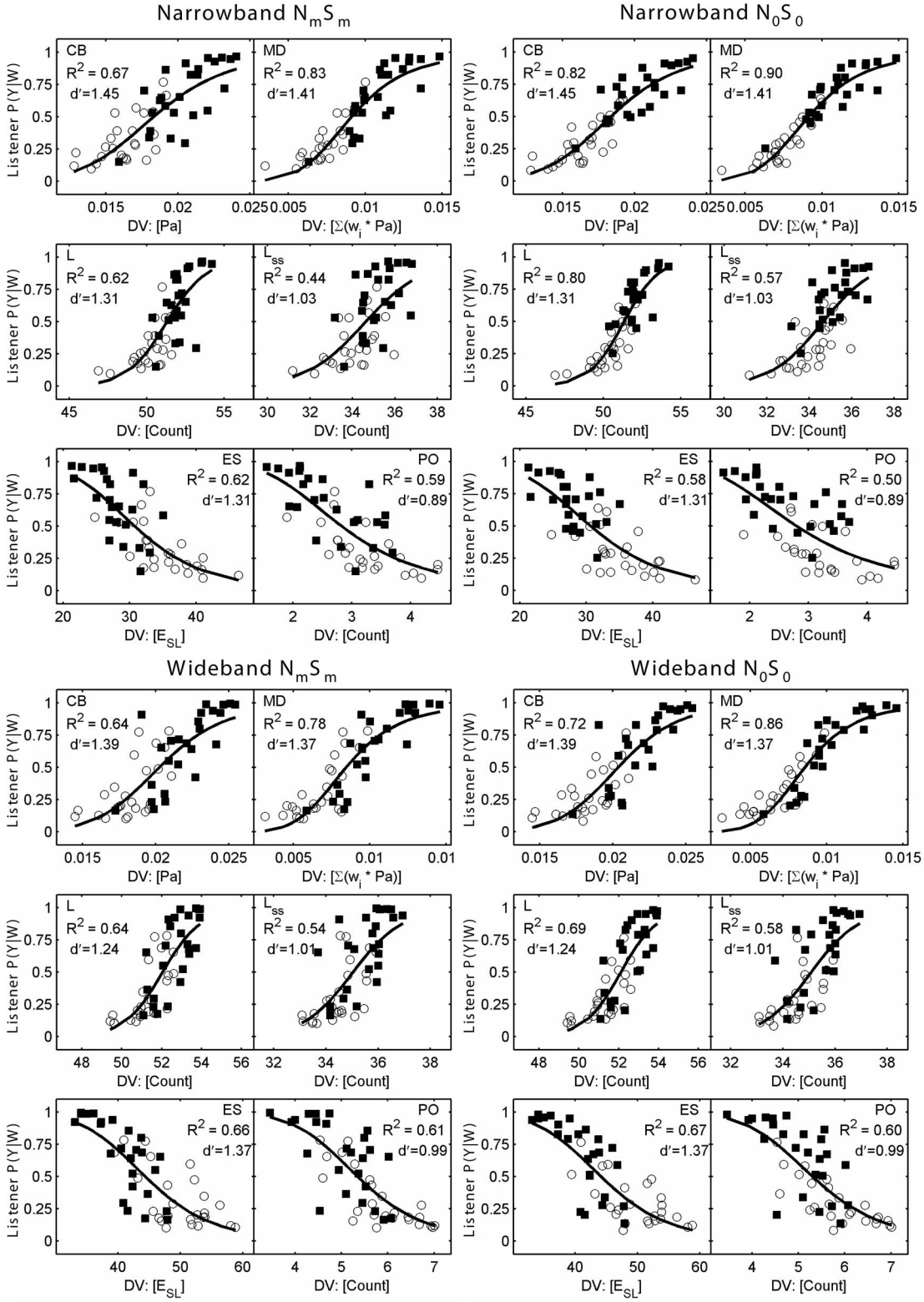


FIG. 3. Model predictions. Logistical regression was used to quantify the ability of the models in Fig. 2 to predict the average subject's $P(Y|W)$ values. Each panel shows the resulting ogive, and a scatter plot of model decision variables vs the average subject's $P(Y|W)$ values, subdivided into $P(Y|T+N)$ (squares, the probability of responding yes to a tone-plus-noise waveform) and $P(Y|N)$ (circles, the probability of responding yes to a noise-alone waveform). R^2 values and model d's are also shown. A perfect model would be indicated by all points in the scatter plot falling along the ogive and $R^2=1$. Model abbreviations are as in Fig. 2. Note that only the low-spontaneous-rate AN models are shown because of the relatively poor performance of high-spontaneous-rate AN models (see text).

plex direct search (MATLAB fminsearch) was used to minimize the sum of squared deviations between the set of subject $P(Y|W)$ values and the set of model $P'(Y|W)$ values. The search produced threshold and slope parameters corresponding to the best model fit for each individual subject and for the average subject. Model fits to the average subject are shown in Fig. 3. Note that although the plots in Fig. 3 explicitly show tone-plus-noise and noise-alone data, models were fit to the combined tone-plus-noise and noise alone data, that is, all 50 $P(Y|X)$ values. The resulting threshold and slope parameters were used in Eq. (5) to transform the model decision variable corresponding to each tone-plus-noise waveform into a model $P(Y|T+N)$, and the model decision variable corresponding to each noise-alone waveform into a model $P(Y|N)$, yielding a set of model hit and false-alarm rates. The following R^2 statistic was used to quantify the ability of each model to predict the data. The proportion of variance accounted for by the fit is defined as

$$R^2 = \frac{\sum_{i=1}^{50} [P'(Y|W_i) - P(Y|W_i)]^2}{\sum_{i=1}^{50} [P(Y|W_i) - \overline{P(Y|W)}]^2}, \quad (6)$$

where $P(Y|W)$ is the subject's probability of a "yes" response and $P'(Y|W)$ is the model's probability of a "yes" response. The numerator of Eq. (6) is the sum of squared errors between individual model and subject probabilities across waveforms, where i indicates the waveform index. The denominator is the sum of squared deviations between subject probabilities corresponding to individual waveforms and the mean subject probability across waveforms. The symbol R^2 is used to indicate the nonlinear fit between the model decision variable and the subject $P(Y|W)$ values, but is otherwise analogous to the r^2 values between sets of $P(Y|W)$ values reported in Sec. II. The use of model parameters in addition to the slope and threshold based on Eq. (5) is specifically reported in each model description below.

Each model was run at an E_S/N_0 of 10.8 dB, the median signal-to-noise ratio of all listeners combined across all conditions. Model d' values were computed as a relative sensitivity measure from the raw decision variables (i.e., before any fitting was performed). Thus, a single d' was computed for each model in each bandwidth as the difference in the means of the model decision variable between tone-plus-noise and noise-alone waveforms, divided by the square root of the average variance of model decision variables for tone-plus-noise and noise-alone waveforms. Individual model d' values were expected to be identical under the N_mS_m and N_0S_0 listening conditions because the models were monaural and d' values were computed from decision variables based on the same stimulus waveforms (before any fitting procedures were completed).

None of the models in this study included internal noise. Accordingly, if we assume the subjects' internal-to-external noise ratios are equal to 1 (internal-to-external noise ratios for the average subject ranged from 1.0 to 1.3), and that the model sensitivity equals the average listeners' sensitivity, then we expect model d' values to be larger than those of the subjects by a factor of $\sqrt{2}$ (i.e., since the model contained no internal noise, and the models and subjects experienced the same external noise, the overall model noise was half that of the listener, and thus the model d' was larger by a factor of $\sqrt{2}$).

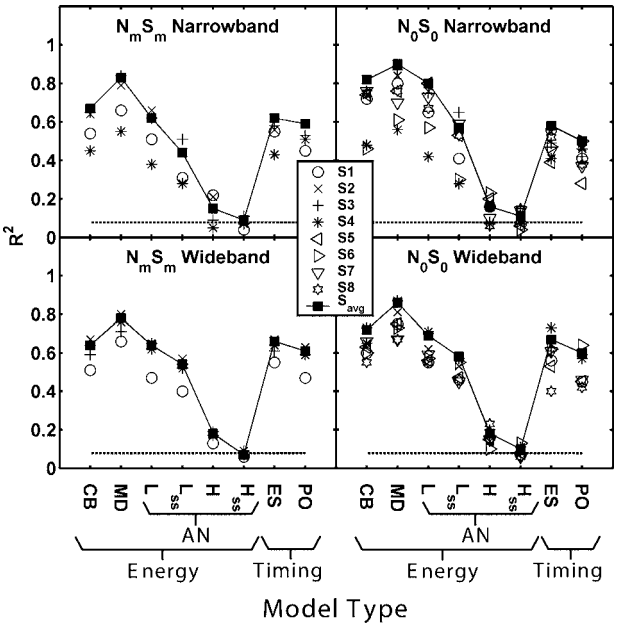


FIG. 4. Summary of model-subject R^2 values for combined tone-plus-noise and noise-alone trials for the eight models tested in this work. Individual subject data are shown by different symbols. The average-subject data are shown by the solid line with closed squares. Note that the horizontal axis is not a continuous variable; lines connect symbols corresponding to the average subject to facilitate comparisons between models. Model abbreviations are as in Fig. 2. Recall that CB, MD, and AN models use energy-based decision variables. ES and PO use temporally-based decision variables. R^2 values exceeding the $p<0.05$ significance levels are above the thick-dashed line.

Modeling results are presented in Figs. 3 and 4. Each panel in Fig. 3 shows the ogive fit of an individual model to the average subject. Each panel also shows a scatter plot of the average subject's $P(Y|T+N)$ (squares) and $P(Y|N)$ (circles) values versus the model decision variable for each waveform. Model abbreviations are as in Fig. 2. Values for R^2 and d' are reported in the panels of Fig. 3 for each model in all four listening conditions. Note that the high-spontaneous-rate auditory-nerve models were omitted from Fig. 3 because of their relatively poor performance. In Fig. 4, model types are listed on the abscissa and the ordinate shows the proportion of variance in subject detection patterns predicted by each model (R^2). Scatter plots for individual-subject model predictions are available at http://web.syr.edu/~lacarney/auditory.htm. In the following sections, each model is explicitly described and evaluated. A final section addresses effects of roving the stimulus level.

B. Critical-band model

1. Methods

Several energy-based models were implemented and compared to monotic and diotic psychophysical data collected from this study and from Evilsizer *et al.* (2002). The first energy-based model, the simple critical-band (CB) model, was composed of a 4th-order linear gammatone filter

centered at the tone frequency (500 Hz). The basic structure of the CB model is shown in the CB panel of Fig. 2. The equivalent-rectangular bandwidth was fixed at 75 Hz to correspond to the auditory-filter estimates of Glasberg and Moore (1990). The rms energy of the filter output was used as the decision variable, similar to the "Energy Mnemonic" described in Gilkey and Robinson (1986). This model had no free parameters in addition to the slope and threshold of the ogive fit [Eq. (5)].

2. Results and Discussion

This model predicts 64%–82% of the variance in the average subject's detection patterns as shown in Figs. 3 and 4 (CB). Model sensitivity was slightly higher for narrowband stimuli ($d'=1.45$) than for wideband stimuli ($d'=1.39$), however both of the resulting d' values were consistent with what would be expected for a model with no internal noise (i.e., $d' \approx \sqrt{2}$, as described above). The finding that stimulus energy was related to subject responses is consistent with previous studies (e.g., Pfafflin and Mathews, 1966; Ahumada *et al.*, 1975; Gilkey and Robinson, 1986). Critical-band model predictions were significant for all subjects in all conditions; however, a substantial portion of the variance in the average subject's detection pattern was unaccounted for under wideband conditions. The lower R^2 values under wideband conditions (with respect to narrowband conditions) can be understood by considering that the model, in contrast to the experimental data (see Table III), predicts a high correlation between narrow and wideband results ($r^2=0.98$).

C. Multiple-detector model

As noted previously, the results of Ahumada and Lovell (1971), Gilkey and Robinson (1986), Isabelle and Colburn (1991), and Evilsizer *et al.* (2002) suggest that information outside the critical-band influences detection patterns for wideband, tone-in-noise detection tasks. In the current section, a more sophisticated energy-based model is explored. Referred to here as the multiple-detector model (MD), this model is capable of incorporating energy outside the critical-band centered at the tone frequency.

1. Methods

The basic model structure is shown in Fig. 2. A linear combination of the RMS output of seven 4th-order, 75-Hz equivalent-rectangular bandwidth (ERB) linear gammatone filters, spaced at 75-Hz intervals from 275 to 725 Hz, was used as the decision variable (adapted from Gilkey and Robinson, 1986). The coefficients for the linear combination were determined using a two-stage fit. The Matlab fminsearch function was used to iteratively adjust six of the seven weights in the linear combination (the weight for the 500 -Hz channel was fixed at 1.0); within each iteration the fminsearch function was called again to find the best fitting ogive between the model decision variable (i.e., the linear combination with the "current" weights) and the subject's $P(Y|W)$ values.

2. Results and discussion

The best-fitting weights had a characteristic shape for wideband stimuli; the strongest weight, which was positive, was observed for the band centered at the tone frequency, and smaller magnitude negative weights were observed for frequency bands above and below the tone frequency.

Because our study also included narrowband (100-Hz) stimuli, it was of interest to consider the importance of filters located more than a critical band away from the tone frequency and centered outside the region of greatest stimulus energy. It was found that dramatically different weights could be obtained for filters more than one critical band above and below the tone frequency with little impact on the variance explained by the model for narrowband stimuli. In order to investigate this phenomenon, the results from a seven-filter MD model were compared to those from a three-filter MD model, each fit to the average subject using narrowband stimuli. The three-filter MD model used only three 4th-order, 75-Hz equivalent-rectangular bandwidth (ERB) linear gammatone filters centered at 425, 500, and 575 Hz, and was otherwise identical to the 7-filter model. Partial correlation coefficients were computed to measure the variability explained by the 7-filter model, while already accounting for the variability explained by a model using only the 3 filters closest to the tone frequency. These coefficients were found to be insignificant ($p>0.1$) for both N_mS_m and N_0S_0 stimulus conditions, indicating that, as expected, energy at filters more than a critical band from the tone frequency did not significantly improve MD predictions for narrowband stimuli. Weights for wideband stimuli were found to be consistent across listeners for all seven filters. The MD model was therefore restricted to the three filters closest to the tone frequency for narrowband stimuli, while all seven filters were included for the wideband stimuli. This model used a total of two free parameters for narrowband stimuli, and six free parameters for wideband stimuli (the central filter weight was always fixed at 1.0), in addition to the slope and threshold of the ogive fit. Figure 5 shows weight patterns across the outputs of the filters in the multiple-detector model for the average subject for all four stimulus conditions.

None of the other models explored in this study included additional model parameters (beyond the slope and threshold of the ogive function) that were *fit to subject data*. In order to more closely compare the multiple-detector model to other models in this work, weight patterns derived for the average subject in the N_0S_0 conditions were used for all MD model predictions. Using the average subject's weights for all subjects served to eliminate the process of fitting additional model parameters to individual subject data, while still fitting the model to the subject data in an average sense. Filter weights were very consistent across listeners.

Results for the MD model using the average subject's weights are shown in Figs. 3 and 4. This model accounted for more variance in the average subjects' detection patterns than any of the other models discussed here. The MD model was able to predict 78%–90% of the variance in the average subject's detection patterns. Figures 3 and 4 show, interestingly, that this model made better predictions for narrowband

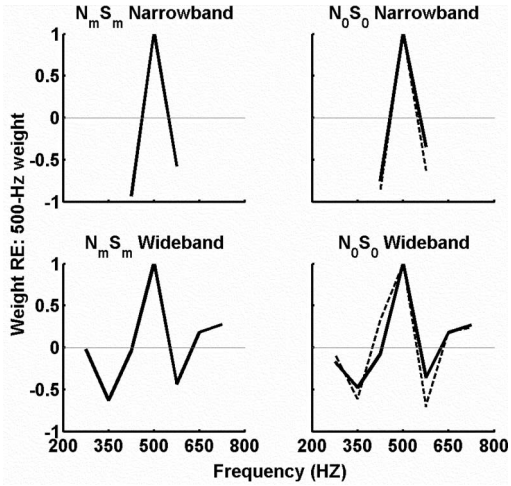


FIG. 5. Spectral weights from the multiple-detector model. The average subject is shown by the thick line. Subject data were taken from this work and that of Evilsizer *et al.* (2002). The thick, dashed line denotes the weighting pattern resulting from a multiple-detector model fit to phase-opponency model hit and false-alarm rates for the narrowband and wideband N_0S_0 stimulus configurations.

stimuli than wideband stimuli. Also note that N_0S_0 predictions were better than N_mS_m predictions, which was expected, because weight patterns were derived in the N_0S_0 condition and the data were averaged across more subjects in the N_0S_0 condition than in the N_mS_m condition. This model was slightly less sensitive under the wideband condition ($d'=1.37$) than the narrowband condition ($d'=1.41$).

Incorporating energy present in channels outside a single critical bandwidth improved model-subject correlations with respect to the critical-band model, which used the energy at the output of a *single* 75-Hz bandwidth filter centered at 500 Hz. This improvement occurred for both wideband and narrowband stimuli. Although little stimulus energy existed outside the critical-band for narrowband stimuli, the reduced weights in the filters adjacent to the filter centered at 500 Hz must have been responsible for improving model predictions with respect to the critical-band model, and thus this information should not be ignored. Implications of the weight pattern are discussed below.

Unlike other models discussed in this study, which had only two free parameters (the slope and intercept of the psychometric function), the MD model incorporated either four or eight free parameters (the three narrowband or seven wideband filter weights with central weights fixed at one, and the slope and mean of the psychometric function). Therefore, it is possible that the increases in R^2 for the MD model with respect to the CB model could have resulted simply from the additional free parameters. To examine this issue, a MD model was fit to the data an additional 100 times, using randomly-created independent sets of 25 noise-alone and 25 tone-plus-noise waveforms within each of the 100 fits. Three filters were used for narrowband stimuli and seven filters were used for wideband stimuli. Central to the test was the fact that the energy at the output of the filter centered at the tone frequency was always computed from the *original sets* of noise waveforms, while energies at the adjacent filters were computed from the *independent sets* of random noise waveforms. The independent sets of random-noise waveforms were not related to those used in the experiment; therefore, any improvement in the fits of the MD model over that of the CB model for the random sets of waveforms must be explained by the fact that the MD model incorporated more free parameters. Using this analysis, R^2-values from the MD model were compared to the R^2 values from the 100 fits under narrowband and wideband, N_0S_0 and N_0S_π listening conditions.

It was found that for 397 out of 400 random-noise fits (for the average subject across all conditions) the R^2 values were below the R^2 value obtained when the MD model was fit using the original set of waveforms. This result suggests that the increases in R^2 for the MD model relative to the CB model was significant, and that the subjects had indeed used energy in the frequency regions outside the critical band as a basis for their decisions about the presence of the signal.

One might attempt to interpret the shape of the MD model's weighting function (Fig. 5) in terms of physiological mechanisms. For example, the combination of central positive and surrounding negative weights is compatible with a neural lateral-inhibitionlike mechanism that may be present in cells with characteristic frequencies (CFs) tuned to frequencies surrounding the 500-Hz stimulus. Consistent with this interpretation, Kopp-Scheinpflug *et al.* (2002) report the presence of broadly tuned inhibitory inputs to spherical bushy cells (SBCs) in the AVCN. They propose that these inputs could be responsible for increased frequency selectivity seen in SBCs. However, it is possible to obtain a weighting function including both negative and positive weights with a purely excitatory model. The dashed, black lines in Fig. 5 show the results of a multiple-detector model fit to a purely-excitatory model (described in detail in Sec. III F) in the N_0S_0 listening condition. Note that the overall shape is similar to that produced by the multiple-detector model fits to psychophysical data. The multiple-detector modeling results could predict about 49% of the variance in narrowband and about 65% of the variance in wideband detection patterns generated by the excitatory model. Thus, the shape of the weighting function cannot be used explicitly to identify excitatory vs inhibitory mechanisms. The negative weights for the filters adjacent to the filter centered at the tone frequency may be explained by phase differences (associated with filter shape) between the central filter and adjacent filters.

D. AN-count-based models

1. Methods

A physiologically-based energy model was tested using the Heinz *et al.* (2001b) AN model and a simple neural-count-based detection strategy. One motivation for testing this model was that the relatively broad tuning of AN fibers at high sound levels may explain the bandwidth effects in the psychophysical results. High-spontaneous-rate (HSR; 50 spikes/s) and low-spontaneous-rate (LSR; 1 spike/s) model fibers were tested. For LSR simulations, the parameters of the inner-hair-cell-to-auditory-nerve synapse were set to reproduce the dynamic range of low-frequency LSR

rate-level functions in response to tones at the fiber's characteristic frequency (Heinz, personal communication). The decision variable was the mean discharge rate (count) of the model fiber, either including or excluding the onset-response (first 5 ms), such that four models were considered. Models denoted as steady-state (SS) excluded the first and last 50 ms of the AN response (onset and offset response). Mean counts were computed from the AN model's rate function; no spikes were generated, and thus the internal noise generally associated with the Poisson properties of AN fibers was not included here.

2. Results and discussion

Results for the low-spontaneous-rate AN-count-based model are shown in Figs. 3 and 4 [L (low-spontaneous rate); L_{SS} (low-spontaneous-rate steady-state)], while high-spontaneous-rate AN count based models are shown in Fig. 4 only [H (high spontaneous rate); H_{SS} (high-spontaneous rate steady-state]. Predictions for HSR fibers, which saturate at low SPLs (i.e., have small dynamic ranges), accounted for only 7%–18% of the variance in the average subject's detection patterns and thus were not included in Fig. 3. HSR-fiber models produced d' values ranging from 0.47 to 0.56, well below the expected value of 1.41, due to effects of rate saturation.

For the narrowband results, the LSR model had $d' = 1.31$, and for wideband results, the LSR model had $d' = 1.24$. The LSR_{SS} model had lower sensitivity for both narrowband ($d' = 1.03$) and wideband ($d' = 1.01$) stimuli, indicating that information in the onset response of the fibers is useful for the detection process. Model LSR fibers predicted between 44% and 80% of the variance in the average subject's narrowband detection patterns and between 54%–69% of the variance in the average subject's wideband detection patterns.

In general, the predictions of LSR AN count-based models were similar to the CB model discussed in Sec. III B and the LSR_{SS} model made slightly poorer predictions. Model predictions based on HSR fibers were barely significantly better than chance at the $p < 0.05$ level and predictions based on HSR_{SS} fibers were insignificant in some conditions (see Fig. 4). In particular, the wider-than-critical-band tuning of AN fibers at high sound levels did not improve upon the performance of the critical-band model for describing detection patterns in response to wideband stimuli. The count-based detection strategy is essentially a peripheral transformation of the single-critical-band detector discussed earlier. However, unlike the critical-band model, the AN-count-based models are subject to saturation and rate-dependent variance (Colburn *et al.*, 2003). This characteristic poses problems for HSR fibers coding high-level stimuli such as those used here, or for coding stimuli with broad dynamic range (e.g., when the level is roved). No improvement was observed using model AN fibers over the CB or MD models. Thus, it was of interest to determine if temporal models for tone-in-noise detection could improve upon these predictions.

E. Envelope-slope model

A number of findings in the literature suggest that subjects can use cues other than energy for detection in noise, whether or not energy cues are available. For example, Richards and Nekrich (1993) found that listeners are able to perform a two-alternative forced-choice tone-in-noise detection task when no reliable level differences between noise-alone and tone-plus-noise stimuli are present. Richards and Nekrich (1993) also measured effects of energy on detection performance with separate level-discrimination tasks for noise-alone and tone-plus-noise stimuli. Performance in these conditions did not fully explain results in which level cues are present, suggesting the existence of cue(s) that do not rely on energy. One such cue is derived from temporal envelope fluctuations in the stimulus waveform and is known as the envelope-slope statistic (see Fig. 2, ES).

1. Methods

We investigated envelope cues using the following modified version of the Zhang (2004) envelope-slope statistic, which was adapted from the statistic derived by Richards (1992),

$$E_{SL} = \frac{\sum_t |x[t - \Delta t] - x[t]|}{\sum_t x[t]} \tag{7}$$

where $x[t]$ is the Hilbert envelope of the output of a 4th-order linear gammatone filter centered at the stimulus frequency (75-Hz ERB), and Δt is the time step over which differences were computed (the sampling frequency). The envelope of the gammatone filter was computed using a Hilbert transform and then low-pass filtered (10th-order maximally-flat IIR filter with a cutoff of 250 Hz) to ensure that any fine structure was removed and that the model decision variable was based solely on envelope fluctuations. The statistic was normalized to remove overall level and duration effects. When the tone was added to the noise waveform, the envelope-slope statistic of the tone-plus-noise complex decreased relative to the envelope-slope statistic of the noise-alone waveform, indicating tone presence. (A tone has an envelope slope equal to 0.) This model used no additional parameters other than the slope and threshold of the ogive fit.

2. Results and discussion

The ES model was slightly more sensitive for wideband stimuli ($d' = 1.37$) than for narrowband stimuli ($d' = 1.31$). Figures 3 and 4 (ES) reveal that this decision variable[4] was significantly correlated with the psychophysical data across reproducible maskers for all conditions. Envelope-slope model predictions approached the critical-band model predictions for wideband stimuli. The envelope-slope model accounted for approximately 65% of the variance in the average subject's wideband detection patterns and approximately 60% of the variance in the average subject's narrowband detection patterns. All predictions were significant at the $p < 0.05$ level.

F. Phase-opponency model

1. Methods

We also tested a single-cell version of the phase-opponency model (Carney *et al.*, 2002), which is a physiologically-based model that takes advantage of the temporal fine structure in AN responses. A block diagram of the single-cell phase-opponency model is shown in Fig. 2 (PO). Two model auditory-nerve fibers were used (Heinz *et al.*, 2001b). The center frequencies of the two fibers were selected (459 and 542 Hz) such that they were symmetrically spaced around the 500-Hz target frequency and had phase responses that differed by 180° at the target frequency at the level used. These AN model outputs were cross correlated. When the tone was present, both AN fibers tended to synchronize with the 500-Hz tone; and the 180° phase shift between the fibers made it unlikely that discharges would occur simultaneously for the two fibers. That is, coincidences were less likely to occur when the target was present. The coincidence-detector output counts were approximated with Eq. (8), which was based on more general expressions for neural coincidence detector output [e.g., Eq. (2) in Colburn (1977), Eq. (6) in Colburn (1996); and Eq. (B1) in Heinz *et al.* (2001a)], specifically,

$$C_M = n_{\mathrm{fib}}^2 T_{\mathrm{CW}} \int_{T_{\mathrm{DUR}}} r_i(t,F_i) r_j(t,F_j) dt, \tag{8}$$

where C_M is the monaural coincidence-detector count, r_i and r_j are rate functions of AN fibers with differing CFs, n_{fib} is the number of AN fiber inputs for each CF, T_{DUR} is the interval over which the coincidence detector response is computed, and T_{CW} is the time window for coincidences. This equation assumes that the width of the time window T_{CW} is narrow compared with the rate of fluctuation in the rate functions in Eq. (8). As noted above, the current model did not generate discharge times; thus the internal noise associated with AN responses was not included. When the two onset probabilities were multiplied in Eq. (8), they exceeded realistic levels and did not produce decision variables correlated to subject data. Therefore, only the sustained portion of the response was used (i.e., the first and last 50 ms of the coincidence detector response were excluded). The $n_{\mathrm{fib}}^2 T_{\mathrm{CW}}$ term at the left of Eq. (8) accounted for the number of fibers at each CF incident on the coincidence-detector cell [Eq. (4.2) in Zhang, 2004] and for the coincidence-window duration [Eq. (6) in Colburn (1996)]. Ten identical fibers were incident on each model coincidence-detecting cell at each CF, and the coincidence-window duration was 20 μs (Carney *et al.*, 2002). Three parameters [AN center frequencies (2) and the number of fibers incident on each cell], were used in this model, in addition to the slope and threshold of the ogive fit.

2. Results and discussion

The PO model was less sensitive than the other models described here for narrowband ($d'=0.89$) and wideband ($d'=0.99$) conditions, as well as less sensitive than expected for a model with no internal noise. As shown in Figs. 3 and 4 (PO) this model was able to predict 50%–59% of the variance in the average subject's narrowband detection patterns, and about 60% of the variance in the average subject's wideband detection patterns.

The PO panels in Fig. 4 show that the phase-opponency model predicted less variance in the psychophysical data than the multiple-detector, envelope-slope and critical-band models. A population version of the PO model (Carney *et al.*, 2002) was also tested, but showed no improvement over the single-cell version for these conditions.

G. Effects of roving stimulus level in a 2-interval detection task

Model predictions were also considered under conditions in which stimulus levels were roved within trials in a 2-interval, 2-alternative forced-choice detection task using random maskers (i.e., not reproducible-noise maskers). Although no psychophysical data were obtained in this study for the roving-level condition, there is little doubt about the nature of these results and that they have important implications for models of detection.

Although both the CB and MD models were able to account for a large amount of variance in the reproducible-noise psychophysical data, these models (in their current implementations) did not correctly predict published thresholds in a roving-level task. For example, Kidd *et al.* (1989) reported psychophysical roving-level data that do not match threshold predictions for a critical-band model. The MD model requires that the negatively-weighted filters be excited in order to be robust to the roving-level paradigm. That is, in the case of a wideband stimulus, the random amount of energy contributed by the noise to the filter at the tone frequency can be effectively removed by adjacent filters that encompass only noise and have approximately equal but negative weights. However, if the stimulus bandwidth is narrow enough that negatively-weighted filters are effectively not stimulated, the MD model shows an increase in threshold that is similar to the critical-band model. Since the current form of the MD model incorporated only the weights derived in this work, it would not be expected to describe this roving-level experiment. The MD model could be improved to make better predictions using roving-level stimuli. That is, the detection mechanism could be modified to adjust either the number of filters and/or their bandwidths for each stimulus bandwidth, or adjust the weights of a fixed set of filters for each stimulus bandwidth, such that the scaled output of the negatively-weighted filters effectively cancels the noise energy at the filter centered at the tone frequency.

Counts of AN fibers are monotonically related to energy; therefore, like the critical band model, the AN models cannot successfully predict thresholds for a roving-level task. If the AN model were expanded into a population model (i.e., that incorporated some sort of weighted sum of AN outputs, analogous to the MD model), its ability to explain results for the roving-level task would involve the same constraints as the MD model (i.e., inaccurate predictions if the bandwidth of the noise was insufficient to excite negatively-weighted filters, and the requirement to readjust weights for each stimulus bandwidth), in addition to rate-saturation problems.

The ES model was unaffected by roving the stimulus level until very narrow noise bandwidths (10 Hz), for which there was an increase in threshold that was greater than that observed in psychophysical data (Kidd *et al.*, 1989).

Carney *et al.* (2002) showed that a phase-opponent detection mechanism can successfully explain results (within 3 dB) for the roving-level detection task for maskers with bandwidths of at least 300 Hz. Further, Zhang and Carney (2004) and Zhang (2004) have shown that a more general cross-frequency coincidence-detection model can explain thresholds in a narrowband condition with roving-level maskers. The general coincidence-detection model is capable of predicting detection thresholds within 3 dB of human listeners in conditions with a 32-dB rove and a 100-Hz stimulus bandwidth, and does so using only HSR fibers. Recall that HSR fibers cannot explain thresholds based on average discharge rate due to saturation, and that HSR fibers comprise the majority of AN fibers.

Thus, it appears that the multiple-detector, phase-opponency, and envelope-slope models are capable of for predicting results from roving-level tasks with wideband noise, however each of these models breaks down when the bandwidth of the stimulus becomes narrow [bandwidth $<$300 Hz for the single cell PO model, $<$75 Hz (1 critical band) for the MD model, and $<$10 Hz for the ES model].

IV. SUMMARY AND FUTURE DIRECTIONS

This study measured detection in monotic and diotic masking conditions with reproducible noise waveforms. Measurements showed that using either monotic or diotic conditions resulted in similar distribution of performance over the individual waveforms. This result adds to the evidence that these conditions involve the same decision processing.

Predictions of energy-based, auditory-nerve, envelope, and phase-opponent cross-frequency coincidence-detection models were evaluated by comparing model decision variables and subject detection patterns. Results showed that the multiple-detector model could explain the most variance in subject detection patterns followed by the critical-band model. The critical-band model, however, cannot explain performance in a roving-level paradigm. The multiple-detector model will require a mechanism to adapt its filters and/or filter weights to specific stimulus conditions in order to predict thresholds under conditions where the stimulus level is roved. The phase-opponency and envelope-based models, although explaining less variance in the subject's detection patterns than energy-based models, are robust to changes in stimulus level. In contrast to the temporal models, HSR AN-count-based physiological implementations of energy-based models saturate, preventing accurate prediction of detection patterns. LSR AN-count-based models accurately predicted subject detection patterns, but were unable to explain performance in a roving-level task. In general, these results suggest the need to consider alternative temporal models for tone-in-noise detection, as well as to pursue the multiple-detector model's potential to describe roving-level results. Ongoing studies designed to manipulate specific detection cues will help direct future temporal-modeling efforts. The energy-based multiple-detector model will also be explored further.

ACKNOWLEDGMENTS

This work was supported by NIDCD R01-01641 (L.H.C.), and R01 DC00100 (H.S.C.). Part of this work was supported by a Research Challenge grant from the Ohio Board of Regents (R.H.G.). We thank Susan Early for her editorial comments and Valerie Shalin and Julio Mateo for assistance with statistical analyses.

APPENDIX: PREDICTABLE VARIANCE FOR INDIVIDUAL-SUBJECT FITS

Extending the analysis presented in Sec. II B 7, it is possible to derive the expected proportion of predictable variance available when fitting models individually to each subject's data, as compared to fitting models to data that have been averaged across subjects, as is the emphasis in this paper. To obtain an estimate of the error variance that is separate from the waveform by subject interaction, the data were rearranged as follows: N_0S_0 and N_mS_m data were combined (to increase the number of trials in the analysis; this decision is justified because the differences between the two conditions are small), and the 192 trials obtained for each combination of waveform and bandwidth were randomly grouped into 12 blocks of 16 trials. Separate Waveform $\times$ Subject $\times$ Block random-effects ANOVAs were performed for each bandwidth. Working from the expected values of the Mean Squares, as described by Kirk (1995, p. 462), and combining terms appropriately, we obtained separate estimates for the variances associated with the effect of waveform (σ_W^2), subject (σ_S^2), block (σ_B^2), the interaction between waveform and subject (σ_{WS}^2), the interaction between waveform and block (σ_{WB}^2), the interaction between subject and block (σ_{SB}^2), and the interaction among waveform, subject, and block combined with error ($\sigma_{\mathrm{WSB}}^2+\sigma_\varepsilon^2$). This procedure lead to some negative variance estimates (which, for our data, were small in magnitude) that were set to zero, as is common practice (e.g., Maxwell and Delaney, 2004, p. 487). For the case in which the models were fit to data averaged across subjects, as in the current study, the proportion of predictable variance is given by

$$\frac{\sigma_P^2}{\sigma_T^2}=\frac{\sigma_W^2}{\sigma_W^2+\frac{\sigma_{\mathrm{WS}}^2}{N_S}+\frac{\sigma_{\mathrm{WB}}^2}{N_B}+\frac{\sigma_{\mathrm{WSB}}^2+\sigma_\varepsilon^2}{N_SN_B}}, \tag{A1}$$

where σ_P^2 is the predictable variance in the average detection pattern,σ_T^2 is the total variance in the average detection pattern, N_S is the number of subjects, and N_B is the number of 16-trial blocks combined to estimate each value in the detection pattern. For the case in which the models were fit separately to the data of individual subjects, predictable variance is given by

TABLE VI. Estimated proportion of predictable variance in $P(Y|W)$ values averaged across the N_mS_m and N_0S_0 stimulus configurations under both narrow and wide stimulus bandwidths. Results are presented for Eq. (A1) which estimates the proportion of predictable variance for the average subject, and equation Eq. (A2), which estimates the proportion of predictable variance for the case in which models are fit to the data of individual subjects. Results are also presented for Eq. (4) using the subject data that was combined across the N_mS_m and N_0S_0 stimulus configurations (see text).

Bandwidth	Eq. (A1)	Eq. (A2)	Eq. (4) Combined
NB	0.944	0.989	0.945
WB	0.982	0.991	0.982

$$\frac{\sigma_P^2}{\sigma_T^2} = \frac{\sigma_W^2 + \sigma_S^2 + \sigma_{\mathrm{WS}}^2}{\sigma_W^2 + \sigma_S^2 + \sigma_{\mathrm{WS}}^2 + \frac{\sigma_{\mathrm{WB}}^2}{N_B} + \frac{\sigma_{\mathrm{SB}}^2}{N_B} + \frac{\sigma_{\mathrm{WSB}}^2 + \sigma_\varepsilon^2}{N_B}}, \quad \text{(A2)}$$

where σ_P^2 is the predictable variance in the individual detection patterns, and σ_T^2 is the total variance in the individual detection patterns.

The estimated proportions of predictable variance for Eqs. (A1) and (A2) are shown in Table VI (columns 1 and 2) for the two bandwidth conditions, with $N_S=4$ and $N_B=12$. Note that both Eq. (A1) and Eq. (4) estimate the proportion of predictable variance for the case when models are fit to data averaged across subjects. The values shown for Eq. (A1) (Table VI, column 1) are slightly larger than those shown for Eq. (4) (Table V, column 4). This difference occurred because the analyses in Table VI were based on 192 trials per waveform (i.e., the data were combined across the N_mS_m and N_0S_0 conditions), whereas the analyses in Table V were based on 96 trials per waveform. To allow a direct comparison between Eq. (A1) and Eq. (4), Eq. (4) was applied to the results of data that had been combined across binaural conditions, such that 192 trials per waveform occurred; these results are shown in Table VI (column 3) and are in agreement with the results of Eq. (A1).

Comparing results from Eq. (A1) and Eq. (A2) (shown in the first two columns of Table VI), we see, as expected, that fitting the models to individual subjects increased the proportion of predictable variance. This was particularly true for the narrow bandwidth case, because there was greater across-subject variability for that stimulus condition.

[1]One model not considered here is that of Dau *et al.* (1996). We were unable to meaningfully apply their model's structure to our single-interval reproducible-noise task.

[2]Although no MLD was found in this study, others have found an $N_mS_m - N_0S_0$ MLD, presumably only at low masker levels. Dolan (1968) observed increasing $N_mS_m - N_0S_0$ MLD for 150-Hz and 300-Hz tones as the spectrum level of a 1000-Hz low-pass filtered noise masker was lowered from 65 to 20 dB SPL. MLDs were less than 1 dB for the 65 dB SPL spectrum level noise and up to 4 dB for the 20 dB SPL spectrum-level noise. Diercks and Jeffress (1962) found a small $N_mS_m - N_0S_0$ MLD (2.8 dB) for a 250-Hz tone, but the masker levels for which this result was achieved were unpublished. Shaw *et al.* (1947) also found an $N_mS_m - N_0S_0$ MLD of 5.8 dB; however masker levels were also unpublished in that work. It is assumed that the latter two studies used relatively low noise spectrum levels in accordance with the results of Dolan (1968). Thus, if this experiment were repeated at lower masker levels that yielded MLDs, it is possible that detection patterns would differ between the two conditions.

[3]We also computed internal-to-external noise ratios for the Evilsizer *et al.* (2002) study. In that study, N_0S_π noise ratios were larger than diotic noise ratios, with the exception of S7, who had the largest estimated noise ratios (ranging from 2.6 to 3.6 in narrowband N_0S_0 and N_0S_π conditions, respectively) and had the lowest first-half, last-half correlations [ranging from 0.52 to 0.76; see Table I of Evilsizer *et al.* (2002)]. This finding indicates that this subject may have changed strategies, thereby increasing internal noise estimates.

[4]An envelope-slope based model was also considered with the basilar-membrane and inner-hair-cell portions of the Heinz *et al.* (2001b) AN model. Results are not presented here because predictions were not significantly different from the envelope-slope model incorporating a gammatone filter. A model based on the standard deviation of the envelope (Richards, 1992) was also considered, but was not included here because predicted detection patterns were not correlated to the psychophysical data.

Ahumada, A., and Lovell, J., (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.

Ahumada, A., Marken, R., and Sandusky, A., (**1975**). "Time and frequency analyses of auditory signal detection," J. Acoust. Soc. Am. **57**, 385–390.

Blodgett, H. C., Jeffress, L. A., and Taylor, R. B., (**1958**). "Relation of masked threshold to signal duration for various interaural phase combinations," Am. J. Psychol. **71**, 283–290.

Bruning, J. L., and Kintz, B. L., (**1968**). *Computational Handbook of Statistics* (Scott, Foresman and Company, Glenview, IL).

Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S., (**2002**). "Auditory phase-opponency: A temporal model for masked detection at low frequencies," Acta. Acust. Acust. **88**, 334–347.

Colburn, H. S., (**1977**). "Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise," J. Acoust. Soc. Am. **61**, 525–533.

Colburn, H. S., (**1996**). "Computational models of binaural processing," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer, New York), pp. 332–400.

Colburn, H. S., Carney, L. H., and Heniz, M. G., (**2003**). "Quantifying the information in auditory-nerve responses for level discrimination," J. Assoc. Res. Otolaryngol. **4**, 294–311.

Costalupes, J. A., Young, E. D., and Gibson, D. J., (**1984**). "Effects of continuous noise backgrounds on rate response of auditory-nerve fibers in cat," J. Neurophysiol. **51**, 1326–1344.

Dau, T., Püschel, D., and Kohlrausch, A., (**1996**). "A quantative model of the 'effective' signal processing in the auditory system," J. Acoust. Soc. Am. **99**, 3615–3622.

Diercks, J. K., and Jeffress, L. A., (**1962**). "Interaural phase and the absolute threshold for tone," J. Acoust. Soc. Am. **34**, 981–984.

Dolan, T. R., (**1968**). "Effect of masker spectrum level and masking-level differences at low signal frequencies," J. Acoust. Soc. Am. **44**, 1507–1512.

Egan, J. P., (**1965**). "Masking-level differences as a function of interaural disparities in intensity of signal and noise," J. Acoust. Soc. Am. **38**, 1043–1049.

Egan, J. P., Lindner, W. A., and McFadden, D., (**1969**). "Masking-level differences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.

Evilsizer, M. E., Gilkey, R. H., Mason, C. R., Colburn, H. S., and Carney, L. H., (**2002**). "Binaural detection with narrowband and wideband reproducible noise maskers: I. Results for human," J. Acoust. Soc. Am. **111**, 336–345.

Fletcher, H., (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Gilkey, R. H., Robinson, D. E., and Hanna, T. E., (**1985**). "Effects of masker waveform and signal-to-masker phase relation on diotic and dichotic masking by reproducible noise," J. Acoust. Soc. Am. **78**, 1207–1219.

Gilkey, R. H., and Meyer, T. A., (**1987**). "Modeling subject responses in a reproducible-noise masking task," J. Acoust. Soc. Am. **82**, 92–93.

Gilkey, R. H., and Robinson, D. E., (**1986**). "Models of auditory masking: A molecular phsychophysical approach," J. Acoust. Soc. Am. **79**, 1499–1510.

Glasberg, B. R., and Moore, B. J. C., (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Green, D. M., (**1984**). "Profile analysis: A different view of auditory intensity discrimination," Am. Psychol. **38**, 133–142.

Green, D. M., (**1988**). *Profile Analysis* (Oxford University Press, New York).

Hall, J. W., Haggard, M. P., and Fernandes, M. A., (**1984**). "Detection in

noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. **76**, 50–56.
Hawkins, J. E., and Stevens, S. S., (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.
Heinz, M. G., Colburn, H. S., and Carney, L. H., (**2001a**). "Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection," J. Acoust. Soc. Am. **110**, 2065–2084.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H., (**2001b**). "Auditory-nerve Model for Predicting Performance Limits of Normal and Impaired Listeners," ARLO **2**, 91–96.
Hirsh, I. J., and Burgeat, M., (**1958**). "Binaural effects in remote masking," J. Acoust. Soc. Am. **30**, 827–832.
Isabelle, S. K., and Colburn, H. S., (**1991**). "Detection of tones in reproducible narrow-band noise," J. Acoust. Soc. Am. **89**, 352–359.
Jeffress, L. A., (**1968**). "Mathematical and electrical models of auditory detection," J. Acoust. Soc. Am. **44**, 187–203.
Kiang, N. Y., and Moxon, E. C., (**1974**). "Tails of tuning curves of auditory-nerve fibers," J. Acoust. Soc. Am. **55**, 620–630.
Kidd, G., (**1987**). "Auditory discrimination of complex sounds: The effects of amplitude perturbation on spectral shape," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NY), pp. 16–25.
Kidd, G., Mason, C. R., Brantley, M. A., and Owen, G. A., (**1989**). "Roving-level tone-in-noise detection," J. Acoust. Soc. Am. **86**, 1310–1317.
Kirk, R. E., (**1995**). *Experimental Design: Procedures for the Behavioral Sciences*, 3rd ed. (Brooks/Cole, Pacific Grove, CA).
Kopp-Scheinpflug, C., Dehmel, S., Dörrscheidt, G. J., and Rübsamen, R., (**2002**). "Interaction of excitiation and inhibition in anteroventral cochlear nucleus neruons that receive large endbulb synaptic endings," J. Neurosci. **22**, 11004–11018.
Langhans, A., and Kohlrausch, A., (**1992**). "Differences in auditory performance between monaural and diotic conditions. I. Masked thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.
Levitt, H., (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
MacMillan, N. A., and Creelman, C. D., (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, New York).
Maxwell, S. E., and Delaney, H. D., (**2004**). *Designing Experiments and Analyzing Data: A Model Comparison Perspective* (Lawrence Erlbaum Associates, Mahwah, NJ), pp. 469–517.
Neff, D. L., and Callaghan, B. P., (**1988**). "Effective properties of mulit-component simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1888.
Patterson, R. D., (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.
Pfafflin, S. M., (**1968**). "Detection of Auditory Signal in Restricted Sets of Reproducible Noise," J. Acoust. Soc. Am. **43**, 487–490.
Pfafflin, S. M., and Mathews, M. V., (**1966**). "Detection of auditory signals in reproducible noise," J. Acoust. Soc. Am. **39**, 340–345.
Rhode, W. S., (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique," J. Acoust. Soc. Am. **49**, 1218–1231.
Richards, V. M., (**1992**). "The detectability of a tone added to narrow bands of equal energy noise," J. Acoust. Soc. Am. **91**, 3424–3425.
Richards, V. M., and Nekrich, R. D., (**1993**). "The Incorporation of level and level-invariant cues for the detection of a tone added to noise," J. Acoust. Soc. Am. **94**, 2560–2574.
Ruggero, M. A., (**1973**). "Responses to noise of auditory-nerve fibers in the quirrel monkey," J. Neurophysiol. **36**, 569–587.
Schalk, T. B., and Sachs, M. B., (**1980**). "Nonlinearities in auditory-nerve fiber responses to bandlimited noise," J. Acoust. Soc. Am. **67**, 903–913.
Sever, J. C., and Small, A., M., (**1979**). "Binaural critical-bands," J. Acoust. Soc. Am. **66**, 1343–1350.
Shannon, R. V., (**1976**). "Two-tone unmasking and suppression in a forward-masking situation," J. Acoust. Soc. Am. **59**, 1460–1470.
Shaw, W. A., Newman, E. B., and Hirsh, I. J., (**1947**). "The difference between monaural and binaural thresholds," J. Exp. Psychol. **37**, 229–242.
Siegel, R. A., and Colburn, H. S., (**1989**). "Binaural processing of noisy stimuli: Internal/external noise ratios under diotic and dichotic stimulus conditions," J. Acoust. Soc. Am. **86**, 2122–2128.
Spiegel, M. F., and Green, D. M., (**1981**). "Two procedures for estimating internal noise," J. Acoust. Soc. Am. **70**, 69–73.
Zhang, X., (**2004**). "Cross-frequency coincidence detection in the processing of complex sounds," Ph.D. dissertation, Boston University.
Zhang, X., and Carney, L. H., (**2004**). "Models for monaural cross-frequency coincidence-detectors in tone-in-noise detection tasks," Abstract 842, Association for Research in Otolaryngology.
Zheng, L., Early, S. J., Mason, C. R., Idrobo, F., Harrison, J. M., and Carney, L. H., (**2002**). "Binaural detection with Narrowband and Wideband Reproducible Noise Maskers. II. Results for Rabbit," J. Acoust. Soc. Am. **111**, 346–356.

Sensitivity to isolated and concurrent intensity and fundamental frequency increments by cochlear implant users under natural listening conditions[a)]

Cheryl F. Rogers, Eric W. Healy,[b)] and Allen A. Montgomery
Department of Communication Sciences and Disorders, The Arnold School of Public Health, University of South Carolina, Columbia, South Carolina 28208

(Received 20 April 2005; revised 22 December 2005; accepted 28 December 2005)

Sensitivity to acoustic cues in cochlear implant (CI) listening under natural conditions is a potentially complex interaction between a number of simultaneous factors, and may be difficult to predict. In the present study, sensitivity was measured under conditions that approximate those of natural listening. Synthesized words having increases in intensity or fundamental frequency (F0) in a middle stressed syllable were presented in soundfield to normal-hearing listeners and to CI listeners using their everyday speech processors and programming. In contrast to the extremely fine sensitivity to electrical current observed when direct stimulation of single electrodes is employed, difference limens (DLs) for intensity were larger for the CI listeners by a factor of 2.4. In accord with previous work, F0 DLs were larger by almost one order of magnitude. In a second experiment, it was found that the presence of concurrent intensity and F0 increments reduced the mean DL to half that of either cue alone for both groups of subjects, indicating that both groups combine concurrent cues with equal success. Although sensitivity to either cue in isolation was not related to word recognition in CI users, the listeners having lower combined-cue thresholds produced better word recognition scores. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2167150]

PACS number(s): 43.66.Fe, 43.64.Me, 43.71.Es [JHG] Pages: 2276–2287

I. INTRODUCTION

It is likely that sensitivity to both intensity and fundamental frequency (F0) is important for the perception of complex spectro-temporal signals such as speech. Variations in intensity over time provide temporal information capable of encoding a number of important speech cues. The difference limen (DL) for detecting changes in the intensity of a broadband noise is approximately 0.5 to 1 dB for normal-hearing (NH) listeners across a wide range of overall levels, but larger near threshold (Miller, 1947). Early experiments employing electrical stimulation of human (Shannon, 1983) and animal subjects (Pfingst, Burnett, and Sutton, 1983) showed both a trend for larger DLs at low levels, and smaller DLs for electrical than for acoustic stimulation. When expressed as Weber fractions, sensitivity to changes in electrical current can be extraordinarily fine. Zeng (2004) describes cases in which a cochlear implant (CI) user can reliably produce an electrical-stimulation DL of 0.08 dB—a value far below that detectable in normal hearing. More generally, the DL measured under direct electrical stimulation is roughly one order of magnitude smaller than the acoustic DL, and this difference has been related to the lack of cochlear compression in electrical stimulation (Zeng, 2004, also see Nelson *et al.*, 1996). However, this extremely fine sensitivity is offset by the small dynamic range of most CI users. Whereas the dynamic range of NH is roughly 120 dB (cf. Bacon, 2004), the difference between threshold and maximum loudness in electrical stimulation has been shown to be approximately 10 to 30 dB (Zeng and Shannon, 1994, 1999). This combination of fine sensitivity and reduced dynamic range combines to produce discriminable steps that number roughly 5 to 45 in CI users (Nelson *et al.*, 1996; Zeng and Shannon, 1999; Kreft, Donaldson, and Nelson, 2004), compared to 50 to 200 in normal hearing (Viemeister and Bacon, 1988).

A number of factors have been shown to affect intensity discrimination for stimulation of single electrodes with electrical pulse trains. Discrimination may be affected by pulse rate (Kreft, Donaldson, and Nelson, 2004), electrode configuration (Drennen and Pfingst, 2005), proximity to the modiolus (Cohen, Saunders, and Clark, 2001), and electrode position (Nelson *et al.*, 1996). Nelson *et al.* (1996) reported large variability in the number of discriminable loudness steps within a group of eight CI users. Although some relationship between the size of the discriminable step and threshold or dynamic range was observed, this relation did not hold for different electrode positions within the same subject. More importantly, the number of steps was not simply related to the absolute size of the electrical dynamic range. It has also been found that characteristics of the stimulus can affect discrimination. Wojtczak, Donaldson, and Viemeister (2003) observed smaller DLs for increments applied to continuous carriers relative to those of gated carriers for some subjects and level conditions. Because the meaningful

[a)]Portions of this work were presented in "Sensitivity to acoustic speech cues in acoustic and electric hearing," Proceedings of the 28th Midwinter Research Meeting of the Association for Research in Otolaryngology, New Orleans, Louisiana, February, 2005, and "Sensitivity to combined frequency and amplitude speech cues by cochlear implant users," Proceedings of the 149th Meeting of the Acoustical Society of America, Vancouver, British Columbia, May, 2005.

[b)]Electronic mail: ewh@sc.edu

 0001-4966/2006/119(4)/2276/12/$22.50

intensity variations that occur in speech are often relative to a baseline speech level, these results may be more relevant to everyday signals than those from the gated-carrier conditions usually employed.

The F0 of an utterance carries important information regarding speaker gender, identity, and intent. It is the means by which intonation is encoded, and it is one of the cues signifying lexical stress (cf. Cutler, 2005; Vaissiere, 2005). Normal-hearing listeners generally produce very small F0 DLs when all harmonics exist and both place and temporal cues are present. Classic work demonstrated that the F0 DL for steady-state synthesized vowels is on the order of 0.3 to 0.5 Hz (Flanagan and Saslow, 1958). However, when listeners are forced to rely on temporal cues alone, either by removing the lower resolved harmonics (Cullen and Long, 1986; Houtsma and Smurzynski, 1990) or by employing sinusoidally amplitude modulated (SAM) noise (Formby, 1985), DLs increase by roughly one order of magnitude.

Because the spectral representation provided by the CI electrode array is quite poor, pitch discrimination in these users may depend more strongly upon temporal cues than in normal hearing. Early experiments showed that CI recipients could discriminate the pitch of pulse trains having different repetition rates up to roughly 300 Hz (Shannon, 1983; Tong and Clark, 1985; Townshend *et al.*, 1987). Recent estimates indicate that the DL for rate discrimination (Zeng, 2002) or sinusoidal frequency modulation (FM) detection (Chen and Zeng, 2004—approximately 20 Hz at a standard frequency of 100 Hz) is roughly one order of magnitude larger for CI listeners under direct stimulation relative to NH listeners. Although modern CI devices do not employ pulse trains of varying repetition rate, these values are meaningful because some similarity has been observed between the pitch elicited by repetition rate and that elicited by fixed-rate pulse trains having different SAM rates (McKay, McDermott, and Clark, 1994). Thus, NH listeners have available both spectral and temporal cues to pitch, whereas CI users rely primarily on temporal cues. This is evidenced by the similarity in performance across NH and CI listeners under conditions in which spectral cues are removed and only temporal cues are available.

As with intensity discrimination, a number of factors can influence sensitivity to changes in F0. Although the DL for NH listeners hearing steady-state vowels is quite small, this value increases by roughly a factor of ten when F0 rises or falls in linear fashion over time (Klatt, 1973). Geurts and Wouters (2001) examined F0 discrimination in conditions that more closely resembled those of normal CI listening. The stimuli were synthesized vowels having F0 coded by modulation rate of amplitude-modulated pulse trains. Although sound-field presentation was not employed, the study differed from others by employing direct stimulation of all electrodes in interleaved fashion. The authors found discrimination of a 400-ms steady-state vowel at 150 Hz to average approximately 12 Hz (8%). It was also found that multiple-electrode stimulation, which is more representative of natural CI listening, allowed detection of modulation at smaller depths when compared to stimulation of individual electrodes.

Studies designed to measure discrimination sensitivity in CI users have generally employed direct stimulation of single electrodes using simple pulsatile stimuli. Although essential for isolating particular aspects or stages of processing, direct stimulation bypasses much of the processing that would normally take place in the CI reception of sound. Further, results obtained using simple pulsatile stimuli must be extrapolated to make assumptions about sensitivity to complex spectro-temporal signals such as speech. In sharp contrast, sensitivity to acoustic speech cues by CI listeners under natural listening conditions is a potentially complex interaction between a number of factors. One set of influences is attributable to the CI device and may include pulse rate, electrode configuration, electrode position, proximity to the modiolus, single-versus multi-electrode stimulation, etc. Another set of factors is related to the complexity of the stimulus. Discrimination of various cues may be influenced by nonstationary pitch and/or intensity, and duration of the target segment. Further, despite the large number of studies that have examined sensitivity to individual cues in isolation, little work has examined sensitivity under conditions that simulate the simultaneous frequency and intensity changes that occur in speech.

In the present study, sensitivity to increments in intensity and F0 was examined under pseudonatural listening conditions. The stimuli were synthesized words in which highly controlled manipulations of the cues existed, but these changes had the spectral and temporal characteristics of those that normally occur in speech. In addition, subjects were tested in the free field of a calibrated studio monitor using their everyday clinical speech processors to examine sensitivity when the CI device is performing in its normal mode from microphone to electrode array.

The goals of the present study are (1) to compare sensitivity to increments in intensity and F0 across NH and CI listeners under conditions that more closely approximate those of natural listening, (2) to determine if sensitivity is enhanced when concurrent intensity and F0 increments are present, (3) to determine if this potential enhancement is similar in magnitude across NH and CI listener groups, (4) to examine the relation between sensitivity to intensity increments, F0 increments, and concurrent F0 and intensity increments, to word recognition in CI users.

II. EXPERIMENT 1. SENSITIVITY TO INDIVIDUAL CUES

A. Method

1. Subjects

Two groups of subjects participated. The CI group was composed of six individuals ranging in age from 30–70 years. Five were fitted with the Nucleus 24—Esprit 3G device and one (CI2) was fitted with the Clarion II—PSP device. The subjects with the Nucleus device all used the ACE strategy, monopolar (MP1+2) stimulation, a Q value of 20, and a stimulation rate of 900 Hz/channel. The subject with the Clarion device used the HiRes-P strategy with a stimulation rate of 5156 Hz/channel. Table I provides additional information. As the speech recognition scores reflect,

TABLE I. Characteristics of CI participants.

Subject	Gender	Age	Age Dx	Etiology	Implanted ear	Number maxima	Number electrodes	Freq. table	Recognition percent			
									Wrds	Phonms	Sent quiet	Sent noise
CI 1	F	30	3	Unknown	L	9	20	7	54	84	92	85
CI 2	M	47	22	Unknown	L	—	16	—	58	74	NA	NA
CI 3	M	59	58	Auto-immune	R	8	18	8	80	88	100	90
CI 4	M	70	31	Unknown	L	8	20	7	76	87	96	94
CI 5	F	57	32	Unknown	R	8	20	7	64	80	98	92
CI 6	M	47	26	Meniere's	L	8	20	7	82	89	98	88

the CI users were all relatively good performers. The second group was composed of five individuals with NH [audiometric thresholds of 20 dB HL or better at octave frequencies between 250 and 8000 Hz (ANSI, 1996)].[1] All were native speakers of American English and were paid for their participation. Only one subject (NH4) had previous experience in psychoacoustic tasks.

2. Stimuli and apparatus

The stimuli consisted of the synthesized words, "potato," "sufficient," and "allowance" (Words 1–3). These words have a middle stressed syllable that is embedded in different phonetic environments. Productions of each word were made by a male speaker having a standard American dialect, and acoustic measurements of these productions were entered into a Klatt-based speech synthesizer operating at 20 kHz sampling. Each of the resulting synthesized words was clearly identifiable and even allowed NH listeners familiar with the speaker to recognize his voice. Figure 1 displays the amplitude envelopes of the three synthesized words. The

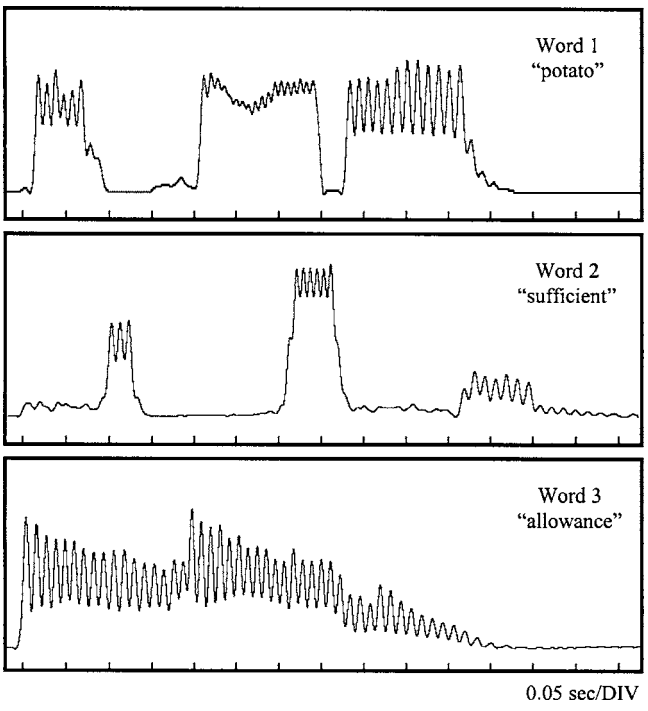


FIG. 1. Amplitude envelopes of the three stimulus words employed.

standard durations of the middle vowels for Words 1–3 were 140, 70, and 150 ms, respectively. Additional versions of each stimulus word were also created in which the middle vowel had systematic increases in either intensity or F0.

a. Intensity increments. In this series, the intensity of the middle vowel was increased in 0.25-dB steps. Because the speech synthesis application provided an intensity resolution of only 1 dB, stimuli having 1-dB steps were synthesized, and these files were scaled using a waveform editor (Adobe Audition) to obtain the desired 0.25-dB steps. These modifications were performed at zero crossings and included a 5 -ms linear rise/fall to eliminate clicks or other artifacts that could potentially aid judgments. Considerable care was taken to ensure that desired specifications were met and that the stimuli varied only in the manner intended. Level measurements (Larson Davis 800B) of the transduced middle syllables, played in loop mode, indicated that all intensities were within 0.1 dB of the desired value.

b. Fundamental-frequency increments. This series was created by increasing the F0 of the middle syllables in 1 Hz steps, up to 40 Hz, using the speech synthesizer. Prior to modification, the F0 of the middle syllables for Words 1–3 was 132 (±8), 125 (±1), and 91 (±5) Hz, respectively. To assess the accuracy of the F0 modifications, the frequency of each increment step was measured using the waveform editor. Although the speech synthesizer allowed F0 to be specified at high resolution, measurements of the output waveforms indicated that the actual frequencies varied in some cases considerably from the specified values. These actual F0 values were determined using the waveform editor and measurements across the syllable at sample-point resolution, which allowed measurement within 0.03 to 0.2 Hz across conditions. By manipulating the synthesis parameters and remeasuring the resultant sound files, F0-increment stimuli were created for each word that were within 0.3 Hz of the target value for each step in the series up to 15 Hz and within 0.7 Hz for steps above 15 Hz. There was no cumulative drift across the F0 increment values. Fundamental frequency DLs were calculated based upon both the nominal and the measured F0 increment values to examine the influence of these small deviations.

Previous research involving synthesized speech has indicated that a change in F0 may be accompanied by a change in intensity (e.g., Kewley-Port and Watson, 1994), and similar artifacts were observed here. They were relatively small

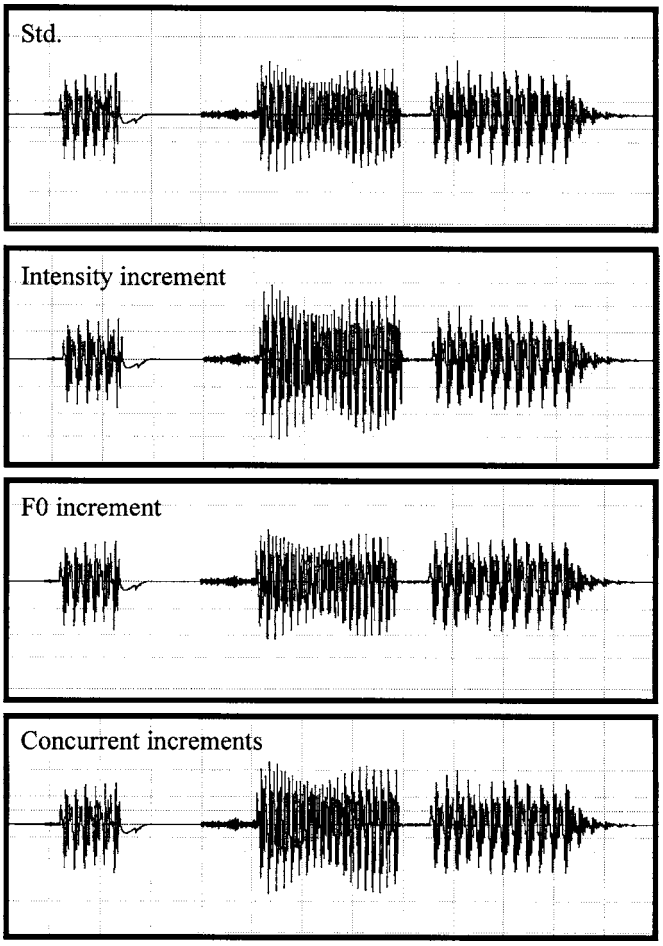


FIG. 2. Waveforms for the stimulus word "potato" having the increments indicated.

at F0 increases of 1–3 Hz, but could be as great as 1.5 dB at the larger frequency increments for all three stimulus words. To correct for these artifacts, each frequency increment stimulus was processed to ensure that its amplitude envelope matched exactly that of the corresponding standard stimulus. Using MATLAB, the envelopes for both the standard and the increment were extracted via full-wave rectification and low-pass filtering at 50 Hz. The waveform of each frequency increment stimulus was then scaled by multiplying it with a correction vector consisting of the ratio between the standard and the particular frequency increment envelopes. The resulting stimuli had intensities within 0.1 dB of that of the standard stimulus for F0 increments up to 10 Hz and within 0.3 dB of the standard for increments up to 40 Hz.

It was also observed that the manipulation of F0 in the middle syllable affected the onset timing of the subsequent syllable for Words 1 and 2. Therefore, the final syllable for each of the F0 increment stimuli for these words was replaced with that from the standard version. This process was completed using the waveform editor by replacing the final syllable at a particular fixed sample value in the center of the closure of the /t/ for Word 1 ("potato") or the low-intensity aperiodic phoneme /f/ for Word 2 ("sufficient") using a 5-ms crossfade to ensure a smooth transition.

Waveforms for stimulus Word 1 ("potato") are shown in Fig. 2. The upper panel shows the standard synthesized word, and the two panels below it show the stimulus having an increment in either intensity or F0. Throughout the course of stimulus preparation, a pair of individuals having experience preparing psychoacoustic stimuli examined each stimulus carefully and found them to be free of clicks or other artifacts that could possibly aid performance. The digital files were played back from a PC (Edirol UA-5 D/A converter) and presented using an amplified loudspeaker (Mackie HR824) having a flat frequency response (±1.5 dB, 39 Hz–20 kHz). Each standard stimulus was set to play back at a slow-response rms peak level of 70 dB SPL measured 1 m directly in front of the speaker.

3. Procedure

Subjects were administered the Consonant Nucleus Consonant test (CNC, Peterson and Lehiste, 1962) to assess identification of monosyllabic words and their constituent phonemes. The 50-item test was administered in soundfield at 65 dB SPL within a double-walled sound booth. Results are presented in Table I. Also provided are scores for the Hearing in Noise Test (HINT, Nilsson, Soli, and Sullivan, 1994) sentences presented both in quiet and with the standard noise background at +10 dB S/N (CI2 was unavailable for HINT testing).

Difference limens (DLs) were measured for each isolated cue for each of the three stimulus words. Presentation of conditions and words was random, with the exception that listeners heard all three words in either the intensity or F0 condition before advancing to hear all three words in the other condition. Thresholds were obtained using an adaptive three-interval, two alternative forced-choice (3I-2AFC) paradigm with a two-down, one-up decision rule to track the 71% correct point on the psychometric function (Levitt, 1971). The standard always appeared in the first interval and again in either the second or third. The increment step appeared randomly in either the second or third interval and 500 ms of silence separated each interval. A track began with the step stimulus clearly distinguishable from the standard. Subjects selected the different stimulus from the final two intervals by responding on a computer keyboard. A block of trials was terminated after 12 reversals, and the mean of the last eight was taken as the threshold estimate. The step sizes were 2 dB or 3 Hz prior to the fourth reversal, and 0.25 dB or 1 Hz thereafter.

A minimum of three threshold estimates were collected in each condition. Estimates comprising a mean were collected across different days, with each session lasting one to two hours. Testing took place with the subject seated 1 m directly in front of the calibrated loudspeaker within a double-walled sound booth. Trial-by-trial feedback was provided using a computer monitor located adjacent to the speaker. Presentation of the sound files in adaptive fashion and collection of responses were performed using custom software written in MATLAB. All subjects received a minimum of five hours of practice prior to data collection.

Although a criterion for inclusion of intensity threshold estimates based on the standard deviation (SD) of the reversals was set at 1.5 dB, no estimate from the highly-practiced subjects exceeded this value. Further, 89% of the adaptive

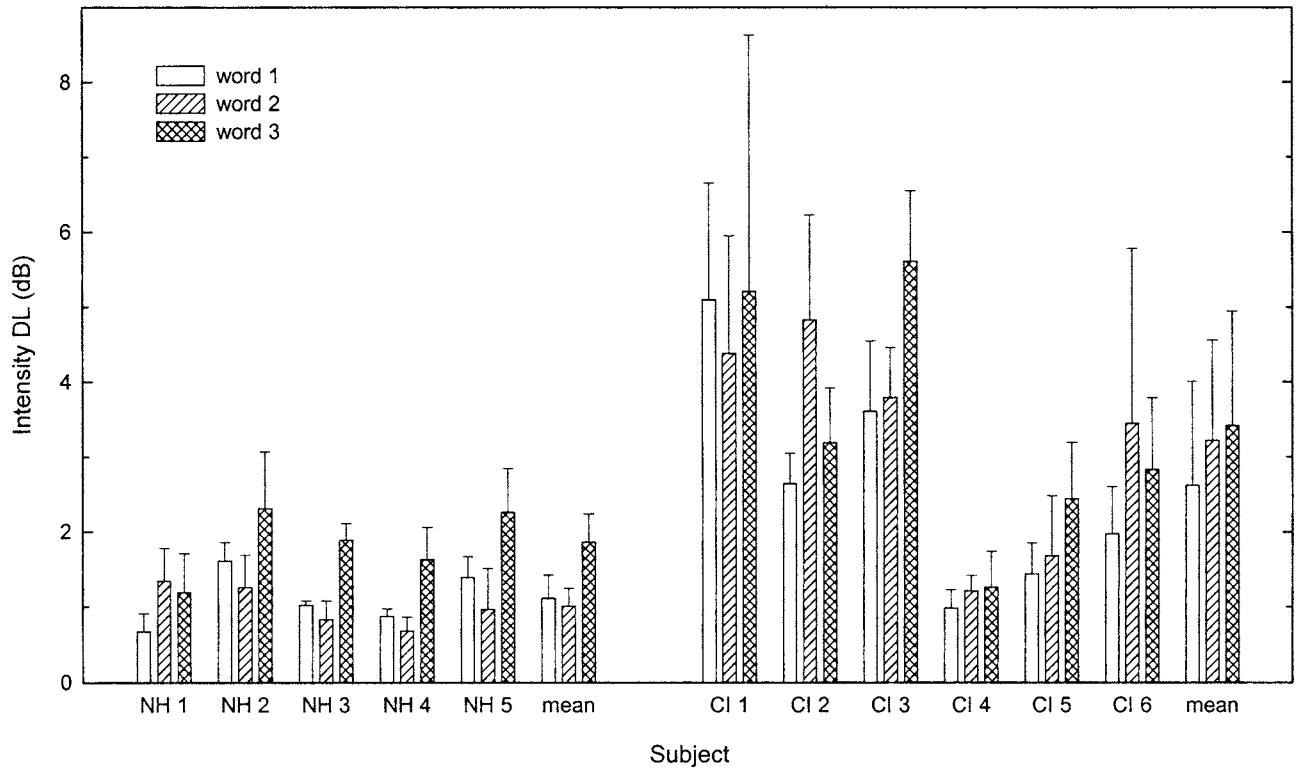


FIG. 3. Difference limens for individual subjects hearing synthesized words having an intensity increment in the middle syllable.

tracks had a SD below 0.75 dB. For the NH listeners, the frequency estimate criterion was set to 3 Hz, but no estimate exceeded this value. Further, 96% of the estimates had a SD below 2 Hz. For the CI listeners, the SD criterion defining an acceptable estimate was increased to 6 Hz and estimates not conforming to this (9% of the total) were discarded and replaced. The majority (84%) of the estimates following replacement had a SD below 3 Hz.

The CI subjects completed the experiment using their usual clinical processor, map, and strategy. It was important to ensure that the stimuli fell within the linear input range of the device and were not affected by the limiting of input signals to a maximum (C) output current level. To ensure this operation and still provide stimuli having levels that approximated those of normal speech, autosensitivity was disabled and subjects were instructed to adjust sensitivity within a range that ensured linear operation. It was found that subjects generally chose a similar setting, and that all reported stimulus levels approximating those of normal conversation at that setting. A sensitivity setting of 1 was therefore standardized for all subjects using the Nucleus device. This setting placed the standard 70 dB (SPL) stimulus near the center of the 60–90 dB input dynamic range of the implant. The subject with the Clarion device was programmed with sensitivity set to zero and input dynamic range set to 60 dB (spanning input between 25 and 85 dB SPL).

B. Results

1. Sensitivity

a. Intensity increments. Intensity DLs and corresponding SDs are shown in Fig. 3. The mean intensity DL (increment size) across words and listeners was 1.3 dB for the NH listeners and 3.1 dB for the CI listeners. This planned comparison between listener groups was significant [$t(9)=-2.65$, $p=0.03$], indicating that the NH listeners were on average more sensitive to increments in intensity within the speech stimuli than were the CI listeners.[2] Although DLs were larger on average for the CI listeners, substantial variability was observed within this group, and two of the listeners (CI4 and CI5) produced DLs within the range of the NH means.

The three stimulus words were chosen to have different phonetic characteristics, which produced substantially different amplitude envelope shapes (see Fig. 1). To examine possible effects related to these differences, one-way repeated-measures analyses of variance (ANOVAs) were performed. The NH listeners [$F(2,8)=11.77$, $p=0.004$] produced intensity DLs that were larger ($p<0.05$ by Tukey post-hoc) for Word 3 ("allowance"), which was continuously voiced and had a smooth amplitude envelope lacking the steplike increment characteristic of the other two words. However, no significant difference was observed across words for the CI listeners [$F(2,10)=2.05$, $p=0.18$].

b. Fundamental frequency increments. Fundamental frequency DLs and corresponding SDs are presented in Fig. 4. The mean DL across words and listeners was 3.2 Hz for the NH listeners and 25.0 Hz for the CI listeners. This planned comparison between groups was significant, indicating that the CI users were less sensitive to increments in F0 in the speech stimuli [$t(9)=-7.44$, $p=0.001$]. The values are expressed as Weber fractions in Table II.

An examination of Fig. 4 shows that, although the NH listeners produced generally similar DLs across the three words, several of the CI listeners produced substantially

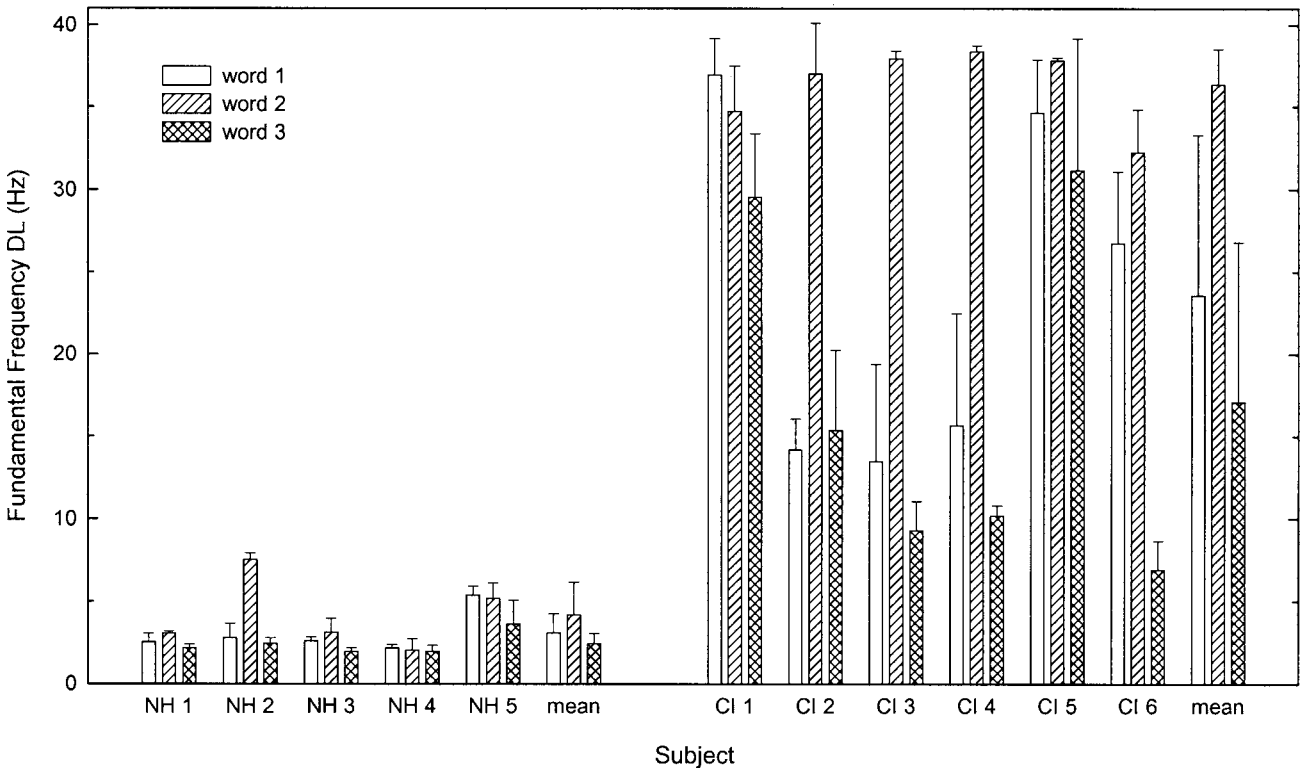


FIG. 4. Difference limens for individual subjects hearing synthesized words having a fundamental-frequency increment in the middle syllable.

higher thresholds for stimulus Word 2 ("sufficient"). Indeed, four of the CI listeners produced threshold estimates near the ceiling frequency value of 40 Hz for this word, and so it is possible that even these large DLs slightly underestimate their true thresholds for this word. This observed difference between words is supported by statistical analyses, in which no effect of word was found in a one-way repeated-measures ANOVA for the NH listeners [$F(2,8)=2.80$, $p=0.12$]. However, a second ANOVA showed that the CI subjects [$F(2,10)=11.35$, $p=0.003$] produced larger DLs for Word 2 relative to the other words ($p<0.05$ by Tukey post-hoc). These results did not change when the tests were performed on Weber fractions.

To examine the influence of the small deviations from nominal F0 values observed for some of the stimuli, the DLs were recalculated using the measured F0 value of each increment stimulus. Mean thresholds were on average within 0.1 Hz of the thresholds based on nominal frequency and shown in Fig. 4, and only three estimates exceeded a difference of 0.2 Hz from the nominal value.

2. Relation to speech recognition

Further analyses were completed to determine if there was a relationship between sensitivity to intensity increments within the speech stimuli and performance on a standard speech-recognition test for the CI users. First it was important to establish that the recognition scores covered an appropriate range. By examining the binomial probabilities (Thornton and Raffin, 1978), it was found that significant differences existed across the word recognition scores produced by the CI users. The lowest of the six scores (54%) was significantly different ($p<0.05$) from the top three scores (76%, 80%, 82%); the top score (82%) was significantly different from the bottom three scores (54%, 58%, 64%); and the two middle scores were significantly different from the top and bottom scores, respectively.

Figure 5 shows the relationship between the mean inten-

TABLE II. Mean fundamental frequency DLs as Weber fractions (%).

	Word 1 "Potato"	Word 2 "Sufficient"	Word 3 "Allowance"	Mean
NH	2.34	3.35	2.68	**2.79**
CI	17.88	29.08	18.77	**21.91**

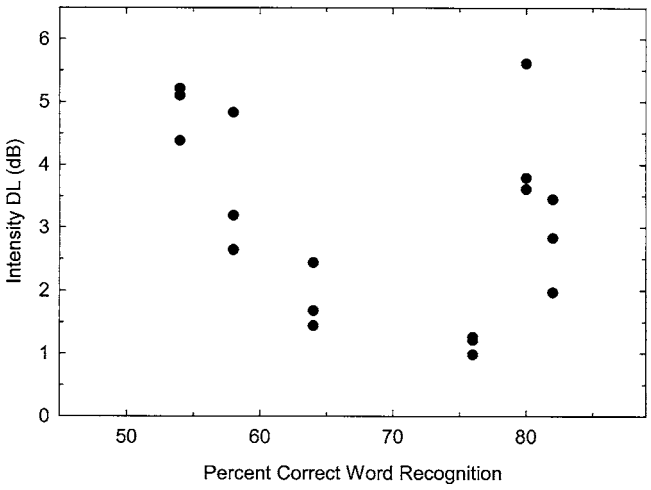


FIG. 5. Mean intensity DLs produced by the CI subjects as a function of their performance on the CNC word recognition test.

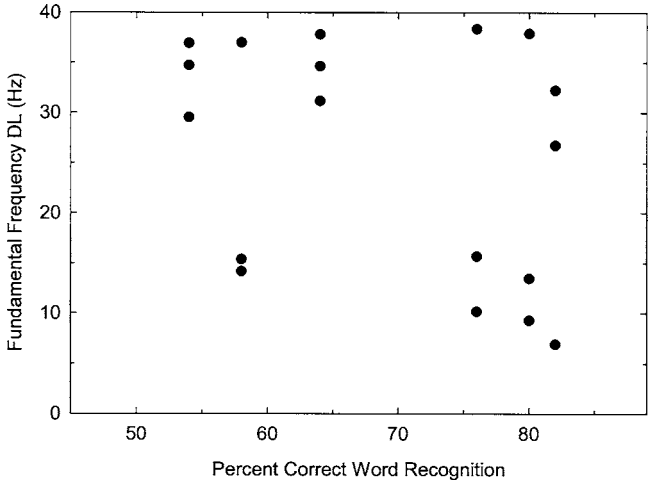


FIG. 6. Mean fundamental frequency DLs produced by the CI subjects as a function of their performance on the CNC word recognition test.

sity DLs for each word as a function of CNC word recognition for each CI listener. No relation is apparent from the figure, and the correlation between mean DL across the three words for each listener and word recognition was nonsignificant ($r=-0.35$, $p=0.50$). Figure 6 shows the relationship between the mean F0 DLs for each word as a function of the word recognition score for each CI listener. Again, no relation is apparent from the figure ($r=-0.67$, $p=0.14$, nonsignificant).[3]

III. EXPERIMENT 2. SENSITIVITY TO CONCURRENT CUES

In this experiment, intensity and F0 increases were presented concurrently within a single stimulus word and potential sensitivity enhancement resulting from the ability to combine cues was assessed.

A. Method

The same six CI subjects and five NH subjects who completed Experiment 1 participated in this experiment. The DLs obtained in Experiment 1 were used in this experiment to equate the perceptual equivalence of intensity and F0 increments for each of the three stimulus words. For the NH subjects, the median DL for each word was employed. Because of the increased variability of the CI subjects, different combined-cue stimuli were created for each subject, tailored to their individual DLs for each word.

Psychometric functions relating d' to increment size are generally linear and pass through the origins (e.g., Turner and Nelson, 1982; Nelson and Freyman, 1986; Nelson *et al.*, 1996; Lentz and Richards, 1997; Moore *et al.*, 1999; Jesteadt, Nizami, and Schairer, 2003). Indeed, these functions are often fit using lines through the origins (e.g., Nelson and Freyman, 1986; Nelson *et al.*, 1996; Moore *et al.*, 1999). These studies provide some assurance that the methodologically simple procedure of dividing the increment at DL into physically equal steps produces increments that are also perceptually equivalent.

For the NH subjects, stimulus Words 1 and 3 had concurrent-cue steps in which the intensity and F0 increases matched 1/8 of the corresponding median DLs. For Word 2, careful measurement of F0 indicated that 1/4-DL steps were the smallest that could be reliably created using the maximum 0.1-Hz nominal resolution of the synthesizer. Frequency increment stimuli (using steps of 0.32, 0.78, and 0.27 Hz for Words 1–3, respectively) were created using the procedures employed in Experiment 1. Corresponding 1/8- or 1/4-DL intensity increases (steps of 0.13, 0.24, and 0.24 dB) were then applied to each F0-increment file. Incrementing series up to two or three DLs were created.

To create the combined-cue stimuli for the CI subjects, the integer F0 increment (one to four hertz) that was closest to 1/8 of the F0 DL for each word and listener was chosen, yielding step sizes that ranged across listeners from 1/10 to 1/7 DL. Corresponding intensity step values were then calculated for each word and listener and applied to the F0 increments. Combined-cue stimuli were not created for the few conditions in which F0 DLs were near the ceiling value in Experiment 1. Each intensity and F0 value was confirmed using the procedures employed in Experiment 1, and found to be within 0.1 dB and 0.1 Hz of the nominal values. The same pair of experienced listeners confirmed that all stimuli were free from artifacts. The bottom panel of Fig. 2 shows the waveform for Word 1 ("potato") having concurrent intensity and F0 increments.

Difference limens were obtained using the procedures employed in Experiment 1. The step size was increased by a factor of two until the first four reversals were obtained. Words were presented in random order. Individual estimates were discarded and replaced on the rare occasion that the SD of the reversals exceeded 0.35 DL. Three or more estimates were averaged to produce each threshold.

B. Results

1. Sensitivity

Thresholds are expressed as proportions of individual-cue DLs in Table III. Therefore, a DL of 1.0 indicates no advantage of concurrent cues over individual cues. When averaged across words and listeners, the NH subjects produced a DL of 0.50 and the CI subjects produced a DL of 0.51 [planned comparison $t(9)=-0.10$, $p=0.92$]. Thus, these data indicate no overall difference across listener groups in the sensitivity enhancement that results from the presence of concurrent cues. Further, the value of 0.5 indicates that the two cues combined to produce a threshold that was half that of either cue alone.

2. Relation to speech recognition

Figure 7 shows the relationship between the mean combined-cue DLs for each word as a function of CNC word recognition for each of the CI users. These data show a tendency for individuals who were better able to combine the concurrent cues to perform better on the speech recognition tests. In contrast to the results of Experiment 1, the correlation between mean DL across the three words and word recognition was significant ($r=-0.87$, $p=0.02$).

TABLE III. Mean combined-cue DLs (relative to individual-cue DLs) and standard deviations.

	Word 1 "Potato"		Word 2 "Sufficient"		Word 3 "Allowance"		
S	M	*SD*	M	*SD*	M	*SD*	Mean
NH 1	0.54	*0.10*	0.43	*0.02*	0.38	*0.12*	0.45
NH 2	0.87	*0.22*	0.42	*0.12*	0.66	*0.11*	0.65
NH 3	0.67	*0.22*	0.29	*0.04*	0.41	*0.26*	0.46
NH 4	0.45	*0.08*	0.26	*0.04*	0.39	*0.06*	0.37
NH 5	0.63	*0.20*	0.36	*0.31*	0.69	*0.08*	0.56
Mean	**0.63**	*0.14*	**0.35**	*0.07*	**0.51**	*0.14*	**0.50**
CI 1	—	—	0.54	*0.13*	0.74	*0.37*	0.64
CI 2	0.52	*0.05*	—	—	0.72	*0.07*	0.62
CI 3	0.24	*0.00*	0.28	*0.06*	0.22	*0.03*	0.25
CI 4	0.38	*0.05*	—	—	0.60	*0.15*	0.49
CI 5	0.52	*0.19*	—	—	0.76	*0.16*	0.64
CI 6	0.24	*0.04*	0.47	*0.09*	0.49	*0.02*	0.40
Mean	**0.38**	*0.13*	**0.43**	*0.11*	**0.59**	*0.19*	**0.51**

IV. DISCUSSION

Sensitivity under natural conditions is a potentially complex interaction between a number of factors. As a result, it would be difficult or impossible to predict "real-world" sensitivity based upon the results of studies employing direct stimulation of single electrodes using simple stimuli. In the present study, this question was addressed under conditions in which these various simultaneous factors were allowed to interact in a more natural way. It is important to note that, although they produced a modest range of word recognition scores, the CI users employed in the present study were all relatively good users, and so the current results may be restricted to representing this population.

In Experiment 1 it was found that intensity DLs for spectro-temporally complex stimuli presented in soundfield were larger on average for the CI users than for the NH listeners. This sensitivity for CI users is a result of tradeoffs between fine sensitivity to increments in electrical current and limited dynamic range, combined with differential sensitivity across electrodes within individuals and differences in the number of discriminable steps across users. These tradeoffs together yielded a real-world intensity sensitivity that was poorer than that of NH by a factor of 2.4 in decibel units.

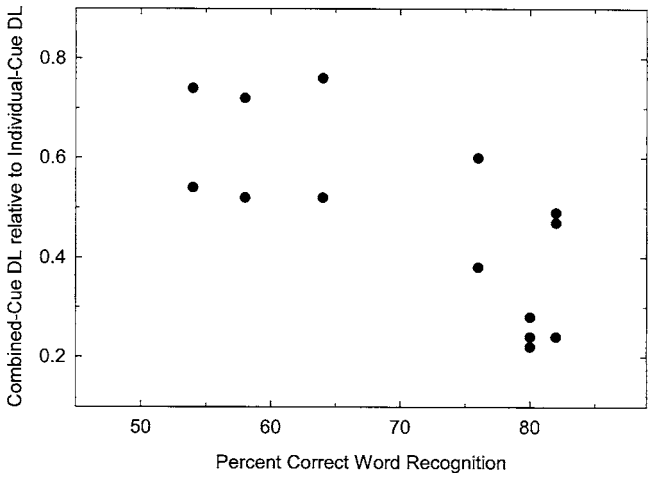


FIG. 7. Mean DLs for detection of concurrent intensity and fundamental frequency increments produced by the CI subjects as a function of their performance on the CNC word recognition test.

Further, while the NH subjects produced larger intensity DLs for the stimulus word that had an amplitude envelope lacking the abrupt steplike changes characteristic of the other two words, the CI subjects produced DLs that were more uniform. These data therefore suggest that CI users may be less able than NH listeners to exploit abruptness of change in envelope to detect intensity increments. Although gated stimulus presentations have been found to produce larger intensity DLs than conditions in which increments were applied to continuous or semi-continuous pedestals (Wojtczak *et al.*, 2003), this difference was restricted to Clarion users and was not observed in users of the Nucleus device. However, because intensity discrimination in the present study required comparison across a silent interstimulus interval, and because the preceding and following "fringe" syllables provided no cue, the current conditions can be considered variations of gated conditions. It may be worthwhile to examine systematically the influence of abruptness of change on intensity discrimination by different listener groups, by employing simple stimuli and gated- or continuous-stimulus conditions in which increments have systematically varying rise/fall durations.

Discrimination of F0 by the NH listeners in the present study is in accord with classic work employing isolated vowels having nonstationary F0 (Klatt, 1973). Values produced by the CI subjects are roughly one order of magnitude larger and are in accord with recent estimates of rate or FM discrimination (e.g., Zeng, 2002; Chen and Zeng, 2004). Under normal listening conditions, F0 may be encoded not only by temporal cues, but also by changes in the spectral balance of activity across the electrodes as the signal increases in F0 and harmonics move up in frequency. Geurts and Wouters (2001) showed that this "average amplitude cue" could indeed be used by some CI subjects to discriminate F0 in conditions where only spectral balance cues existed. However,

TABLE IV. Mean DLs (and standard deviations) for synthesized speech and complex tone control stimuli.

	F0 DLs (Hz)			Intensity DLs (dB)		
S#	Word 1	Complex Tone	Proportion	Word 1	Complex tone	Proportion
NH 1	2.0 (0.2)	1.2 (0.2)	**0.60**	0.9 (0.2)	0.8 (0.1)	**0.89**
NH 2	2.4 (0.7)	1.0 (0.1)	**0.42**	0.8 (0.1)	0.5 (0.1)	**0.63**
CI 1	22.0 (4.5)	22.4 (0.9)	**1.02**	3.4 (1.1)	3.1 (1.5)	**0.91**
CI 2	11.9 (2.6)	12.0 (3.5)	**1.01**	2.9 (1.3)	3.1 (0.9)	**1.07**
CI 3	24.7 (1.0)	23.0 (2.1)	**0.93**	3.3 (0.8)	3.5 (1.3)	**1.06**

the difference between NH and CI listeners in the present study is approximately the same as that seen when place cues are removed from NH listeners. This result serves to highlight the heavy reliance on temporal cues and the especially poor spectral resolution of modern CIs by suggesting a relative lack of contribution by spectral balance (place) cues to F0 discrimination even when current balance across electrodes is allowed to vary in normal fashion.

Under conditions that were relatively similar to the F0 conditions of the present study, Geurts and Wouters (2001) obtained a mean F0 DL of approximately 8%, compared to the value of roughly 22% in the present study. Although there could be influences of CI processor coding not accounted for in the direct-stimulation technique of the previous study, it is also possible that the values observed here were affected by the natural durations of the speech stimuli and more accurately reflect the real-world abilities of CI users, than do the long-duration stimuli employed in the previous study.

There were differences, sometimes quite large, between the F0 DLs produced by the CI listeners across the three stimulus words. The presence of a higher DL for Word 2 ("sufficient") is especially clear in the three CI subjects who tended to produce the lowest DLs (CI2, CI3, and CI4). The three center syllables differ (by design) along a number of dimensions, including overall bandwidth, position of formants, and amount of spectral change within the syllable. Although Word 2 does have the smallest amount of spectral change within the center syllable, the changes that occur within the other two words are small compared to the F0 DL of the CI listeners. There are also differences in the intensities of the middle syllables relative to the preceding and following syllables, and differences in the depth of modulations when extracted using identical procedures (see Fig. 1).

The word that produced the highest F0 DL was also that having the briefest middle-syllable duration at 70 ms. One possibility is that by equating the slow-response rms peak levels of the three words, the instantaneous peak of the brief middle syllable in Word 2 was nearer the top of the input dynamic range of the CI device, and so the temporal fluctuations were compressed to some degree not present in the other two words. However, intensity DLs were not larger for this word, indicating that any limiting of the brief middle syllable that did occur was not severe enough to disrupt intensity discrimination.

Another possibility is that accurate F0 information was simply less available to the CI users because of the brief duration of the target segment. The results of other studies have not shown this dependence of F0 discrimination on duration. Discrimination by NH subjects is relatively consistent across the range of 70 to 150 ms, even when the stimulus is a complex tone composed of unresolved harmonics (White and Plack, 2003). There is one relevant early study involving two CI subjects (Tong *et al.*, 1982). These authors evaluated effects of duration on rate pitch for stimuli that varied linearly in pulse rate during the course of the stimulus pulse train. Results were mixed, with some tendency for performance to improve with increasing duration, but only for higher initial pulse rates. Thus, although substantial differences exist between the present investigation and the study by Tong *et al.*, those early data would not allow the prediction of the large F0 DLs observed here for Word 2. The cause of these differences in F0 DLs across words may warrant further investigation.

While the present study provides some indication of the sources of the observed sensitivity differences between listener groups, its main goals involved an examination of the magnitude of these differences. However, to address the general question involving influences of the stimuli versus influences of the CI device, control conditions were prepared to compare sensitivity to increments in speech to those in simple stimuli. A complex tone having a F0 of 125 Hz, a bandwidth up to 6000 Hz, and a duration of 150 ms was synthesized and incrementing intensity and F0 series were prepared as before. The F0, spectral range, and duration were selected to resemble those of the middle syllable of Word 1, and the complex tone was presented at an intensity matching that of the middle syllable of Word 1. Therefore, differences between the stimuli primarily involved the presence in Word 1 of surrounding syllables, spectral shaping, a F0 that changed (±8 Hz) over time, and a nonconstant amplitude envelope.

Difference limens for Word 1 and the complex tone were collected from two additional NH listeners and three CI users using the procedures employed earlier. These data are presented in Table IV, along with the proportion of the speech DL produced by the nonspeech stimuli. It was found that the NH listeners experienced a reduction in F0 DLs when simple stimuli replaced the speech stimuli (mean DL for complex tone was on average 0.51 that for speech), but the CI users

experienced no such improvement (mean DL for complex tone was on average 0.99 that for speech). Similarly, intensity DLs were reduced, especially for one of the NH listeners, with the change to simple stimuli (mean DL for complex tone was on average 0.76 that for speech), but the CI users experienced no improvement (mean DL for complex tone 1.01 that for speech). These results, although preliminary, indicate that the sensitivity differences observed across listener groups in the present study may be related more to factors associated with the implant, rather than dynamics of the stimuli. However, it should be noted that identifying in detail the relative contributions of the various influences that combine to produce real-world sensitivity may not be a simple matter.

In Experiment 2, concurrent intensity and F0 increments were presented within single words. It was found that, on average, CI users and NH listeners were equally able to combine concurrent cues to improve sensitivity. This combination of two cues served to improve sensitivity to half that of either cue in isolation for both groups. This result indicates that CI users should not be especially hindered by the presence of concurrent cues. Further, these results may potentially serve to simplify our understanding of CI sensitivity to cues that change in more complex fashion. It is clear that sensitivity to some cues in isolation is poorer than that of NH. Because it was found that CI users in the current study were able to exploit concurrent cues in similar fashion to the NH listeners, it is suggested that overall sensitivity to stimuli having concurrent changes in more than one cue may potentially be estimated based simply upon sensitivity to the individual cues.

The CI listeners demonstrate larger variability than their NH counterparts including, as Table III shows, in the combined-cue conditions of Experiment 2 where the stimuli were customized relative to individual sensitivity. It may be potentially interesting to examine psychometric functions relating d' to increment size for individual cues and listeners in an attempt to examine in more detail the cue-combining mechanism and to predict combined-cue sensitivity enhancement for individual subjects. Particularly large or small combined-cue DLs may potentially be explained by differences in the slopes of the psychometric functions for the two contributing cues, together with differences in the relative contributions of the two cues.

It is not uncommon for basic psychoacoustic abilities to relate only weakly to speech reception in hearing-impaired listeners, and a lack of correspondence has been seen previously for CI users (Cazals *et al.*, 1990; Busby and Clark, 1999). However, the lack of a significant relationship between intensity or F0 DLs and word recognition in the present study may be considered somewhat surprising given the nature of the stimuli employed and the fact that some correspondence for CI users has been observed previously (Cazals *et al.*, 1991, 1994; Muchnik *et al.*, 1994; Dorman *et al.*, 1996; Hanekom and Shannon, 1998; Fu, 2002).

In contrast to the lack of relationship in Experiment 1, a significant relationship was observed between individual differences in the ability to combine intensity and F0 cues, and word recognition. Although, on average, CI users combine cues to improve sensitivity with the same success as NH listeners, individual differences within the CI group in this ability predicted performance on word recognition. Because the combined-cue stimuli in Experiment 2 were tailored for each CI listener to match their sensitivity to the individual cues in Experiment 1, these data indicate that it is the differential ability to combine cues, separate from the sensitivity to the cues themselves, that correlates to speech reception. Because this pattern of absence versus presence of correlation was observed within identical groups of subjects, using similar stimuli, identical word recognition scores, and identical procedures, the usual difficulty encountered when attempting to interpret results of different studies is avoided. These data therefore suggest, at least for relatively good CI users, that the complex processing required to perform the multidimensional task of combining concurrent cues may more closely predict performance on speech recognition than does simple psychoacoustic sensitivity.

The cues manipulated in the present study are those that signify lexical stress in English. Thus, the present study may have implications for the perception of prosodic cues by CI users. Although the relative weights of the acoustic cues are a matter of some debate (cf. Howell, 1993; Cutler, 2005), sensitivity to intensity and F0 was found to be poorer by factors of 2.4 and 8, respectively. Because the relative sensitivity enhancement resulting from concurrent cues was found to be equal on average across listener groups, these sensitivities may possibly be used to estimate the relative availability of cues to lexical stress in modern CI users in natural settings.

V. CONCLUSIONS

(1) Sensitivity to basic acoustic cues in NH and CI listeners under natural listening conditions may be estimated using stimuli having characteristics that approximate those of normal speech, and presentation methods that allow the interaction of multiple simultaneous factors.

(2) The factors governing intensity discrimination in modern CI devices combine to produce sensitivity in natural settings that is poorer than that of NH by a factor of 2.4 in decibels. This mean value represents relatively good users.

(3) Unlike NH listeners, the CI users did not produce smaller intensity DLs for words having abrupt changes in amplitude envelope.

(4) Fundamental frequency DLs for these relatively good CI users were poorer than those of NH listeners by nearly one order of magnitude, which indicates the relative lack of availability of spectral balance cues and the heavy reliance on temporal cues in CI listening.

(5) The CI subjects produced significantly larger F0 DLs for one of the three stimulus words (Word 2), which happens to have the briefest syllable duration. However, this effect was not observed in intensity DLs.

(6) The sensitivity enhancement resulting from combining concurrent cues was similar across NH and CI listeners. Thresholds were reduced, on average, to half those of either cue alone.

(7) The relation between sensitivity to the individual cues and performance on a standard word recognition test was weak for the CI users. However, individual differences in the ability to combine concurrent cues to improve sensitivity were significantly related to word recognition.

ACKNOWLEDGMENTS

This work was supported by NIDCD Grant No. DC05795 and by a grant from the University of South Carolina Office of Research. Portions were drawn from a dissertation submitted by the first author to the Graduate School at the University of South Carolina. The authors thank Wendy Potts for audiological assistance and Sid Bacon for helpful comments. We are also grateful for the constructive comments of two anonymous reviewers and the Associate Editor on a previous version of this manuscript.

[1]One of the NH listeners (NH 4) had slightly to moderately elevated thresholds at the highest audiometric frequencies in one ear. The decision to include this individual was based upon the unilateral nature of the threshold elevation and the fact that the elevation was restricted to frequencies above that of the stimuli. His DLs were found to be the lowest of the group in almost every condition.

[2]All differences observed here using planned comparison are robust enough to also be significant when assessed using ANOVA.

[3]To investigate whether the high F0 DLs associated with Word 2 may have obscured the relationship between word recognition and F0 sensitivity, correlations were performed between word recognition and F0 DLs for each of the words individually. However, none of these relationships reached statistical significance ($p > 0.05$).

ANSI (**1996**). ANSI-S3.6, 1996, "Specifications for audiometers." (American National Standards Institute, New York).

Bacon, S. P. (**2004**). *Compression: From Cochlea to Cochlear Implants*, R. R. Fay, A. N. Popper (series editors), S. P. Bacon, vol. editor (Springer-Verlag, New York).

Busby, P. A., and Clark, G. M. (**1999**). "Gap detection by early-deafened cochlear-implant subjects," J. Acoust. Soc. Am. **105**, 1841–1852.

Cazals, Y., Pelizzone, M., Kasper, A., and Montandon, P. (**1990**). "Multichannel cochlear implant patients with different open speech understanding show some similar basic psychophysical results," Acta Oto-Laryngol. **469**, 150–155.

Cazals, Y., Pelizzone, M., Kasper, A., and Montandon, P. (**1991**). "Indication of a relation between speech perception and temporal resolution for cochlear implantees," Ann. Otol. Rhinol. Laryngol. **100**, 893–895.

Cazals, Y., Pelizzone, M., Saudan, O., and Boex, C. (**1994**). "Low-pass filtering in amplitude modulation detection associated with vowel and consonant identification in subjects with cochlear implants," J. Acoust. Soc. Am. **96**, 2048–2054.

Chen, H., and Zeng, F. G. (**2004**). "Frequency modulation detection in cochlear implant subjects," J. Acoust. Soc. Am. **116**, 2269–2277.

Cohen, L. T., Saunders, E., and Clark, G. M. (**2001**). "Psychophysics of a prototype peri-modiolar cochlear implant electrode array," Hear. Res. **155**, 63–81.

Cullen, J. K., and Long, G. R. (**1986**). "Rate discrimination of high-pass-filtered pulsetrains," J. Acoust. Soc. Am. **79**, 114–119.

Cutler, A. (**2005**). "Lexical stress," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Oxford), pp. 264–289.

Dorman, M. F., Smith, L. M., Smith, M., and Parkin, J. L. (**1996**). "Frequency discrimination and speech recognition by patients who use the Ineraid and continuous interleaved sampling cochlear-implant signal processors," J. Acoust. Soc. Am. **99**, 1174–1184.

Drennen, W. R., and Pfingst, B. E. (**2005**). "Current-level discrimination using bipolar and monopolar electrode configurations in cochlear implants," Hear. Res. **202**, 170–179.

Flanagan, J. L., and Saslow, M. (**1958**). "Pitch discrimination for synthetic vowels," J. Acoust. Soc. Am. **30**, 435–442.

Formby, C. (**1985**). "Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners," J. Acoust. Soc. Am. **78**, 70–77.

Fu, Q. J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**, 1635–1639.

Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.

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.

Houtsma, A. J., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.

Howell, P. (**1993**). "Cue trading in the production and perception of vowel stress," J. Acoust. Soc. Am. **94**, 2063–2073.

Jesteadt, W., Nizami, L., and Schairer, K. S. (**2003**). "A measure of internal noise based on sample discrimination," J. Acoust. Soc. Am. **114**, 2147–2157.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Klatt, D. H. (**1973**). "Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception," J. Acoust. Soc. Am. **53**, 8–16.

Kreft, H. A., Donaldson, G. S., and Nelson, D. A. (**2004**). "Effects of pulse rate and electrode array design on intensity discrimination in cochlear implant users," J. Acoust. Soc. Am. **116**, 2258–2268.

Lentz, J. J., and Richards, V. M. (**1997**). "Sensitivity to changes in overall level and spectral shape: An evaluation of a channel model," J. Acoust. Soc. Am. **101**, 3625–3635.

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

McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.

Miller, G. A. (**1947**). "Sensitivity to changes in the intensity of white noise and its relation to masking and loudness," J. Acoust. Soc. Am. **19**, 609–619.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1999**). "Effects of frequency and duration on psychometric functions for detection of increments and decrements in sinusoids in noise," J. Acoust. Soc. Am. **106**, 3539–3551

Muchnik, C., Taitelbaum, R., Tene, S., and Hildesheimer, M. (**1994**). "Auditory temporal resolution and open speech recognition in cochlear implant recipients," Scand. Audiol. **23**, 105–109.

Nelson, D. A., and Freyman, R. L. (**1986**). "Psychometric functions for frequency discrimination from listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **79**, 799–805.

Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). "Intensity discrimination as a function of stimulus level with electric stimulation," J. Acoust. Soc. Am. **100**, 2393–2414.

Nilsson, M., Soli, S. D., and Sullivan, J. A. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.

Peterson, G. E., and Lehiste, I. (**1962**). "Revised CNC lists for auditory tests," J. Speech Hear Disord. **27**, 62–70.

Pfingst, B. E., Burnett, P. A., and Sutton, D. (**1983**). "Intensity discrimination with cochlear implants," J. Acoust. Soc. Am. **73**, 1283–1292.

Shannon, R. V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157–189.

Thornton, A. R., and Raffin, M. J. M. (**1978**). "Speech-discrimination scores modeled as a binomial variable," J. Speech Hear. Res. **21**, 507–518.

Tong, Y. C., and Clark, G. M. (**1985**). "Absolute identification of electric pulse rates and electrode positions by cochlear implant patients," J. Acoust. Soc. Am. **77**, 1881–1886.

Tong, Y. C., Clark, P. J., Blamey, P. J., Busby, P. A., and Dowell, R. C. (**1982**). "Psychophysical studies for two multi-channel cochlear implant patients," J. Acoust. Soc. Am. **71**, 153–160.

Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant patients," J. Acoust. Soc. Am. **82**, 106–115.

Turner, C. W., and Nelson, D. A. (**1982**). "Frequency discrimination in regions of normal and impaired hearing," J. Speech Hear. Res. **25**, 34–41.

Vaissiere, J. A. (**2005**). "Perception of intonation," in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Oxford), pp. 264–289.

Viemeister, N. F., and Bacon, S. P. (**1988**). "Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones," J. Acoust. Soc. Am. **84**, 172–178.
White, L. J., and Plack, C. J. (**2003**). "Factors affecting the duration effect in pitch perception for unresolved complex tones," J. Acoust. Soc. Am. **114**, 3309–3316.
Wojtczak, M., Donaldson, G. S., and Viemeister, N. F. (**2003**). "Intensity-discrimination and increment detection in cochlear-implant users," J. Acoust. Soc. Am. **114**, 396–407.
Zeng, F. G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.
Zeng, F. G. (**2004**). "Compression and Cochlear Implants," in *Compression: From Cochlea to Cochlear Implants*, R. R. Fay and A. N. Popper, series editor and S. P. Bacon, volume editor (Springer-Verlag, New York), pp. 184–220.
Zeng, F. G., and Shannon, R. V. (**1994**). "Loudness-coding mechanisms inferred from electric stimulation of the human auditory system," Science **264**, 564–566.
Zeng, F. G., and Shannon, R. V. (**1999**). "Psychophysical laws revealed by electric hearing," NeuroReport **10**, 1931–1935.

Compensation following real-time manipulation of formants in isolated vowels

David W. Purcell[a] and Kevin G. Munhall
Department of Psychology, Queen's University, Kingston, Ontario, K7L 3N6, Canada

(Received 4 July 2005; revised 23 November 2005; accepted 14 January 2006)

Auditory feedback influences human speech production, as demonstrated by studies using rapid pitch and loudness changes. Feedback has also been investigated using the gradual manipulation of formants in adaptation studies with whispered speech. In the work reported here, the first formant of steady-state isolated vowels was unexpectedly altered within trials for voiced speech. This was achieved using a real-time formant tracking and filtering system developed for this purpose. The first formant of vowel /ɛ/ was manipulated 100% toward either /æ/ or /ɪ/, and participants responded by altering their production with average F1 compensation as large as 16.3% and 10.6% of the applied formant shift, respectively. Compensation was estimated to begin <460 ms after stimulus onset. The rapid formant compensations found here suggest that auditory feedback control is similar for both F0 and formants. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2173514]

PACS number(s): 43.66.Jh, 43.70.Aj [AL] Pages: 2288–2297

I. INTRODUCTION

Auditory feedback plays an important role in speech production. Post-lingually deafened adults show differences relative to the normal hearing population in the fundamental frequency or F0 of the voice, vocal intensity, and speaking rate (Leder *et al.*, 1987; Waldstein, 1990; Lane and Webster, 1991; Cowie and Douglas-Cowie, 1992; Leder and Spitzer, 1993). There are also differences in the formants of post-lingually deafened speakers after long-term hearing loss (Waldstein, 1990; Schenk *et al.*, 2003), or with shorter-term alteration of auditory feedback through changes in the cochlear implant status (Svirsky and Tobey, 1991; Economou *et al.*, 1992; Perkell *et al.*, 1992; Svirsky *et al.*, 1992). Shorter-term changes in auditory feedback also affect F0 (Binnie *et al.*, 1982; Economou *et al.*, 1992; Perkell *et al.*, 1992; Svirsky *et al.*, 1992). In summary, there is evidence showing both relatively rapid and slower changes in speech parameters like F0 and formants in the absence of auditory feedback. In this paper we investigate whether relatively rapid changes in production can be induced by altering formants in the auditory feedback of normal hearing individuals.

There is some debate, however, about the relative weight of auditory feedback in the control of suprasegmental speech parameters like F0, and segmental aspects of speech like formants (Svirsky *et al.*, 1992; Perkell *et al.*, 1992; Perkell *et al.*, 2000). Perkell *et al.* (2000) suggest that auditory feedback is used differently by a postural system operating on a suprasegmental time scale, and a phonemic settings system at the segmental level. The postural system adjusts average suprasegmental parameters such as SPL, F0, rate, and the degree of prosodic influence on SPL and F0, in order to maintain intelligibility. Phonemic settings control mechanisms responsible for producing the distinctions in different phonemes, and employ auditory feedback intermittently for maintenance purposes. The two systems use feedback on different time scales, and perhaps have different functional organization for different speech attributes. Intelligibility is reduced with post-lingual deafness (Cowie *et al.*, 1982), but the preservation of functionally adequate intelligibility lends evidence to the argument that auditory feedback is not necessary for the moment-to-moment control of segmental speech characteristics. Svirsky and Tobey (1991) note, however, that intelligibility is maintained until relevant acoustic parameters cross phonemic boundaries, making intelligibility a relatively insensitive measure of subtler changes. The present paper is part of a research program that investigates the use of feedback in the control of parameters like F0 and formants.

Two different types of experimental paradigms have been used to investigate the characteristics of the auditory feedback system. In order to study the influence of auditory feedback over longer time spans, a sensorimotor adaptation technique has been used. Drawn directly from earlier literature on visuomotor control (e.g., Held, 1965), the experiments involve the manipulation of sensory feedback. Two aspects of the response are examined: the immediate compensation for the perturbation and, more importantly, the adaptation or persistence of the effect after the perturbation is eliminated.

Using whispered speech, Houde and Jordan (1998; 2002) have shown that gradual manipulation of the feedback of formants F1 through F3 can lead to compensatory responses. For example, a subject may whisper the vowel /ɛ/, but is provided with auditory feedback, where the formants have been adjusted such that the feedback vowel sounds more like the subject's own /i/. Without conscious knowledge of their response, the subject may compensate by producing a vowel closer to /æ/ (Houde and Jordan, 2002). After sufficient training, adaptation was observed when auditory feedback was eliminated using masking (Houde and Jordan, 2002). Similar results have been reported recently for voiced

[a]Author to whom correspondence should be addressed. Electronic-mail: purcelld@nca.uwo.ca

 0001-4966/2006/119(4)/2288/10/$22.50

speech instead of whispered using improved formant transformation systems (Villacorta *et al.*, 2004; 2005).

Subjects show both compensation and adaptation effects to the perturbation of the fundamental frequency of their voice. Jones and Munhall (2000) in a study of English-speaking subjects showed that gradual changes in pitch feedback were countered by F0 production in the opposite direction, and that an adaptation effect persisted after feedback was returned to normal. This effect has been replicated with native speakers of Mandarin (Jones and Munhall, 2002). Both the formant and F0 adaptation effects have been interpreted as evidence for a central representation or internal model that is used in motor control in a feedforward manner (e.g., Miall and Wolpert, 1996). In this mechanism, feedback's primary role is to tune the "model" that the nervous system uses to predict the effects of movements and thus plan more effectively. This similarity between F0, particularly in English, and the Houde and Jordan vowel formant data suggests that similar systems are involved (or the same system).

The second experimental paradigm that has been used to examine the role of auditory feedback in speech production is the introduction of sudden, unexpected auditory feedback changes. This technique tests auditory feedback as part of a servomechanism that is involved in the moment-to-moment control of articulation. When sudden changes to the pitch of the voice are unexpectedly introduced, most individuals will change their production within about 100 to 225 ms by a shift of F0 in the opposite direction (Burnett *et al.*, 1997). The response magnitude and latency have shown an indifference to stimulus pitch-shift magnitude (Burnett *et al.*, 1998), indicating that the response has generally not been elicited in a linear range. When the duration of the pitch-shift stimulus is extended beyond 500 ms, the existence of a second response is observable, and is more directly under voluntary control (Hain *et al.*, 2002). The rapid response is observed in speakers of Mandarin (Jones and Munhall, 2002) and English (Burnett *et al.*, 1997). In the present paper we extend this work by using sudden perturbations of formant frequency during vowel production.

These two experimental paradigms (employing sudden and gradual manipulations of auditory feedback) can be used to study the potential differences and similarities in the control of segmental and suprasegmental speech characteristics. For F0, the presence of both adaptation effects and compensation to rapid perturbations suggests that the neural mechanisms for its control operate on multiple different time scales. By examining the presence of immediate responses to formant perturbations, this study extends our understanding of the way in which the sound of our voice contributes to fluent speech. If both F0 and formant perturbations show the same sensitivity to immediate feedback, it will suggest that the basic form of the feedback system is similar across the various properties of speech acoustics.

II. METHODS

A. Subjects

Subjects were undergraduate students ($n=28$, 21 females), and varied in age from 17 to 28 yr. Each person participated in a single experimental condition completed in one session. Thresholds of hearing were evaluated individually in each ear at octave frequencies from 500 to 4000 Hz. The majority of participants had normal thresholds ($<=20$ dB HL), but four individuals were included who had thresholds as high as 35 dB HL at only one frequency. Other audiometric frequencies were normal in these individuals near the frequency with a high threshold, and they did not have atypical responses to the altered auditory feedback. They were retained to have as much data as possible, and because the accuracy of speech processing is not predictable from the audiogram (Surprenant and Watson, 2001), and the speech feedback was well above threshold across frequency. All subjects learned English as a first language and had no known language or speech impairments. Some subjects beyond the reported 28 were not included in the analysis because they failed one of the above criteria, had difficult to track formants as described below, or their vowels /ɪ/, /ɛ/, and /æ/ were not separated by at least 50 Hz in F1.

B. Equipment

During the experiment, subjects produced isolated steady-state vowels that were transduced into an electrical signal using a Shure headset microphone type WH20. This microphone signal was amplified using a Tucker-Davis Technologies MA3 microphone amplifier with the +20 dB gain switch active. After amplification, the microphone signal was low-pass filtered with cutoff frequency 4500 Hz using an analog Frequency Devices type 901 filter with a gain of 0 dB. A National Instruments PXI-6052E input/output board mounted in a PXI-1002 chassis digitized this signal at 10 kHz with 16-bit precision. The voice was analyzed and filtered (details explained below) in real time using a National Instruments PXI-8176 embedded controller, which is essentially an Intel Pentium III processor running at 1.26 GHz. The filtered voice signal was converted back to analog by the digitizer at 10 kHz with 16-bit precision. The processed analog signal was again low-pass filtered as above with a second Frequency Devices unit, and subsequently amplified using a Madsen Midimate 622 audiometer where speech noise was added. The microphone MA3 amplifier was adjusted between 30 and 50 dB for each individual during training such that the Madsen input VU meter read approximately 0 dB during vocalization. The Madsen output gain controllers were set such that the vocal auditory feedback at each ear was approximately 80 dBA sound pressure level (SPL), and the speech shaped noise was approximately 50 dBA SPL. Participants heard their vocalization in real time, along with the speech-shaped noise, through Sennheiser "HD 265 linear" headphones with the same signal presented to each ear. These headphones are somewhat acoustically open and strike a reasonable balance between the contradictory requirements of shielding the listener from the airborne sound of their unaltered voice and minimizing the emphasis of bone conducted sound by the occlusion effect (Békésy, 1932; Tonndorf, 1972). The headphones were calibrated using a Brüel & Kjær sound level meter and artificial ear Type 4153. Prompts were shown to the participant

on a monitor that also displayed a bar graph representation of their speaking level. This visual feedback allowed the subject to maintain a relatively consistent level throughout the experiment.

C. Experimental conditions

Two experimental conditions were employed where the first formant (F1) was shifted for the vowel /ɛ/. The altered auditory feedback was always introduced 300 ms after the onset of voicing, and was cross-faded in linearly over 500 ms to replace the normal feedback for the duration of the trial. This slow stimulus onset was chosen to maximize compensation under the assumption that it may be similar to the pitch-shift response. Larson *et al.* (2000) found responses were larger with a 500 ms onset compared to those for step changes in feedback pitch.

The two experimental conditions evaluated whether speakers would compensate for shifts of F1 in either direction. In condition A, 16 participants (13 female) had their F1 shifted upward 100% toward /æ/. In condition B, 12 subjects (8 female) had their F1 shifted downward 100% toward /ɪ/. The actual shift in Hz was determined from the individual's average formant values. The upward shift was the individual difference between F1 of /æ/ and F1 of /ɛ/. Similarly, the downward shift was the difference between /ɛ/ and /ɪ/. For the subjects in this experiment, the average upward shift was +136 Hz (standard deviation SD=46.2 Hz), and the average downward change was −135 Hz (SD=42.7 Hz).

D. Experimental protocol

Participants were seated in a comfortable chair located in an Industrial Acoustics Company (IAC) sound insulated room for the experiment, which lasted one hour. Participants were instructed to try to produce isolated steady-state vowels using their normal speaking voice. A normal speaking level was used to set microphone gain, and a headphone level was set sufficiently high to maintain a favorable relative level between the contributions of bone-conducted and headphone delivered sounds to the net voice signal at the cochlea. If subjects had increased their speaking level significantly during the experiment, the headphone signal could have clipped and thus not maintained the desired relative level at the cochlea (or fidelity; level feedback was therefore provided to the participants and was monitored by the experimenter to avoid problems of this type). Subjects were prompted during roughly 25 practice trials to produce the isolated vowels /ɛ/, /æ/, and /ɪ/ for the duration of a 2.5 s prompt, as they would sound in the consonant-vowel-consonant (CVC) contexts "head," "had," and "hid," respectively. For formant tracking (via the same method as described in Sec. II E), the spectral model order was always in the range 8 through 12, and was chosen manually based on the practice trials of /ɛ/ such that F1 was most continuous (Vallabha and Tuller, 2002). When the practice period was complete, 10 utterances of each of /ɛ/, /æ/, and /ɪ/ were recorded and the average steady-state F1 of each was determined (for the purpose of setting the individual F1 shift size).

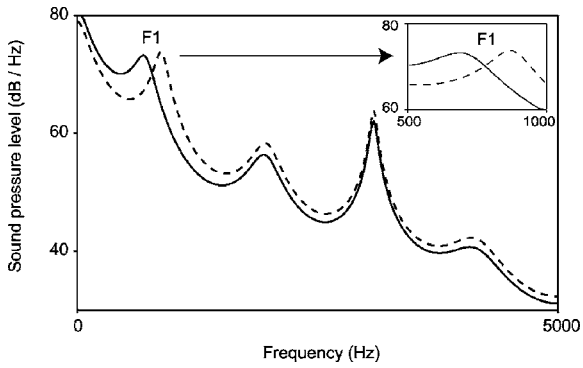


FIG. 1. Example formant shift used in a single test trial for one participant's steady-state utterance of vowel /ɛ/. The solid line shows the spectral envelope of the microphone signal, where the first formant F1 was estimated online at 657 Hz. Formant F1 was shifted upward +142 to 799 Hz for the headphone feedback signal shown with the dashed line. This shift is the difference between the participant's average F1 for vowels /æ/ and /ɛ/. Although she spoke /ɛ/, the perturbation placed the feedback F1 near the F1 frequency of her average production of /æ/. The inset shows the F1 formant perturbation with a smaller frequency range.

The experiment itself consisted of four blocks of 100 trials with an interprompt interval of approximately 1.5 s. Each trial consisted of the presentation of 1 of 11 different vowels in a CVC context (/hVd/) that were chosen to use the entire English vowel space (Ladefoged, 1982). Auditory feedback was normal for all trials except rare test trials, where the target vowel /ɛ/ was perturbed. This approach was intended to minimize learning effects such that perturbed trials could be averaged together. In each block of 100 trials, there were generally five test trials where the target vowel /ɛ/ was manipulated; however, due to random spacing and order constraints, as few as four or as many as six test trials could occur in a block. The test trials were intermingled randomly with the other 10 distractor vowels. The test trials were always separated by at least three instances of the target vowel /ɛ/, where auditory feedback was normal. These target instances with normal auditory feedback were themselves also separated by at least one distractor vowel. Thus, it was presumed that production of the target vowel had returned to normal before auditory feedback was again manipulated.

E. Online formant shifting and detection of voicing

With the feedback modification system, F1 was shifted by filtering out the signal at the frequency where the formant was estimated, and emphasizing the signal at the new desired formant frequency. This was accomplished using a filter transfer function with a pair of spectral zeros in the numerator to attenuate the energy (harmonics of the glottal fundamental frequency F0 for vowels) near the existing formant, and a pair of spectral poles in the denominator to amplify energy near the new formant. Figure 1 shows the spectral envelope of the microphone and feedback signals for a single trial where F1 of /ɛ/ was pushed upward 142 Hz.

To shift formants, it is necessary to estimate them from the speech signal, and then to filter the speech signal and deliver it to the subject as altered auditory feedback. Online real-time formant tracking and filtering of the speech signal

was implemented on the National Instruments PXI computer using the LabView Real-Time language. Formants were tracked online using an iterative Burg algorithm for estimating spectral parameters (Orfandidis, 1988). This method uses a sliding analysis window where the weights applied to older samples decay exponentially. The decay parameter was chosen such that 50% of the area under the weighting curve applied to samples less than 8.6 ms old. While tracking two formants, and shifting the first formant, the system was able to estimate the formants and update the filter coefficients approximately every 900 μs or every nine speech data points. Although the analysis window was infinitely long, using exponentially decaying weights the effective delay of the system was less than about 20 ms, or two pitch periods for a deep male voice of F0=100 Hz. This approach is slightly different from filtering methods that employ double buffers of fixed length, since here the formants are updated as often as possible with whatever new microphone data points are available after the completion of each estimate. In contrast, Houde and Jordan (1998) synthesized whispered speech by filtering a noise source with formant estimates from windows of whispered speech.

The onset of voicing was detected using a simple statistical amplitude threshold technique. On visual inspection of the microphone record, this triggering technique typically detected voicing within 10 ms of its true start.

F. Offline formant analysis

The subject's response to altered auditory feedback was evaluated offline using the same formant-tracking algorithm that was employed online. Formants were estimated offline from the recorded speech 1000 times per s (every ten speech samples). Although rare, trials were removed from the analysis if the participant stumbled or produced the wrong vowel, or if the voicing trigger was more than 150 ms early. Early triggers occurred rarely due to lip smacks or other noises that preceded vocalization. Since the manipulation of feedback commenced 300 ms post-trigger, trials with early triggers still provided at least 150 ms of unperturbed baseline data. Continuous and stable formant tracking in the target vowel was a problem for some subjects, and manifested itself as the misinterpretation of F2 as F1. If this occurred excessively during a test trial, then the filter coefficients were not stable and the auditory feedback was garbled. These trials were removed from the analysis, and in some cases a subject (six beyond the reported 28) had to be removed from the analysis if too few usable test trials remained (<10). On average, there were 20 test trials available for further analysis, but 25% of subjects had only 10 to 15 usable test trials.

The formant tracks in the test /ɛ/ trials were averaged together, as were those in the unperturbed /ɛ/ trials immediately preceding the test trials (termed "pretest" trials). Prior to averaging however, the formant tracks were conditioned to remove any individual formant estimates in each trial that were in error by evaluating the histogram of each formant track, and eliminating those estimates that were well outside the norm (estimate bins that had a count $<1\%$ of the average count of the mode bin across trials). The average formant tracks and their SDs within a given group (test or pretest) were then determined by averaging the valid formant estimates at each point in time across trials. The average trials were truncated at the end of the vowel to the duration of the shortest individual trial used in the analysis. While the subjects were prompted for 2.5 s, the average trials were typically of 1.5 s duration due to a slow reaction time or shorter vocalization in one or more trials over the course of the experiment.

G. Response evaluation

To determine if the subjects responded to the altered auditory feedback, the formants in the test trials were compared to those in the pretest trials. Average formant values were calculated for eight consecutive 150 ms blocks of data for each subject beginning at −100 ms relative to the manipulation onset. The manipulation onset will henceforth be defined as time zero. Block averages were used to smooth and reduce the data to a manageable set size for analyses of variance (ANOVA). The block size was chosen to be smaller than average pitch-shift response delays (Burnett *et al.*, 1998; Larson *et al.*, 2000), and smaller than the minimum time for a speaker to interrupt their speech found by Ladefoged *et al.* (1973). The block differences between test and pretest trials were averaged across subjects, and two-way ANOVAs with Greenhouse-Geisser corrections (due to sphericity concerns) were calculated for factors of the formant shift direction (up and down) and block number (one through eight). One ANOVA was determined separately for each formant F1 and F2.

The slow stimulus onset used in this experiment was not designed to measure the minimum response time. However, an upper limit of when the compensatory response began can be estimated using the change point test (Donath *et al.*, 2002; Siegel and Castellan, 1988). The F1 difference between the (unblocked) grand average test and pretest trials was evaluated to determine the time at which the formants began to change.

III. RESULTS

A. Formant changes in response to altered feedback

Subjects tended to change F1 in their production in the direction opposite to the manipulation of F1 in the auditory feedback. This was true for both experimental conditions. When F1 of /ɛ/ was made to sound more like /æ/ in the auditory feedback, participants tended to produce vowels with F1 in the direction of /ɪ/. Correspondingly, when F1 of /ɛ/ was pushed toward /ɪ/, subjects tended to produce vowels with F1 in the direction of /æ/. These trends can be observed in Fig. 2 for both manipulation directions. In this figure, the dark lines represent the grand average of F1 for the test trials in each experimental condition after the grand average of F1 for unperturbed pretest trials was subtracted. This normalized the data for trends within vowels and across the experiment. In addition, the one SD band indicates the between-subject variability. On average, the compensation in F1 production was small compared to the change of F1 in the auditory feedback. The variability in the response was also large rela-

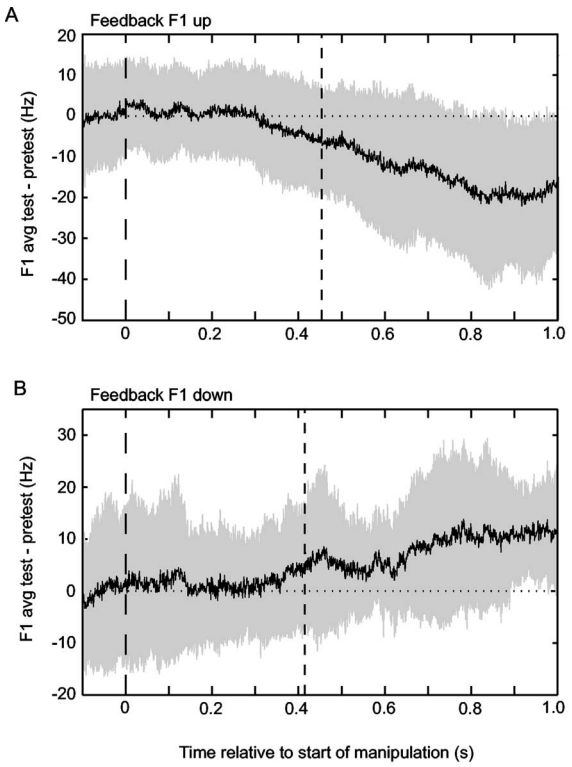


FIG. 2. The F1 formant difference responses are shown for grand averages across subjects (thick black lines) with 1 SD bounds (grey area). These curves are the F1 differences between the test and pretest trials of /ɛ/ plotted against time with respect to the perturbation onset at 0 s. Panel (a) shows the responses of 16 subjects where F1 was shifted upward toward /æ/. Panel (b) shows the responses of 12 subjects, where F1 was shifted downward towards /ɪ/. The horizontal dotted lines at zero mark no difference between test and pretest trials. The widely spaced dashed vertical lines at 0 s mark the time when altered auditory feedback began linearly cross-fading into the headphones over 0.5 s. The narrowly spaced dash vertical lines mark the change points where a significant change was first detected in the grand average F1 difference curve. For panels (a) and (b), the change points occurred at 0.454 and 0.415 s, respectively.

tive to the size of this compensation. The second formant (F2) showed small concomitant changes toward the opposite vowel category as with F1, but these were not statistically significant and again demonstrated relatively large variability. For the condition where F1 feedback was shifted downward in frequency, one subject correctly identified that /ɛ/ was being altered occasionally in an exit interview. His data were not atypical and he was not excluded from the analysis. No other participants recognized the manipulation.

The difference between average test and pretest trials was determined for each block number for formants F1 and F2, as well as the percent compensation and percent change relative to the average perturbation. These are given in Table I for the condition where feedback F1 was pushed upward on average +136 Hz, and Table II, where F1 was pushed downward on average −135 Hz. In addition, the tables report the SDs across subjects for the differences between average test and pretest trials. Variability in the absolute estimates of F1 and F2 for /ɛ/ was also large across trials within each subject. For each subject, SDs of the formant estimates were calculated separately for test and pretest trials. These SDs were then averaged across subjects for both manipulation directions, resulting in an F1 SD=38 Hz, and an F2 SD =53 Hz.

For F1, a two-way ANOVA [time block (sequential blocks of 150 ms, numbered one to eight), and direction of F1 perturbation (upward versus downward)] showed significant main effects of the block number [$F(2.8,71.7)=3.1$, $p<0.04$], and manipulation direction [$F(1,26)=22.2$, $p<0.001$]. There was also a significant interaction between these two factors [$F(2.8,71.7)=27.5$, $p<0.001$]. For F2, the same ANOVA did not quite show a significant main effect of the manipulation direction [$F(1,24)=3.8$, $p<0.062$], but had a significant interaction between the direction and block number [$F(2.8,68)=4.3$, $p<0.01$]. The interactions between the block number and manipulation direction are because the changes with block number are either negative or positive depending on the manipulation direction. Both manipulation directions had one subject without valid F2 estimates, so the degrees of freedom were lower than in the F1 ANOVA.

For each stimulus condition and formant number, Scheffé's method was used to evaluate whether the test-pretest differences of block one were different from the other blocks two through eight (given in Tables I and II). For the experi-

TABLE I. Grand average block means and standard deviations (SDs) across subjects of the difference between test and pretest trials for stimulus condition A, where F1 was manipulated upward 100% toward /æ/. Blocks are sequential windows of 150 ms. Time zero was chosen as the time when the manipulation commenced, so Block 1 shows the mean formants prior to the response. Percent compensation and percent change were with respect to the group average F1 manipulation of +136 Hz. One subject did not have valid F2 estimates and is not included in the F2 summary.

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
Block center re manipulation start (s)	−0.025	0.125	0.275	0.425	0.575	0.725	0.875	1.025
				Formant F1				
Mean $F1_{test}-F1_{pretest}$ (Hz)	−1.4	−1.8	−2.5	−7.6	−12.7	−16.9	−22.2	−21.5
SD $F1_{test}-F1_{pretest}$ (Hz)	10.2	8.9	10.1	11.7	15.0	16.1	17.3	14.8
% F1 compensation	1.0	1.3	1.8	5.6	9.4	12.4	16.3	15.8
				Formant F2				
Mean $F2_{test}-F2_{pretest}$ (Hz)	4.2	5.7	2.7	5.0	4.5	8.8	10.7	11.7
SD $F2_{test}-F2_{pretest}$ (Hz)	11.9	15.4	12.5	14.6	16.1	16.1	20.7	22.6
% F2 change	3.1	4.2	2.0	3.7	3.3	6.5	7.9	8.6

TABLE II. Grand average block means and standard deviations (SDs) across subjects of the difference between test and pretest trials for stimulus condition B, where F1 was manipulated downward 100% toward /ɪ/. Blocks are sequential windows of 150 ms. Time zero was chosen as the time when the manipulation commenced, so Block 1 shows the mean formants prior to the response. Percent compensation and percent change were with respect to the group average manipulation of −135 Hz. One subject did not have valid F2 estimates and is not included in the F2 summary.

	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
Block center re manipulation start (s)	−0.025	0.125	0.275	0.425	0.575	0.725	0.875	1.025
				Formant F1				
Mean $F1_{test}-F1_{pretest}$ (Hz)	3.3	4.2	3.6	7.5	7.9	12.2	13.3	14.3
SD $F1_{test}-F1_{pretest}$ (Hz)	13.4	12.2	8.9	11.4	8.0	13.1	9.3	6.4
% F1 compensation	2.4	3.1	2.7	5.6	5.8	9.1	9.8	10.6
				Formant F2				
Mean $F2_{test}-F2_{pretest}$ (Hz)	−1.7	1.0	0.4	−3.3	−4.9	−6.9	−8.0	−11.0
SD $F2_{test}-F2_{pretest}$ (Hz)	13.5	11.3	14.3	14.3	15.5	18.6	18.1	20.3
% F2 change	1.2	−0.8	−0.3	2.4	3.6	5.1	5.9	8.2

mental condition where feedback F1 was moved upward, blocks five through eight had significantly larger absolute test-pretest formant changes in F1 than block one ($p < 0.05$). There were no significant differences for F2, or for F1 of the downward condition. Given the small response and large variability, the lack of significance in the downward post-hoc tests was probably because there were fewer participants in this condition.

B. Formant F1 change points and comparison of manipulation direction

Figure 2 also shows the change points calculated for F1 of each experimental condition. In Fig. 2(a), where feedback F1 was pushed upward, a significant change was found in F1 at a delay of 0.454 s with respect to stimulus onset. Figure 2(b), where feedback F1 was pushed downward, shows a significant change point at a delay of 0.415 s with respect to the stimulus onset.

Separate variance t tests showed that the absolute magnitude of the stimulus manipulation was not different between conditions (+136 Hz vs −135 Hz). The absolute response magnitude and percent compensation/change for F1 and F2 were also not different between conditions for the mean formants in block 8 (see Tables I and II). The F1 responses for the up/down directions can be compared qualitatively in Fig. 2 between panels (a) (F1 feedback up) and (b) (F1 feedback down).

Whereas Fig. 2 shows the average trial data, histograms of the compensation in each test trial from all participants are given in Fig. 3. The percent compensation in each test trial was estimated as 100 times the difference between blocks eight and one divided by the size of the manipulation. Note that for both panels, a positive percent compensation indicates that production changed to counter the feedback manipulation. A negative percent compensation on the horizontal axis indicates that F1 production followed the shift in F1 feedback. In panel (a), where feedback F1 was increased, trials with positive percent compensation had a decrease in F1 production. For panel (b), feedback F1 was decreased, so trials with positive percent compensation had an increase in produced F1. A correction was derived for participants from their pretest trials, in order to reduce bias in the percent compensation estimate due to the normal trajectory of F1 over the course of an utterance (the average percent change between blocks eight and one in the pretest trials, relative to the size of the test trial manipulation, was added to the values calculated for the test trials). These histograms show that the size of the response to feedback perturbation varies from trial to trial, as discussed further below. Both histograms are shifted slightly right of zero, indicating that both feedback manipulations were compensated in many trials, but very few trials approach complete compensation (positive values near 100%). Both distributions include zero and negative percent compensation because there were trials with no com-

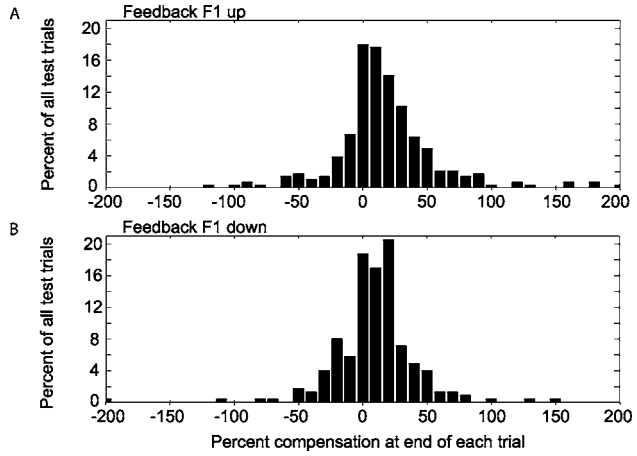


FIG. 3. Histograms of the percent compensation in all test trials from all subjects. The vertical axis shows the bin counts as a percentage of all test trials, rather than the absolute number of trials. Percent compensation on the horizontal axis was calculated as 100 times the F1 difference between blocks eight and one, divided by the (participant dependent) manipulation size. Each bin is 10% wide, where the zero bin spans from −5% to +5%. Positive values of percent compensation indicate that the change in production countered the feedback manipulation. Negative percent compensation indicates that F1 production changed in the same direction as the manipulation (a following response). Panels (a) and (b) show responses for the experimental conditions where feedback F1 was increased and decreased, respectively. Both distributions are slightly right of zero because both conditions elicited a majority of compensatory responses.

pensation, and those where F1 followed the manipulation. Figure 3 does not explicitly separate between- and within-subject variability, however, all participants had mean percent compensations $<30\%$. The spread of the histogram toward higher compensations is due to trial-to-trial variability within subjects, not because of a few participants with consistently high percent compensation.

IV. DISCUSSION

The data reported here reveal a modest sensitivity to unexpected formant perturbations in auditory feedback. Subjects, on average, produced compensatory changes in the frequency of their first formant in the opposite direction to unexpected changes in the auditory feedback of their own voice. When the manipulation of the feedback raised F1, the compensations were slightly larger on average (although not significantly so) than those in response to perturbations that lowered F1. However, considerable variability in the compensatory responses for both directions was evident within and between participants. An examination of the total distribution of responses (Fig. 3) found that the perturbation was completely compensated in a few of the test trials. In combination with reported rapid F0 compensations (e.g., Burnett *et al.*, 1998), these results suggest a general similarity between F0 and formant behavior when feedback is suddenly modified.

Incomplete compensation has been found previously in pitch-shift studies. For example, Fig. 3(b) of Burnett *et al.* (1998) shows that the response magnitude is largely independent of stimulus magnitude from 25 to 300 cents, leading to a lower percent compensation for the larger stimulus pitch shifts. Recent data (Leydon *et al.*, 2003) suggests that the pitch-shift response may have a linear operating range where a single gain relates output to input, whereas most studies have employed stimulus shifts that saturate the response at the limit of compensation. Since only a single formant shift value was used in this study, a more comprehensive comparison would be required to make any conclusions about the relative responsiveness of the speech motor system to formant modification.

Incomplete compensations have also been observed in the study of formant shifts using whispered steady-state vowels in an adaptation paradigm (Houde and Jordan, 1998; 2002). In the data of the eight reported subjects, the observed compensation varied between about 10% and 110% for a large feedback shift of two vowel categories (the cross-subject mean of their 1998 Fig. 3 is roughly 50%). Houde and Jordan carefully produced formants near hypothetical lines in formant space that connect adjacent vowels. According to Houde and Jordan, this path defines "perceptually salient directions" between different vowels, and identifies what formant frequencies can be realized by the vocal system. Manipulations should produce feedback vowels that are on the linear segments connecting vowel categories, if the manipulations are to be perceived and if compensation is to be possible. In the experiment reported here, only F1 was manipulated, and for most individuals this would push /ɛ/ off the path advocated by Houde and Jordon. They suggest that feedback vowels with off-path manipulations are projected onto the nearest intervowel category line. If this were the case, then manipulations of F1 alone would be effectively smaller as interpreted by the auditory vocal feedback system because the projected manipulation must be smaller in magnitude than the original manipulation vector.

Figure 3 shows that there were a small percentage of complete compensations and many responses near the zero percent value. One possible explanation is that the perturbations are not always detected by the speech motor system. Kewley-Port and colleagues (e.g., Kewley-Port and Watson, 1994; Kewley-Port and Zheng, 1999; Kewley-Port, 2001) have studied the psychoacoustic thresholds for the detection of changes in single formants. These studies have shown that well-trained individuals can detect changes in F1 as small as 14 Hz under ideal conditions (Kewley-Port and Watson, 1994). While this is significantly larger than the threshold for detecting changes in the frequency of tones (Moore, 1973), it is smaller than the manipulations that were employed here. Lack of training can raise the threshold for detecting F1 formant changes to about 45 Hz (Kewley-Port, 2001), and the presence of noise may increase this 20% to over 300% depending on SNR and noise characteristics (Liu and Kewley-Port, 2004). These thresholds are somewhat smaller than the average manipulations used here of approximately ±135 Hz with naiive listeners and some background noise. It is unknown whether thresholds for the detection of changes in one's own voice during production are similar to those for listening to the voices of others. It is possible that for some individuals the manipulation was subthreshold, but this seems tenuous given that F1 was moved an entire vowel category. However, only one subject identified that /ɛ/ was being manipulated, and only a few recognized (in retrospect) that there was some change in their vowel sounds after the experiment was explained in detail during an exit interview. It is also unknown whether compensation occurs for manipulations smaller than the psychoacoustic threshold for the detection of changes in one's own voice during production.

The variability in response magnitude observed here does not seem to be unique to the formant shift. While unaveraged individual trial data is unpublished, pitch-shift studies have previously reported that from 5% to 20% of subjects' averaged responses do not meet criteria to be "valid" for all stimulus conditions (e.g., Burnett *et al.*, 1998; Bauer and Larson, 2003; Sivasankar *et al.*, 2005). These invalid (non)responses from individuals are typically dropped from further analysis. In most pitch-shift reports, the between-subject response SDs are a large fraction of the average response magnitudes. Response magnitude varies with experimental design but some example values are approximately 50 (SD=20) cents (Donath *et al.*, 2002), 60 (SD =32) cents (Natke *et al.*, 2003), 38 (SD=33) cents (Larson *et al.*, 2001), and 28 (SD=15) cents (Burnett *et al.*, 1998). These example intersubject response magnitude SD values are relatively large, which is noted by Burnett *et al.* (1998) as "considerable between- and within-subject variability." The histograms in Fig. 3 show that there is a wide range in the percent compensation in individual test trials. A large proportion of test trials exhibit no compensation, whereas some

perturbations are robustly compensated. Overall, the distributions are shifted toward compensation (i.e., right of zero in Fig. 3), but in a sizable proportion of trials participants actually followed the manipulation, indicated by negative compensation bins.

A possible explanation why formant and F0 perturbations appear to be similar (other than actually being similar or the same) is that the task used in the sudden perturbation studies may induce a uniformity of response that does not exist in natural speech; the production of prolonged vowels may be influencing the results in these studies. This relatively unnatural speech task might alter the way in which auditory feedback is processed or change the importance of the information. In pitch-shift studies it has been found that task relevance has an influence on the characteristics of the compensatory response. Whereas many studies have applied pitch shifts to single syllable vocalizations that were produced for several seconds, another approach has been the use of multisyllabic nonsense words with different stress patterns (Natke and Kalveram, 2001). The latency of the compensatory response is long enough that it does not affect short, unstressed initial syllables, but it does, however, influence long, stressed initial syllables and subsequent syllables. Interestingly, the response persists after the pitch-shift stimulus is terminated, and influences the following word that may be uttered up to six seconds later (Donath *et al.*, 2002). Responses were found to be larger and the aftereffect longer, when subjects sang nonsense words in tune with a piano compared to speaking them (Natke *et al.*, 2003). This is presumably in part because F0 control is more important in singing than speaking for nontonal languages. Although Jones and Munhall (2002) found similar compensation latencies for speakers of the tonal language Mandarin relative to English speakers, there were no Mandarin speakers with the pitch-following response that has been reported in English participants (Burnett *et al.*, 1997). Comparisons within Mandarin have shown that response latency is shorter and magnitude is greater when pitch shifts are applied during dynamic tones, relative to static tones (Xu *et al.*, 2004). This supports the hypothesis that the response is modulated by the importance of pitch to the task. The relationship is, however, complex. In trained singers who spoke a nontonal language, the response latency was longer and the compensation magnitude smaller in singing when the target pitch was a dynamic glissando instead of a steady value (Burnett and Larson, 2002).

These results raise the possibility that the importance or salience of the auditory feedback can be modulated over time. Such modulation may also account for substantial within-subject variability. Average formant SDs were 38 Hz for F1 and 53 Hz for F2. These are large compared to the average changes between test and pretest trials (i.e., compensation) shown in Tables I and II. The within-subject formant SDs can be compared to values from other investigations of vowel production. The SDs here are in the same range as those reported in the vowel imitation literature (e.g., Kent and Forner, 1979; Repp and Williams, 1987; Vallabha and Tuller, 2004), and vowel production studies (e.g., Pisoni, 1980; Perkell and Nelson, 1985; Beckman *et al.*, 1995).

The pitch-shift literature has demonstrated a relatively rapid response to perturbations. Unfortunately the present study was not designed to measure minimum response time. Rather, the relatively slow onset of altered auditory feedback over 500 ms was chosen to try and maximize any compensatory response, under the assumption that formant feedback monitoring may be similar to that in pitch-shift paradigms (Larson *et al.*, 2000). This slow onset made it difficult to determine response latency since it was not known when the stimulus amplitude crossed the hypothetical detection or response initiation threshold. Response delay was estimated using the change point test, but the true physiological delay should be smaller since the early part of the altered feedback would presumably be subthreshold. Indeed, the mean change point times for F1 are longer than previously reported latencies for the pitch-shift response, such as 100 to 225 ms (Burnett *et al.*, 1997), or means of 192 and 266 ms (Burnett *et al.*, 1998). Larson *et al.* (2000) found mean upward responses were 217 vs 273 ms for downward responses. With the stimulus used here, it was not possible to determine whether formant compensation can occur as quickly as for the pitch-shift response. A more rapid transition to altered auditory feedback would allow a better estimate of response latency. In pilot measurements, a cross-fade of 20 ms was employed, and on average individuals showed both compensation and in some cases following responses.

Although only F1 was manipulated, small concomitant changes were also observed in the second formant, as suggested by trends in the F2 data of Tables I and II. This is not unexpected since formants are created by constrictions in the vocal tract (Fant, 1970), and it is unlikely that the map between articulation and acoustic goals learned as an infant would require or employ independent control of F1 and F2. Rather, a movement of vocal tract constriction would likely change both F1 and F2 because they are coupled in the resonant system (Stevens, 1998). The average data in Tables I and II suggest that the formant changes are appropriate for compensation along trajectories between vowel groups. For example, in Table II, F1 of /ɛ/ was perturbed downward toward /ɪ/. The average responses in the later blocks (e.g., six through eight) show the production of something in the direction of the /æ/ category, with higher F1 and lower F2. This would partially counter the lower F1 and higher F2 of the /ɪ/ category, despite the fact only F1 was manipulated. Similar but reversed trends are seen in Table I where F1 was shifted toward /æ/.

The influence of altered auditory feedback could have been attenuated by direct mouth to ear airborne sound, and through bone-conducted sound. A natural unprocessed voice signal that reaches the cochlea competes with the altered auditory feedback provided by the headphones. In this experiment the headphones themselves provided some isolation from airborne sound of the voice emitted at the mouth. It would be inappropriate, however, to use standard sound attenuation headphones with an acoustically closed back. Such headphones emphasize the bone-conducted signal radiated into the ear canal through the occlusion effect (Tonndorf, 1972). Headphones can be carefully designed to reduce airborne signals in the ear canal and simultaneously avoid the

occlusion effect through the use of larger cavities (Békésy, 1932; Khanna *et al.*, 1976), but they are unwieldy in practice. Bone-conducted sound radiated into the ear canal is not the only way for speech vibrations to reach the cochlea; there are other mechanisms of bone conduction (inertial and compressional) that cannot be mitigated with human participants (Tonndorf, 1972). Bone-conducted sound can have a significant influence on the net signal at the cochlea for frequencies in the range of F1 (Porschmann, 2000). To try and control the relative influence of the unprocessed bone-conducted voice, participants were discouraged from speaking above the normal conversational level used to set microphone gain, and headphone feedback was designed to dominate the response by presentation at a relatively high level (80 dBA SPL at each ear). Bone conduction transfer functions are highly individual (Purcell *et al.*, 2003), and it is possible that in some individuals the bone-conducted voice was sufficient to attenuate the desired illusion of production errors.

The present data are generally consistent with the idea that a single type of feedback system governs both F0 and formant production. Under similar testing conditions, both fundamental frequency and formants show rapid responses to feedback changes and show evidence for adaptation (Jones and Munhall, 2000; Houde and Jordan, 1998). The similarity of operational principles that the data implies does not mean that a single system is necessarily controlling both aspects of speech. Indeed, the intermediate vowel /ɛ/ may be more labile than point vowels since it does not benefit from a saturation effect (Svirsky and Tobey, 1991; Perkell *et al.*, 2000). It also can be shown that the nervous system is capable of learning and simultaneously maintaining more than one independent motor controller (Ghahramani and Wolpert, 1997). In order to test this possibility for speech motor control, studies will have to be specifically designed to test learning and interactions between multiple models of auditory feedback (Wolpert and Kawato, 1998). In addition, the conditions that modulate the actions of these feedback systems over time must be addressed.

V. CONCLUDING REMARKS

The study reported here employed paradigms from the vocal pitch-shift literature to evaluate whether the auditory vocal system compensates for sudden changes in the first formant of vowels. Partial compensation took place similar to that reported for manipulations of F0. The response was quite variable, and was smaller than previously measured in formant adaptation studies with whispered speech. The presence of immediate compensations in F0 and formant frequency, as well as sensorimotor adaptation in both parameters, suggests that the auditory feedback system for both aspects of speech is similar.

ACKNOWLEDGMENTS

The authors would like to thank Reiner Wilhelms-Tricarico and Richard McGowan for their assistance in the development of the formant tracking and filtering algorithms. Research supported by the National Institute on Deafness and Other Communication Disorders, and the (Canadian) Natural Sciences and Engineering Research Council.

Bauer, J. J., and Larson, C. R. (**2003**). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," J. Acoust. Soc. Am. **114**, 1048–1054.

Beckman, M. E., Jung, T. P., Lee, S. H., Dejong, K., Krishnamurthy, A. K., Ahalt, S. C., Cohen, K. B., and Collins, M. J. (**1995**). "Variability in the production of quantal vowels revisited," J. Acoust. Soc. Am. **97**, 471–490.

Békésy, G. v. (**1932**). "Zur theorie des hörens bei der schallaufnahme durch knochenleitung," Ann. Phys. **13**, 111–136.

Binnie, C. A., Daniloff, R. G., and Buckingham, H. W., Jr. (**1982**). "Phonetic disintegration in a five-year-old following sudden hearing loss," J. Speech Hear. Disord. **47**, 181–189.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch-shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**, 1058–1063.

Burnett, T. A., Senner, J. E., and Larson, C. R. (**1997**). "Voice F0 responses to pitch-shifted auditory feedback: A preliminary study," J. Voice **11**, 202–211.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Cowie, R., and Douglas-Cowie, E. (**1992**). *Postlingually Acquired Deafness* (Mouton de Gruyter, New York), p. 304.

Cowie, R., Douglas-Cowie, E., and Kerr, A. G. (**1982**). "A study of speech deterioration in post-lingually deafened adults," J. Laryngol. Otol. **96**, 101–112.

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). "Effects of frequency-shifted auditory feedback on voice F0 contours in syllables," J. Acoust. Soc. Am. **111**, 357–366.

Economou, A., Tartter, V. C., Chute, P. M., and Hellman, S. A. (**1992**). "Speech changes following reimplantation from a single-channel to a multichannel cochlear implant," J. Acoust. Soc. Am. **92**, 1310–1323.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague).

Ghahramani, Z., and Wolpert, D. M. (**1997**). "Modular decomposition in visuomotor learning," Nature **386**, 392–395.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Held, R. (**1965**). "Plasticity in sensory-motor systems," Sci. Am. **213**, 84–94.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.

Kent, R. D., and Forner, L. L. (**1979**). "Developmental study of vowel formant frequencies in an imitation task," J. Acoust. Soc. Am. **65**, 208–217.

Kewley-Port, D. (**2001**). "Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training," J. Acoust. Soc. Am. **110**, 2141–2155.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Kewley-Port, D., and Zheng, Y. (**1999**). "Vowel formant discrimination: Towards more ordinary listening conditions," J. Acoust. Soc. Am. **106**, 2945–2958.

Khanna, S. M., Tonndorf, J., and Queller, J. E. (**1976**). "Mechanical parameters of hearing by bone conduction," J. Acoust. Soc. Am. **60**, 139–154.

Ladefoged, P. (**1982**). *A Course in Phonetics* (Harcourt Brace Jovanovich, Inc., New York).

Ladefoged, P., Silverstein, R., and Papcun, G. (**1973**). "Letter: Interruptibility of speech," J. Acoust. Soc. Am. **54**, 1105–1108.

Lane, H., and Webster, J. W. (**1991**). "Speech deterioration in postlingually deafened adults," J. Acoust. Soc. Am. **89**, 859–866.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparison of voice F0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**, 2845–2848.
Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). "Effects of pitch-shift velocity on voice F0 responses," J. Acoust. Soc. Am. **107**, 559–564.
Leder, S. B., and Spitzer, J. B. (**1993**). "Speaking fundamental frequency, intensity, and rate of adventitiously profoundly hearing-impaired adult women," J. Acoust. Soc. Am. **93**, 2146–2151.
Leder, S. B., Spitzer, J. B., and Kirchner, J. C. (**1987**). "Speaking fundamental frequency of postlingually profoundly deaf adult men," Ann. Otol. Rhinol. Laryngol. **96**, 322–324.
Leydon, C., Bauer, J. J., and Larson, C. R. (**2003**). "The role of auditory feedback in sustaining vocal vibrato," J. Acoust. Soc. Am. **114**, 1575–1581.
Liu, C., and Kewley-Port, D. (**2004**). "Formant discrimination in noise for isolated vowels," J. Acoust. Soc. Am. **116**, 3119–3129.
Miall, R. C., and Wolpert, D. M. (**1996**). "Forward models for physiological motor control," Neural Networks **9**, 1265–1279.
Moore, B. C. (**1973**). "Frequency difference limens for short-duration tones," J. Acoust. Soc. Am. **54**, 610–619.
Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.
Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.
Orfandidis, S. J. (**1988**). *Optimum Signal Processing, An Introduction* (MacMillan, New York).
Perkell, J., Lane, H., Svirsky, M., and Webster, J. (**1992**). "Speech of cochlear implant patients: a longitudinal study of vowel production," J. Acoust. Soc. Am. **91**, 2961–2978.
Perkell, J. S., and Nelson, W. L. (**1985**). "Variability in production of the vowels /i/ and /a/," J. Acoust. Soc. Am. **77**, 1889–1895.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000**). "A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonetics **28**, 233–272.
Pisoni, D. B. (**1980**). "Variability of vowel formant frequencies and the quantal theory of speech: A first report," Phonetica **37**, 285–305.
Porschmann, C. (**2000**). "Influences of bone conduction and air conduction on the sound of one's own voice," Acta Acust. **86**, 1038–1045.
Purcell, D. W., Kunov, H., and Cleghorn, W. (**2003**). "Estimating bone conduction transfer functions using otoacoustic emissions," J. Acoust. Soc. Am. **114**, 907–918.
Repp, B. H., and Williams, D. R. (**1987**). "Categorical tendencies in imitating self-produced isolated vowels," Speech Commun. **6**, 1–14.
Schenk, B. S., Baumgartner, W. D., and Hamzavi, J. S. (**2003**). "Effect of the loss of auditory feedback on segmental parameters of vowels of postlingually deafened speakers," Auris Nasus Larynx **30**, 333–339.
Siegel, S., and Castellan, N. J. (**1988**). *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, Boston).
Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (**2005**). "Voice responses to changes in pitch of voice or tone auditory feedback," J. Acoust. Soc. Am. **117**, 850–857.
Stevens, K. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).
Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.
Svirsky, M. A., and Tobey, E. A. (**1991**). "Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users," J. Acoust. Soc. Am. **89**, 2895–2904.
Svirsky, M. A., Lane, H., Perkell, J. S., and Wozniak, J. (**1992**). "Effects of short-term auditory deprivation on speech production in adult cochlear implant users," J. Acoust. Soc. Am. **92**, 1284–1300.
Tonndorf, J. (**1972**). "Bone Conduction," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), pp. 195–237.
Vallabha, G. K., and Tuller, B. (**2002**). "Systematic errors in the formant analysis of steady-state vowels," Speech Commun. **38**, 141–160.
Vallabha, G. K., and Tuller, B. (**2004**). "Perceptuomotor bias in the imitation of steady-state vowels," J. Acoust. Soc. Am. **116**, 1184–1197.
Villacorta, V., Perkell, J. S., and Guenther, F. H. (**2004**). "Sensorimotor adaptation to acoustic perturbations in vowel formants," J. Acoust. Soc. Am. **115**, 2430.
Villacorta, V., Perkell, J. S., and Guenther, F. H. (**2005**). "Relations between speech sensorimotor adaptation and perceptual acuity," J. Acoust. Soc. Am. **117**, 2618.
Waldstein, R. S. (**1990**). "Effects of postlingual deafness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**, 2099–2114.
Wolpert, D. M., and Kawato, M. (**1998**). "Multiple paired forward and inverse models for motor control," Neural Networks **11**, 1317–1329.
Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**, 1168–1178.

Coherence detection: Effects of frequency, frequency uncertainty, and onset/offset delays

Rong Huang[a] and Virginia M. Richards
Department of Psychology, 3401 Walnut Street, Suite 302C, University of Pennsylvania, Philadelphia, Pennsylvania 19104

(Received 2 June 2005; revised 1 February 2006; accepted 2 February 2006)

The detectability of a sequence of equal-frequency (coherent) tonal components embedded in random, multiburst maskers was evaluated. The masker was comprised of tonal components located in a time-by-frequency spectrogram with eight 30 ms time columns and 29 frequency rows ranging logarithmically from 200 to 5000 Hz. The probability that a tone occurred in any one cell of the spectrogram, p, was the independent variable. The signal and masker components were of equal duration and equal level. Using a yes/no procedure, threshold values of p were estimated for five signal frequencies (220, 445, 1000, 2245, 4490 Hz) and when the signal frequency was random. Thresholds were worst for the random-frequency signal and best for the fixed 1000 Hz signal. In additional conditions, the value of p was fixed and the signal components were delayed relative to the masker components. A 1 ms delay provided better sensitivity (d' grew from 0.5 to 1) for all but the lowest signal frequency tested. An analysis of no-signal trials revealed that false alarm rates were higher when components falling at the signal frequency were consecutive than when they were distributed across bursts. Thus, coherence rather than total energy at the signal frequency is important for signal detection. [DOI: 10.1121/1.2179730]

PACS number(s): 43.66.Mk [JHG] Pages: 2298–2304

I. INTRODUCTION

In a 1995 study, Kidd *et al.* examined listeners' ability to detect a "coherent frequency pattern" (i.e., a sequence of equal-frequency tones) among sequences of random-frequency, multitone bursts. Both the number of random-frequency tones in each burst (M) and the number of coherent signal components (N) were varied. Sensitivity increased systematically as N grew, as M fell, and as the number of bursts increased. Examining the results obtained for $N=1$, Kidd *et al.* noted that sensitivity was somewhat poorer than one might expect based on their earlier work (Kidd *et al.*, 1994). They suggested that this discrepancy in sensitivity might reflect a variety of factors, including the fact that the older study used a signal frequency that was fixed, rather than random, across trials. Finally, Kidd *et al.* (1995) found that listeners could withstand substantial frequency deviation of the coherent signal; when the frequencies of the sequential signal components were perturbed across a 14% range, thresholds were indistinguishable from thresholds obtained when there was no perturbation (coherence). Taking these factors into account, Kidd *et al.* (1995) suggested that their results reflected the sensitivity of the auditory system to frequency continuity between bursts, i.e., sequential grouping or sequential stream segregation (e.g. Moore, 2003, Moore and Gockel, 2002).

Using a task similar to the one used by Kidd *et al.* (1995), Richards and Tang (2005) estimated relative weights using a coherent signal whose frequency was fixed. Their results suggested that listeners attended selectively to the signal frequency (positive weights at the signal frequency when the no-signal trials were used to derive relative weights). Given the potential impact of frequency uncertainty on the listeners' ability to detect a coherent signal (Kidd *et al.*, 1994; Kidd *et al.*, 1995), and given the evidence that for fixed signal frequencies listeners appeared to attend to the signal frequency, it was of interest to provide a direct comparison of the detectability of a coherent signal whose frequency was either fixed or random across trials.

To evaluate the effect of signal-frequency uncertainty, an experiment similar to Kidd *et al.*'s (1995) was designed. The masker was generated using a time-by-frequency spectrogram with eight 30 ms time columns and 29 frequency rows. The independent variable was the probability a tone was present in any one cell of the spectrogram. The signal was a sequence of coherent tones. For the fixed signal frequency five signal frequencies were tested. In additional measurements, the signal frequency was random across trials. The resulting data speak to both the effects of signal frequency and the effects of frequency uncertainty.

The detection of a coherent sequence of signal components in competition with random masker components depends on "sequential grouping" (Bregman and Pinker, 1978; Moore, 2003). In a second series of measurements, the signal components were delayed relative to the masker components. Acknowledging that synchronous onsets contribute to the "simultaneous grouping" of multiple frequencies of a single sound (Bregman and Pinker, 1978; Gockel, 2000), we presumed the difference in the onset (and offset) between the signal and masker components would aid in the segregation of the signal from the masker. A second aim of the experi-

[a]Author to whom correspondence should be addressed: Rong Huang. Telephone; 215-898-4587 electronic mail: rongh@psych.upenn.edu

ments, then, was to provide an empirical estimate of the degree to which onset/offset differences could unmask a coherent signal presented among random maskers.

There is, unfortunately, a potential confound associated with introducing a delay between the signal and masker components to enhance the segregation of the coherent signal. Sensitivity might increase simply because listeners can detect the delay between the signal and masker components—regardless of whether the signal is heard as well separated from the masker. Zera and Green (1993) reported an exhaustive study of sensitivity to onset and offset delays for multitone complexes. When one tone among many was delayed, thresholds of delay (ms) tended to decline as the frequency of the delayed signal increased. Although there are numerous differences between the stimuli tested in the current experiment and those tested by Zera and Green (1993), their paper provides a guideline to answer the question of whether our listeners depended on delay alone versus coherence in making their detection decisions.

II. METHODS

A. Listeners

Four normal-hearing listeners ranging in age from 21–25 were paid for their participation. All had thresholds in quiet of 15 dB HL or less at the octave frequencies from 250 to 8000 Hz. Participants were tested individually in double-walled soundproof booths.

B. Stimuli

The random masker can be described in terms of tones appearing, probabilistically, at cells of a time-by-frequency spectrogram. The spectrogram was defined using eight time columns of 30 ms each and 29 frequency rows ranging from 200 to 5000 Hz on a logarithmic scale. As a result, the components were, at their nearest, separated by 1/6 of an octave. The independent variable was p, which is the probability a tone would be present at each cell of the spectrogram. The signal was defined by altering the masker as follows. For each burst, one of the randomly chosen masker components was shifted to the signal frequency, thereby forming a coherent sequence of tones at the signal frequency. This generation procedure resulted in two important stimulus features. First, the signal and masker components were of equal level; 60 dB SPL. Second, the total expected number of tones in the signal and no-signal stimuli was the same. Note that as p increases, performance is expected to decrease. For example, considering just the no-signal stimulus, large values of p would sometimes lead to series of consecutive tones at the signal frequency.

Across conditions, there were a total of six signal frequencies/types. The five fixed signal frequencies were 220, 445, 1000, 2245, and 4490 Hz. A random signal frequency was also tested; on each trial the signal frequency was randomly chosen from the 29 potential frequencies, except the highest and lowest frequencies. These six signal types will be referred to as the *Low, MidLow, Mid, MidHigh, High*, and *Ran* signals, respectively.

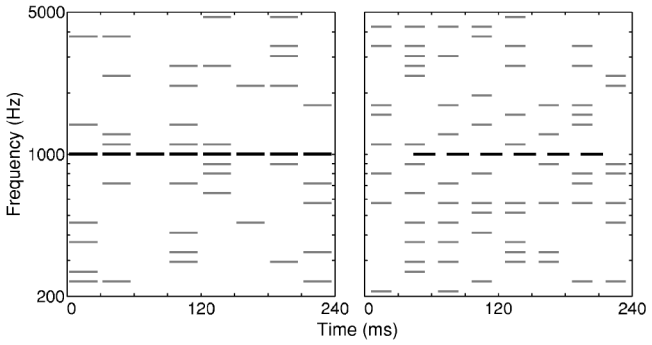


FIG. 1. Examples of the signal stimuli are shown for condition 1 (left) and conditions 3–5 in which the signal tones are delayed relative to the masker tones (right). The abscissa is time (ms) and the ordinate is frequency (Hz; log scale). The signal to be detected is a sequence of coherent 1000 Hz tones, the *Mid* signal. Values of p used to generate this graph were 0.25 and 0.4 for the left and right panels, respectively. The signal-masker *delay* in the right panel is 9 ms.

The stimuli were digitally generated and presented using a single channel of a 16-bit DAC with a sample rate of 20 kHz and low pass filtered at 7 kHz (Stewart VBF 10M). The stimuli were presented diotically via Sennheiser HD410SL headphones.

C. Procedures

A yes/no procedure was used. Data were collected using 50-trial blocks and listeners were encouraged to rest after five of the 50-trial blocks. Listeners indicated the absence/presence of a coherent signal by pressing keyboard keys "1" or "3," respectively. Visual feedback as to the correctness of each response followed each trial, and after a block of 50 trials summary statistics were provided to the listener.

Five conditions are reported. In an effort to examine the effects of signal frequency and the effects of signal frequency uncertainty, the first two conditions estimated the listeners' sensitivity to a coherent signal for six signal types: five fixed-frequency and a random-frequency signal. The remaining conditions, in which the signal components were delayed relative to the masker components, evaluated the contribution of signal/masker delay to the detection of a coherent signal.

Condition 1. The signal was composed of eight 30 ms tones that were turned on and off synchronously with the masker using 5 ms cosine-squared ramps. There was no gap between the offset of one burst and the onset of the next. Six signal types were tested (*Low, MidLow, Mid, MidHigh, High*, and *Ran*) one after another in random order for each listener. Three-point psychometric functions were obtained using three values of p estimated during 200 or more practice trials. The three p's yielded d's of approximately 0.5, 1, and 1.5. Listeners completed five 50-trial blocks for each of the three values of p in random order, and then repeated the measures in reverse order. A least squares linear fit was used to estimate the threshold value of $p(d'=1)$. All listeners finished condition 1 before moving onto conditions 2 and 3.

The left panel of Fig. 1 provides an example of the 1000 Hz signal stimulus for this condition. The ordinate is frequency on a logarithmic scale and the abscissa is time in

ms. The coherent tones are plotted using bold lines—this aid to visualization is not intended to indicate that signal tones had different levels than the masker tones. The apparent gaps between bursts depict onset/offset ramps; there was no delay between bursts in this condition. For this example, the value of p was 0.25.

Condition 2. Condition 2 was similar to condition 1, except as follows. First, all the tones were 20 ms in duration; i.e., there was a 10 ms temporal gap between bursts. Second, in this and the remaining conditions the signal was a sequence of six coherent tones, instead of eight. Note that for the signal stimulus, the six signal tones occupied the central six bursts and masker tones might occupy the first and last temporal epoch at the signal frequency. Third, psychometric functions were generated using five to seven (rather than three in condition 1) values of p's. Fourth, unlike condition 1, data collection was not blocked by p. The resulting psychometric functions, based on a total of 500 to 750 trials, were fitted using a weighted linear least squares fit relating d's and values of p to estimate the threshold ($d'=1$). Listeners ran each of the different signal types in a randomly chosen order. Conditions 2 and 3 were run sequentially for each signal type. Condition 2 and remaining conditions shared many of the features, and as such condition 2 acted as a standard for comparison.

Condition 3. As in condition 2, in this condition the tones were 20 ms in duration and were separated by 10 ms gaps and the signal was defined by six sequential tones. Compared to condition 2, in this and the remaining conditions a delay between the signal and masker components was introduced. Note that because the signal and masker components had equal duration, both the onsets and offsets of the signal tones were delayed relative to the masker tones. In this condition the p was fixed, and the independent variable was the amount of delay between the signal tones and the masker tones.

A two-step procedure was adopted in this experiment. In the first step, the value of p that led to a d' of 0.5 was estimated (from condition 2). This assured that all psychometric functions (i.e., for each signal type and each listener) started at $d' \approx 0.5$ when the delay was 0 ms. In the second step p was fixed and the signal-masker *delay* value was altered to generate psychometric functions. Approximately 500–750 trials contributed to each psychometric function. A weighted linear fit between *delay* and d' was used to estimate the threshold *delay* value ($d'=1$). An example of a delayed 1000 Hz signal stimulus for condition 3 is shown on the right panel of Fig. 1. For this example, $p=0.4$ and *delay*=9 ms.

Condition 4. The signal stimuli in this condition were generated as in condition 3. Compared to the previous conditions, only the 1000 Hz fixed signal frequency was tested. Also in contrast to condition 3, in this condition the value of *delay* was fixed and p was altered to generate psychometric functions. Two values of *delay* were tested: 2 and 4 ms. Approximately 500–750 trials contributed to each psychometric function. Psychometric functions for no delay (*delay*=0) were taken from condition 2. This provided a comparison of psychometric functions as a function of p for three different values of *delay*. The results presented below were taken from a larger dataset in which several delays were evaluated. As a result, the order in which the delays were tested was random across listeners, but not counterbalanced.

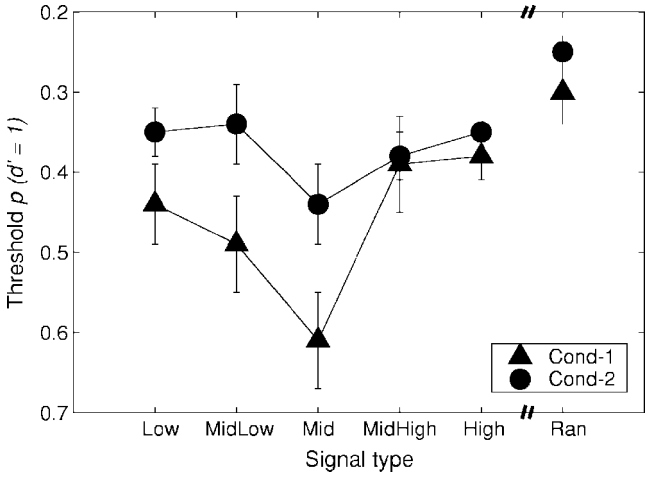


FIG. 2. Threshold ($d'=1$) values of p are plotted for the different signal types for conditions 1 (triangles) and 2 (circles). Error bars indicate the standard errors of the mean across listeners. Note that larger values of p are lower on the ordinate.

Condition 5. In this condition we examined the possibility that there might be an interaction between p and signal-masker *delay*. Toward that end, many (p, *delay*) pairs were tested. The signal was a sequence of 6, 20 ms, 1000 Hz tones. Values of p ranged from 0.1 to 0.9 in steps of 0.1 and values of *delay* ranged from 0 to 6 ms using a 1 ms step size. Not all possible combinations were tested; low values of p paired with high values of *delay* led to near-perfect performance and high values of p paired with low values of *delay* led to near-chance performance. Data from 150 trials were used to estimate d' for each (p, *delay*) pair.

Figure 1 summarizes the main differences among conditions. First, the effect of changing p from 0.25 to 0.4, which increases the total number of components, is apparent by comparing the left and right panels. Second, the figure emphasizes the difference in the component duration (30 *vs* 20 ms) and the number of coherent tones used to define the signal (8 *vs* 6) in condition 1 compared to the other conditions, respectively. Finally, as long as the signal/masker *delay* is less than 10 ms, there would be no temporal overlap between the signal offset and the onset of the subsequent burst.

III. RESULTS AND DISCUSSION

Figure 2 shows the threshold estimates of p averaged across four listeners for conditions 1 and 2 (triangles and circles, respectively). The ordinate is ordered such that higher values of p are plotted toward the bottom; i.e., superior thresholds are plotted as lower on the ordinate. The abscissa indicates the six signal types with the threshold for the random signal frequency being plotted to the right. Error bars indicate the standard errors of the mean across four listeners. Thresholds were higher in condition 1 than in condition 2, especially for signals of 1000 Hz and lower. We had

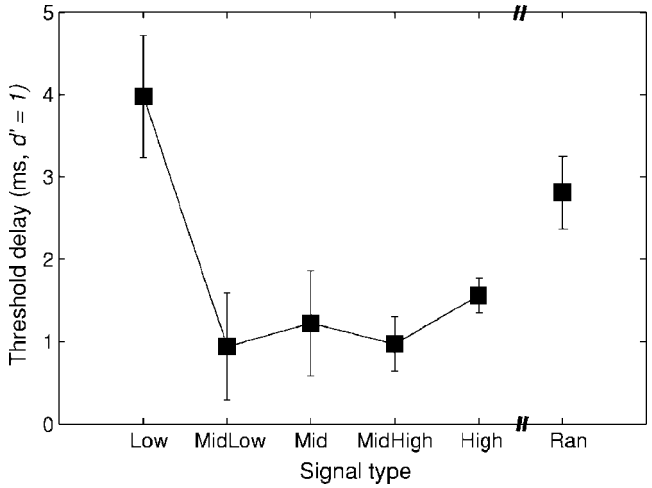


FIG. 3. Threshold ($d'=1$) values of *delay* are plotted for the different signal types for condition 3. Error bars indicate the standard errors of the mean across listeners.

expected higher thresholds in condition 1 for two reasons. First, in condition 1 eight coherent tones defined the signal whereas only six defined the signal in condition 2 (e.g., Kidd *et al.*, 1995). Second, in condition 1 the tones were of longer duration, and there was no gap between bursts (Kidd *et al.*, 2003). Neither of these factors, however, predicts the obtained frequency effect.

In both conditions 1 and 2, sensitivity was better for the *Mid* (1000 Hz) signal than the other signal types. This held true for three of the four listeners in both conditions. The magnitude of the effect, however, was larger in condition 1 than in condition 2. This result might reflect one of at least two factors: that listeners were most sensitive at 1000 Hz *per se*, or that listeners were more sensitive to the 1000 Hz signal frequency because it was central in the range of frequencies tested. Profile analysis experiments using approximately the same frequency range as the current experiments also revealed maximal sensitivity to an intensity increment near the central, 1000 Hz frequency (Green and Mason, 1985; Bernstein and Green, 1987; Green *et al.*, 1987; Green and Berg, 1991). For profile analysis, it appears that the centrality of the signal, not the frequency *per se*, is the important variable. The current data do not speak to this question for coherence detection.

The random-frequency signal yielded the lowest averaged threshold, a result obtained for all listeners in both conditions. The superior sensitivity associated with a central signal frequency coupled with the cost of signal frequency uncertainty suggests that the processes underlying the detection are likely to depend on some form of across-frequency contrast.

Figure 3 shows the threshold *delay* averaged across listeners in condition 3 for the six signal types. Recall that the value of p was chosen separately for each listener such that a d' of 0.5 was approximately achieved for a delay of 0 ms (i.e., synchronous signal and masker tones). Table I lists the values of p that were tested for each listener and each signal type. Because the minimum delay of 0 ms generated a d' of 0.5, the estimated thresholds of *delay* reflect changes in the slope of the estimated psychometric functions. Error bars indicate the standard errors of the mean across the four listeners. Thresholds are near 1 ms for most signal types, but are elevated for the *Low* (4 ms) and *Ran* (3 ms) signal types. For this condition some of the psychometric functions (for one listener the *Low* signal and for another listener the *Ran* and *MidLow* signals) were poorly fitted (the functions accounted for less than 50% of the variance available to be accounted for). Even so, higher thresholds for the *Low* and *Ran* conditions were readily apparent in the individual datasets.

TABLE I. Values of p used in condition 3 which yield $d' \approx 0.5$ for a delay of 0 ms.

	Low	*MidLow*	*Mid*	*MidHigh*	*High*	*Ran*
L1	0.40	0.30	0.50	0.50	0.40	0.30
L2	0.40	0.50	0.50	0.40	0.40	0.25
L3	0.40	0.30	0.40	0.35	0.40	0.30
L4	0.35	0.45	0.65	0.40	0.40	0.30

Although comparisons across very different studies are difficult, the current results are in reasonable agreement with those of Zera and Green (1993). Zera and Green studied the listeners' ability to detect a late onset, and separately a late offset, for one of M tones, where the M tones were logarithmically spaced across a range of 200–4000 Hz. Here we consider the results for their 10-tone stimulus, which is similar to the average of 12 tones per burst in this condition.[1] For delayed onsets, Zera and Green found that threshold *delays* were highest for the lowest frequency tested, and then thresholds fell as the signal frequency increased. For delayed offsets, threshold delays were relatively independent of frequency, except thresholds were slightly higher for the low-frequency signals and slightly lower for the high-frequency signals. For both onset and offset delays, thresholds, at their best, were approximately 2 ms. The current thresholds of *delay* (about 1 ms) are of the same order of magnitude as those reported by Zera and Green, indicating surprisingly good agreement given the vast differences in the experiments. On the other hand, the pattern of threshold *delays* as a function of frequency for the current experiment differs from the patterns obtained by Zera and Green. The contribution of delay *per se* to coherence detection is further considered after the results of condition 4 are described.

Figure 4 plots the results of condition 4 for each listener. Psychometric functions relating d' and p are plotted for three values of *delay*, 0, 2, and 4 ms (asterisks, diamonds, and circles, respectively), of the 1000 Hz signal. For each *delay*, decreases in p led to higher d's. The slope of the psychometric functions were somewhat similar across delays, except for L1 and L4 when the delay was 4 ms. Regarding these two "flat" psychometric functions, two plausible explanations include (a) a ceiling effect and (b) the detectability of the signal relies on the delay within the stimulus, not the presence of coherence *per se*. The latter explanation seems the most likely because d's were high when p was large. The extreme condition is for L4, who achieved a d' of 3 when p was 1. Note that when p was 1 there was no difference be-

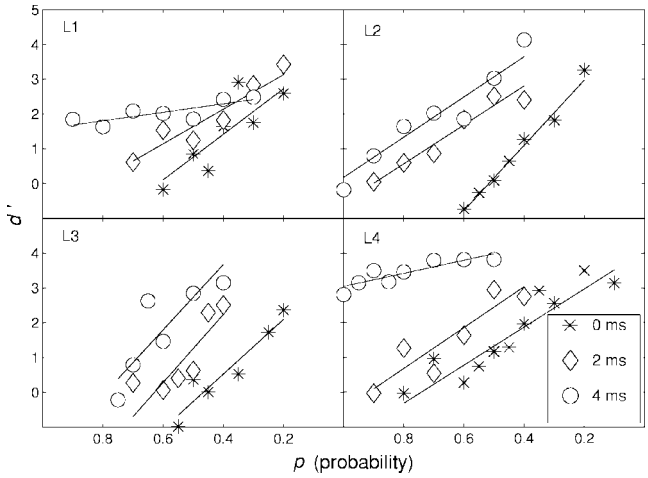


FIG. 4. Psychometric functions, d' as a function of p, are plotted for each listener in four panels for condition 4. The parameter indicates the signal *delay*, 0, 2, and 4 ms (asterisks, diamonds, and circles). The signal frequency is 1000 Hz and the lines show linear least squares fits to the datasets.

tween the no-signal and signal trials, except for the delayed signal components.

The results of condition 5 are in Table II, which shows the averaged d's and the standard errors of the mean across listeners for several (p, *delay*) pairs. The averages listed in Table II are reasonably representative of the individual data, except for L4 when the delay was 3 ms the effect of p was small (i.e., the psychometric function was flat). Sensitivity improved as p decreased or *delay* increased. The bold entries in Table II suggest a weak trade between p and *delay*, in that the d's were approximately equal along the diagonal. We refer to the pattern of data as displaying a "weak" trade because the underlying psychometric functions are not strictly parallel.

To summarize: The results of first two conditions demonstrated a cost associated with signal frequency uncertainty and a benefit associated with the 1000 Hz signal frequency compared to signals of lower and higher frequencies. Drawing a parallel with profile analysis experiments (e.g., Green *et al.*, 1987), it is plausible that the superior sensitivity at 1000 Hz reflected the fact that the 1000 Hz was centrally located in the log-frequency range of potential components. The advantage of delaying the signal components relative to the masker components improved coherence detection, although the advantage was muted for the lowest signal frequency. Finally, at least for a 1000 Hz coherent signal, increases in p could be counterbalanced by increases in signal/masker *delay*. In the next section, we consider whether or not the fact that the signal was comprised of *coherent* tones, compared to increases in *energy* at the signal frequency, was a critical element of the listeners' decision processes.

A. Analysis of no-signal trials

The results show that as p increased, the listeners' ability to detect the coherent signal worsened. Several factors might contribute to this relation (see also Kidd *et al.*, 1995). First, consider the signal trials. As p increased the number of masker components increased, making it more likely that the masker would mask the signal (i.e., increased *energetic* masking). However, because the masker and signal components were equal in level, and the masker components were separated from the signal frequency by approximately 1/6 of an octave, it seems unlikely that energetic masking could account for the reduced sensitivity that occurred when p increased. Second, for no-signal trials, as p increased, the likelihood that a sequence of consecutive tones occurred at the signal frequency also increased, potentially producing increased false alarms (FAs). Critically, for large p's, listeners may have produced more FAs simply because there was substantial energy (i.e., not necessarily consecutive tones) at the known signal frequency.

In an effort to disentangle the contribution of the presence of *coherent* tones at the signal frequency, rather than increased *energy* at the signal frequency, we examined the FA rates as a function of the number of consecutive tones (coherence) versus the total number of tones (energy) at the signal frequency. This analysis was applied only to condition 1 because the other conditions did not have enough trials for such an analysis. Figure 5 shows the results from all listeners for the fixed-frequency signal types of condition 1. The ordinate is the FA rate and the abscissa is the number of tones

TABLE II. Averaged d's (4 listeners) for (p, *delay*) pairs tested in condition 5. Standard errors of the mean are in parentheses.

		p				
		0.7	**0.6**	**0.5**	**0.4**	**0.3**
	0		0.48	0.79	1.08	**2.31**
			(0.22)	(0.30)	(0.54)	(0.40)
	1	0.27	0.74	0.69	**1.65**	2.02
		(0.28)	(0.32)	(0.30)	(0.45)	(0.33)
delay	2	0.57	0.92	**1.66**	2.43	
		(0.18)	(0.41)	(0.35)	(0.36)	
	3	0.99	**1.79**	1.82	2.59	
		(0.19)	(0.65)	(0.47)	(0.50)	
	4	**1.79**	2.88	3.10	3.37	
		(0.55)	(0.56)	(0.38)	(0.40)	
	5	3.21	3.16	3.10	2.23	
		(0.29)	(0.37)	(0.46)	(0.88)	

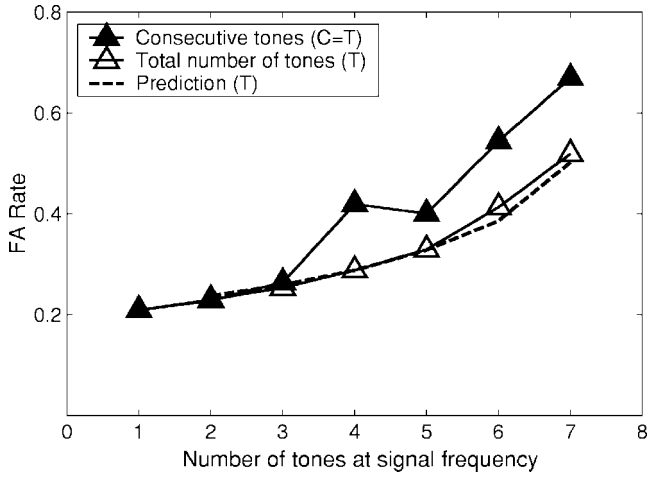


FIG. 5. False alarm (FA) rates are shown as a function of the number of tones at the target frequency. The total numbers of tones are either not assured to be consecutive (unfilled, T) or are consecutive (filled, $C=T$). The dashed line shows the prediction based on the assumption that C consecutive tones drive FA rates when there are a total of T tones at the target frequency.

at the signal frequency. First consider the unfilled symbols. These FA rates were based on the total number of tones at the signal frequency (total energy, T), whether the components were consecutive or not. Next consider the filled symbols. These are the FA rates when the total number of tones was T, and all T of those tones were consecutive (C; and $C=T$). Thus, the filled symbols reflect FA rates based on only a subset of the trials used to generate the open symbols. As an example, consider $T=5$. Across the no-signal trials, when there were a total of five tones at the signal frequency the FA rate was 0.3 (unfilled symbol). For a subset of those trials, there were five consecutive ($C=5$), and the FA rate is higher, 0.4 (filled symbol).

The results of this analysis can be summarized as follows. First, as the number of tones at the signal frequency increased, FA rates grew. This is consistent with an energy model. Moreover, when there were T tones at the signal frequency, and all T tones were consecutive, FA rates were higher than when they were not. This general conclusion holds true for $T>3$. For $T\leqslant 3$, the FA rates did not depend on whether the tones were consecutive. This result indicates that the energy model alone cannot account for the FA rates; whether the tones were consecutive contributed to the listeners' coherence detection decisions.

While the analysis described above provides evidence that consecutive tones at the signal frequency contribute to the production of FAs, the analysis does not address the issue of whether this result holds only for the signal frequency. Do consecutive tones at off-signal frequencies also contribute to the FA rates? This is a question of attention—are listeners successful in attending to only the signal frequency when it is known? In order to address this issue, the analysis described above was repeated for the number of tones at 700 Hz, a frequency that was *not* one of the five fixed signal frequencies. As before, FA rates were obtained using all trials of condition 1 in which the signal frequency was fixed and using the data of all listeners. The results were clear; FA rate was essentially independent of the number of tones present at 700 Hz. As an example, when only signal frequencies were considered, and when all of the tones at the signal frequency were consecutive ($C=T$, filled symbols in Fig. 5), FAs ranged from 0.21 to 0.67 for $T=1$ to $T=7$, respectively. For the 700 Hz off-signal frequency, the range was 0.26 to 0.34. This small range of FA rates indicates that listeners' decisions were based largely on the presence of consecutive tones at the known signal frequency.

For the random-frequency signal, unfortunately, an analogous analysis could not be achieved. This reflects two factors. First, the values of p tested in the random signal condition were small so there were few occasions in which four or more consecutive tones were present at a given frequency. Second, the false alarm rates for small numbers of consecutive tones were not the same for the random- and fixed-signal conditions. As a result, comparisons across conditions are not assured to be meaningful.

The above analysis suggests the presence/absence of consecutive tones contributed to listeners' detection decisions. An additional question is whether energy at the signal frequency contributed to the observed FA rates. That is, did listeners' FA rates depend *only* on the number of consecutive tones at the signal frequency? By this model, energy at the signal frequency would play no role at all except in an epiphenomenal sense—as the number of consecutive tones increased the total number of tones also tended to increase. To address this question, a simple model was used to predict the FA rates as a function of the total number of tones (unfilled symbols in Fig. 5). The model posits that for any number of tones at the signal frequency (T), the FA rate depends only on the length of the largest sequence of consecutive tones.

An evaluation of this model requires a slightly different approach than the analysis described above. The goal is to determine the FA rate, $P(\text{“}S\text{”})$, as a function of T. For a fixed total number of tones, L_S was defined as the length of the longest series of consecutive tones at the signal frequency. The first step was to estimate $P(L_S=C/T)$ for $0<C\leqslant T$. This is the probability that the longest series of consecutive tones is C, given that the total number of tones is T. These probabilities can be derived using enumeration of eight Bernoulli trials (there are eight bursts). A second probability used was $P(\text{“}S\text{”}/L_S=C)$, the FA rate obtained across all trials when the maximum number of consecutive tones was equal to C. These values are not shown in Fig. 5, but the function relating $L_S=C$ and FAs is similar to the function relating $C=T$ and FA (filled symbols). Taking these two key probabilities, the probability of a FA for a given value of T may be computed: $P(\text{“}S\text{”})=\Sigma_{C=1}^{T}P(L_S=C/T)\times P(\text{“}S\text{”}/L_S=C)$. The predictions, shown using dashed lines in Fig. 5, are quite close to the empirical results, suggesting that the FA rate can be modeled reasonably well by assuming that the FA rates depend only on the maximal number of coherent tones at the signal frequency

Overall, the three different analyses of FA rates described above *suggest* three conclusions: (a) the detection of a coherence does not rely solely on the total energy at the signal frequency; (b) when the signal frequency is known,

the number of consecutive tones at off-signal frequencies do not appear to contribute to the generation of FAs; and (c) the patterns of FAs as a function of T are well predicted by a model in which the listeners' signal detection decision depends on the longest series of consecutive tones at the known signal frequency. Unfortunately, there were not sufficient data to extend this *post-hoc* analysis to data collected using the random-frequency signal type, nor to data collected in conditions 2–5.

IV. SUMMARY AND CONCLUSIONS

The detection of a coherent signal embedded in random multicomponent bursts was impaired when the signal frequency was random compared to fixed. When the signal frequency was fixed, listeners were most sensitive to the mid-frequency signal (1000 Hz), suggesting the detection mechanism may depend on contrasts across frequencies as well as time. Analyses of FA rates indicated that signal detection depended, at least in part, on the presence of *coherent* tones at the signal frequency, not just increased *energy* at the signal frequency. Delaying the signal relative to the masker by as little as 1 ms led to enhanced signal detectability, although larger delays were required for the lowest signal frequency tested (220 Hz). Finally, the relation between signal-masker *delay* and p was orderly—increasing the former, or, reducing the latter, balanced to generate equal detectability.

ACKNOWLEDGMENTS

This work was supported by Grant No. RO1 DC02012 from the National Institutes of Health. We thank Zhongzhou Tang for his very helpful assistance on many aspects of this research. We also thank Dr. Gerald Kidd, Jr., and two anonymous reviewers for their insightful suggestions on an earlier draft of this manuscript.

[1]Across all signal types, the average value of p was 0.4 (Table I). Given 29 potential frequencies, one would expect 12 tones per burst. For the time dimension, any one frequency would be expected to contribute energy to three of the eight bursts.

Bernstein, L. R., and Green, D. M. (**1987**). "Detection of simple and complex changes of spectral shape," J. Acoust. Soc. Am. **82**, 1587–1592.

Bregman, A. S., and Pinker, S. (**1978**). "Auditory streaming and the building of timbre," Can. J. Psychol. **32**, 19–31.

Gockel, H. (**2000**). "Perceptual grouping and pitch perception," in *Results of the 8th Oldenburg Symposium on Psychological Acoustics*, edited by A. Schick, M. Mois, and C. Reckhardt (BIS, Oldemberg), pp. 275–292.

Green, D. M., and Berg, B. G. (**1991**). "Spectral weights and the profile bowl," Q. J. Exp. Psychol. **43A**, 449–458.

Green, D. M., and Mason, C. R. (**1985**). "Auditory profile analysis: Frequency, phase, and Weber's law," J. Acoust. Soc. Am. **77**, 1155–1161.

Green, D. M., Onsan, Z. A., and Forrest, T. G. (**1987**). "Frequency effects in profile analysis and detecting complex spectral changes," J. Acoust. Soc. Am. **81**, 692–699.

Kidd, G., Mason, C. R., and Dai, H. (**1995**). "Discriminating coherence in spectro-temporal patterns," J. Acoust. Soc. Am. **97**, 3782–3790.

Kidd, G., Mason, C., and Richards, V. M. (**2003**). "Multiple bursts, multiple looks, and stream coherence in the release from informational masking." J. Acoust. Soc. Am. **114**, 2835–2845.

Kidd, G., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing*, 5th ed. (Academic, San Diego).

Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acta Acustica-Acustica 88, 320–333.

Richards, V. M., and Tang, Z. (2005). "Estimates of effective frequency selectivity based on the detection of a tone added to complex maskers," J. Acoust. Soc. Am. (in press).

Zera, J., and Green, D. M. (**1993**). "Detecting temporal onset and offset asynchrony in multi-component complexes," J. Acoust. Soc. Am. **93**, 1038–1052.

Temporal processing deficits in the pre-senescent auditory system

John H. Grose,[a] Joseph W. Hall III, and Emily Buss
Department of Otolaryngology/Head and Neck Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7070

(Received 22 September 2005; revised 14 December 2005; accepted 13 January 2006)

This study tested the hypothesis that temporal processing deficits are evident in the pre-senescent (middle-aged) auditory system for listening tasks that involve brief stimuli, across-frequency-channel processing, and/or significant processing loads. A gap duration discrimination (GDD) task was employed that used either fixed-duration gap markers (experiment 1) or random-duration markers (experiment 2). Independent variables included standard gap duration (0, 35, and 250 ms), marker frequency (within- and across-frequency), and task complexity. A total of 18 young and 23 middle-aged listeners with normal hearing participated in the GDD experiments. Middle age was defined operationally as 40–55 years of age. The results indicated that middle-aged listeners performed more poorly than the young listeners in general, and that this deficit was sometimes, but not always, exacerbated by increases in task complexity. A third experiment employed a categorical perception task that measured the gap duration associated with a perceptual boundary. The results from 12 young and 12 middle-aged listeners with normal hearing indicated that the categorical boundary was associated with shorter gaps in the young listeners. The results of these experiments indicate that temporal processing deficits can be observed relatively early in the aging process, and are evident in middle age. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2172169]

PACS number(s): 43.66.Mk, 43.66.Sr [AJO] Pages: 2305–2315

I. INTRODUCTION

Temporal processing declines with advanced age, and independently of hearing loss. In human studies, deficits of temporal processing in the senescent auditory system have been measured using gap detection and duration discrimination paradigms (Abel *et al.*, 1990; Fitzgibbons and Gordon-Salant, 1994; Schneider *et al.*, 1994; Fitzgibbons and Gordon-Salant, 1995; Snell, 1997; Schneider *et al.*, 1998; Strouse *et al.*, 1998; He *et al.*, 1999; Schneider and Hamstra, 1999; Snell and Frisina, 2000; Lister and Tarver, 2004). Most of these studies have dealt with within-frequency-channel temporal processing where the acoustic markers of the gap, or the stimuli of varying duration, are spectrally similar. Two key findings can be highlighted from these studies of elderly listeners. First, the effects of advanced age become more pronounced as the durations of the markers bounding the gap are reduced (Muchnik *et al.*, 1985; Schneider and Hamstra, 1999). Second, the decline in duration discrimination performance for suprathreshold gaps is particularly striking when the acoustic markers form part of a tone sequence (Fitzgibbons and Gordon-Salant, 1995; Gordon-Salant and Fitzgibbons, 1999; Fitzgibbons and Gordon-Salant, 2004). This enhanced effect is attributed to increased task complexity. Temporal processing in the senescent auditory system, therefore, appears to be particularly challenged by brief stimuli presented in complex contexts.

The present study examined the hypothesis that deficits in auditory temporal processing may actually begin in middle age. In the experiments conducted here, middle age is operationally defined as 40–55 years of age. From both an anatomical and physiological point of view, there is reason to expect compromised temporal performance in the middle-aged, or pre-senescent, auditory system. For example, neuropharmacological studies that have included psychoacoustic measures of gap detection and/or discrimination have suggested that acuity in the processing of brief auditory events is associated with the level of dopamine activity in the basal ganglia (Artieda *et al.*, 1992; Rammsayer and Classen, 1997). In parallel, imaging studies have pointed to a decline in the dopamine receptors in this brain region throughout adult life (Kaasinen *et al.*, 2000), with a particular acceleration in this decline beginning at about 36 years of age (Mozley *et al.*, 1999). In terms of electrophysiological evidence for pre-senescent deficits in temporal processing, some differences between young and middle-aged listeners are apparent in the N1-P2 response elicited as a function of stimulus duration (Ostroff *et al.*, 2003). Behavioral evidence on the question of a pre-senescent decline in auditory temporal processing is sparse. Abel *et al.* (1990) measured duration discrimination in several populations, including a group of young normal-hearing listeners (20–35 years) and a group of normal-hearing middle-aged listeners (40–60 years). They found that these two groups differed in their ability to discriminate the duration of brief (20 ms) stimuli. Early aging effects have also been noted in studies of gap detection and discrimination. Snell and Frisina (2000) found that, in a group of normal-hearing listeners aged 17–40 years, there was a significant correlation between age and detection threshold for a gap placed in a modulated low-pass noise.

[a]Electronic mail: jhg@med.unc.edu

Muchnik *et al.* (1985) found that gap detection was poorer in middle-aged than in young listeners, particularly when the duration of the acoustic gap markers was brief (10 ms). Evidence for a pre-senescent age effect in across-frequency-channel temporal processing also exists. Grose *et al.* (2001) found that a limited sample of young listeners had consistently lower gap discrimination thresholds than middle-aged listeners for both within- and across-frequency-channel conditions. Lister *et al.* (2002) also measured the discrimination of gaps bounded by disparate marker frequencies in normal-hearing young and middle-aged listeners in a study that included elderly listeners. Although they highlighted the performance deficits in the elderly listeners, and did not identify a decline in performance in their middle-aged listeners, it is evident from their data that the middle-aged and young listeners diverged in performance for the widest (two-octave) marker frequency separation.[1]

In summary, a substantial body of work has established that temporal processing declines with advanced age, and independently of hearing loss, but there has been little focus on the emergence of these temporal deficits in the pre-senescent, or middle-aged, auditory system. The evidence to date suggests that such deficits may be identified in listening tasks that involve brief stimuli, particularly those that span disparate frequency channels. Furthermore, it is likely that any temporal processing deficits will be heightened when the listening tasks are complex. The purpose of this investigation, therefore, was to test the hypothesis that temporal processing deficits are evident in middle age for listening tasks that involve brief stimuli, across-frequency-channel processing, and/or significant processing loads. This hypothesis was tested in a series of three experiments.

II. EXPERIMENT 1: GAP DURATION DISCRIMINATION FOR FIXED-DURATION MARKERS

The purpose of this experiment was to measure the just-noticeable increment in the duration of a silent interval bounded by two brief acoustic markers in normal-hearing young and middle-aged listeners. The three main stimulus parameters manipulated were marker frequency, standard gap duration, and the acoustic context of the gap markers.

A. Method

1. Listeners

Across the various conditions of this experiment, a total of 18 young and 23 middle-aged listeners participated. The young listeners ranged in age from 18–27 years (mean =21.6 years), and the middle-aged listeners ranged in age from 40–55 years (mean=46.6 years). In addition to these listeners, a limited group of eight elderly listeners, ranging in age from 65–83 years (mean=72.8 years), also participated. Their inclusion served to provide a reference for performance in the senescent auditory system. All young and middle-aged listeners had normal hearing, with audiometric thresholds $\leqslant$20 dB HL across the octave frequencies 250–8000 Hz (ANSI, 1996). The elderly listeners had audiometric thresholds $\leqslant$20 dB HL across the octave frequencies 250–4000 Hz, except for one listener whose threshold at 4000 Hz was 40 dB HL. The listeners were tested in two subgroups defined by the conditions to which they listened. Subgroup 1 consisted of 11 young, 10 middle-aged, and all eight elderly listeners. Subgroup 2 consisted of 11 young and 15 middle-aged listeners. It should be noted that inclusion into Subgroup 2 required that the listener pass a screening test for temporal pattern discrimination, as described in more detail below. One potential listener who was middle aged did not pass the screening test and was excluded from participation. A total of four young and two middle-aged listeners were common to both subgroups (i.e., they listened to all conditions).

2. Stimuli

The acoustic markers of the silent gaps were tonebursts consisting of 10 ms cosine-squared rise/fall ramps with no plateau, giving a total duration of 20 ms. The frequency of the leading marker was always 432 Hz; the frequency of the trailing marker was either 458 Hz or 2188 Hz. Relative to 432 Hz, these latter frequencies correspond to spectral separations of one semitone and 13 equivalent rectangular bandwidths [(ERBs), Moore and Glasberg, 1987], respectively. This permitted the assessment of both within-frequency-channel (432–458 Hz) and across-frequency-channel (432–2188 Hz) temporal processing. The duration of the gap between the two acoustic markers was defined as the interval between the 0-voltage points of the leading and trailing marker ramps. Stimuli were generated at a sampling rate of 10 kHz (TDT AP2), output through a digital-to-analog converter (TDT PD1), anti-alias filtered at 4 kHz (Kemo VBF8), and presented monaurally through a Sennheiser 518 headphone. The presentation level was 80 dB peSPL.

3. Procedure

The listener's task was to detect an increment in the duration of the gap between the leading and trailing markers. Because the frequencies of the leading and trailing markers were never identical, and because the gap was defined as the interval between the 0-voltage points on successive 10-ms fall-rise ramps, the two gap markers were never perceived as a continuous sound even for a gap duration of 0 ms. The task was therefore one of gap duration discrimination, hereafter referred to as the GDD task. A 3-interval, forced-choice (3AFC) procedure was used to measure GDD that incorporated a 3-down, 1-up stepping rule to converge on the 79.4% correct point on the psychometric function.

Gap duration was adjusted by a factor of 1.2 at each reversal point. Each threshold track continued for a total of 10 reversals, and the threshold estimate for that track was computed as the geometric mean of the duration increments over the final six reversals. In order to exclude threshold estimates based on tracks that included spuriously large excursions, a statistic was employed which computed the ratio of the gap increment duration one standard deviation above the mean to the mean gap increment duration itself. If this ratio exceeded 1.35, the threshold run was rejected and a replacement run undertaken. At least three valid threshold estimates, but usually four, were collected from a listener for

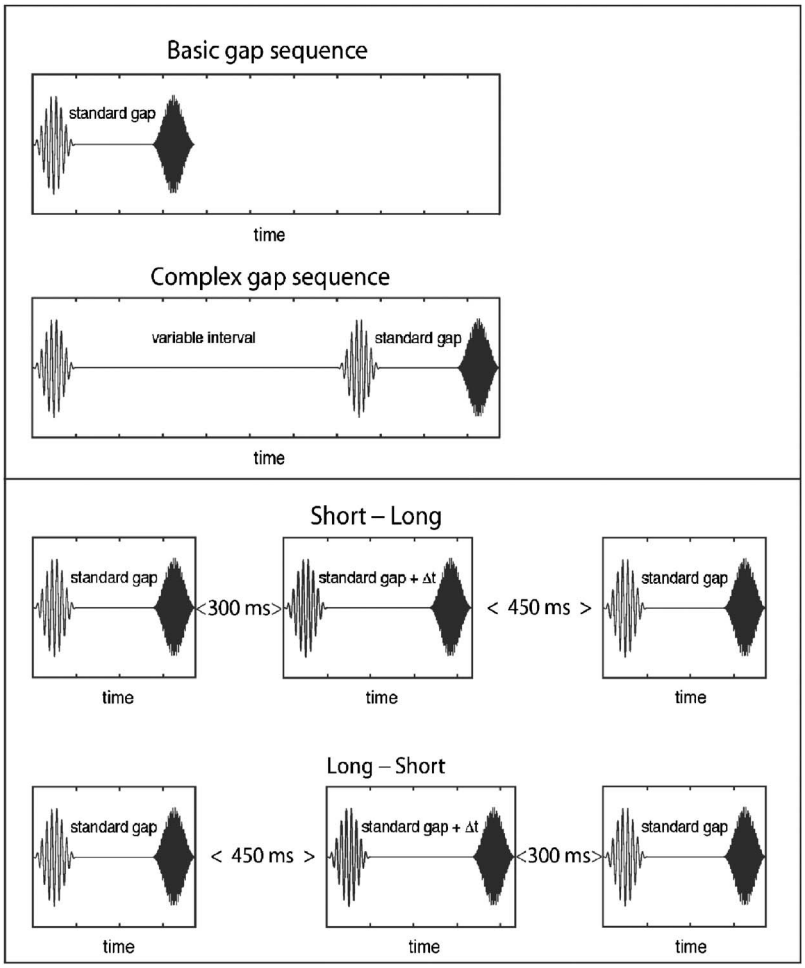


FIG. 1. Stimulus schematics of complex conditions. Upper panel: The top trace shows the basic condition where a pair of tonebursts mark the standard gap; the lower trace shows the complex condition for Subgroup 1 where the gap marker pair is preceded at a variable interval by a preliminary toneburst to form a 3-toneburst sequence. Lower panel: Illustrative 3AFC trials for Subgroup 2 where the pauses between observation intervals provide either a short-long rhythm (top trace) or long-short rhythm (lower trace). In both illustrated trials, the signal interval for the GDD task is the second one.

any particular condition, and the final threshold for that condition and listener was taken as the geometric mean of all estimates collected.

All listeners were well-practiced before data collection commenced. For Subgroup 1, training consisted of repeating the entire experiment twice—once for practice and once for data collection. Since the experiment for this subgroup consisted of eight conditions (see below), all eight conditions were cycled through once (including the multiple repetitions for each condition) and then repeated for data collection. Of the 20 listeners who were unique to Subgroup 2, six underwent the same training regimen wherein the entire experiment was repeated once for practice and once for data collection. The remaining listeners underwent a different approach wherein each condition administered to a listener was replicated until performance appeared to reach an asymptotic level; at this point, data collection for that condition was initiated and at least four threshold estimates were collected.

In the core conditions administered to both subgroups of listeners, GDD was measured for standard gap durations of 0 and 35 ms. Subgroup 1 was also presented with an additional standard gap duration of 250 ms. The gap durations of 35 ms and 250 ms coincide with those tested in Grose *et al.* (2001). For each standard gap duration tested, GDD was measured for both within-frequency-channel (WFC) markers and across-frequency-channel (AFC) markers. These six conditions can therefore be identified with the nomenclature: (1) WFC_0, (2) WFC_35, (3) WFC_250, (4) AFC_0, (5) AFC_35, and (6) AFC_250.

For the standard gap duration of 35 ms, additional conditions were administered which were designed to increase task complexity (and consequentially the processing load placed on the listener). For Subgroup 1, the added complexity consisted of incorporating the pair of gap marker tonebursts into a train of three tonebursts, as shown in the upper panel of Fig. 1. Here, the 432-Hz leading marker of the gap was itself preceded by a 20-ms, 432-Hz toneburst. The interval between the offset of this preliminary toneburst and the onset of the leading gap marker had a nominal duration of 70 ms, but the actual duration varied randomly from 40–100 ms (70 ms±30 ms, with random draws from a rectangular distribution) on each and every presentation. The reason for the random variation was to prevent the listener from using a cue based on the rhythm of the three-tone sequence. The listener now had to discriminate the duration of the gap between the second and third tonebursts of a three-toneburst train. This manipulation of adding a preliminary 432-Hz toneburst was applied to both the within-frequency-channel and across-frequency-channel conditions. Accordingly, these two conditions are identified with the nomenclature WFC_35_CPX1 and AFC_35_CPX1, respectively.

For Subgroup 2, the added complexity for the standard gap duration of 35 ms consisted of requiring the listener to

TABLE I. Conditions and associated listener groups for experiment 1.

Condition	Listeners	Marker frequencies	Standard gap duration	Complexity
WFC_0	Subgroups 1 and 2	432–458 Hz	0 ms	
WFC_35	Subgroups 1 and 2	432–458 Hz	35 ms	
WFC_250	Subgroup 1	432–458 Hz	250 ms	
AFC_0	Subgroups 1 and 2	432–2188 Hz	0 ms	
AFC_35	Subgroups 1 and 2	432–2188 Hz	35 ms	
AFC_250	Subgroup 1	432–2188 Hz	250 ms	
WFC_35_CPX1	Subgroup 1	432–458 Hz	35 ms	+ Prelim. tone
AFC_35_CPX1	Subgroup 1	432–2188 Hz	35 ms	+ Prelim. tone
WFC_35_CPX2	Subgroup 2	432–458 Hz	35 ms	+ Second judgment
AFC_35_CPX2	Subgroup 2	432–2188 Hz	35 ms	+ Second judgment

make two concurrent temporal judgments. The first temporal judgment was the basic GDD decision. The second temporal judgment was the sequential pattern of the three observation intervals in a 3AFC trial. Here, the nominal pauses between the first and second observation intervals and between the second and third observation intervals were either 300 ms and 450 ms, respectively, or 450 ms and 300 ms, respectively, at random. This gave the observation intervals of the 3AFC trial the rhythmic temporal pattern of short-long or long-short, respectively. The lower panel of Fig. 1 depicts a stimulus schematic for this manipulation. Following each 3AFC trial, the listener therefore had to make the basic GDD decision (which interval contained the longer gap) as well as the temporal pattern judgment (short-long or long-short). This manipulation of requiring two concurrent temporal judgments was applied to both the within-frequency-channel and across-frequency-channel conditions. Accordingly, these two conditions are identified with the nomenclature WFC_35_CPX2 and AFC_35_CPX2, respectively.

All responses were entered via button presses on a handheld response box, and feedback was provided after each trial via LED indicators on the box. For Subgroup 2, a rudimentary measure of response time was also included in this procedure wherein the time between the end of stimulus presentation at the conclusion of a 3AFC trial and the moment of button press when selecting the interval with the longest gap was recorded. However, the listener was not aware that response time was being monitored.

Because the manipulation of task complexity for Subgroup 2 required the listener to discriminate a short-long vs long-short temporal pattern, it was necessary to first determine that the listener could indeed make this judgment in isolation, apart from the GDD task. Prior to being enrolled in the experiment, therefore, prospective listeners were administered a simple screening test in which they had to judge the temporal patterns of the pauses between 3AFC observation intervals (300 ms–450 ms or 450 ms–300 ms) where the intervals themselves contained clearly audible 150-ms, 432 -Hz pure tones. Performance at a level of 90% correct or better was required to participate in the experiment.

In summary, this experiment measured GDD as a function of listener age, where the stimulus parameters of marker frequency, standard gap duration, and the acoustic context of the gap markers were systematically manipulated. Table I summarizes the various conditions tested.

B. Results and discussion

The core conditions of experiment 1 measured GDD for standard gap durations of 0 ms and 35 ms for both within-frequency-channel markers and across-frequency-channel markers (conditions WFC_0, WFC_35, AFC_0, and AFC_35). Subgroup 1 listened to an additional standard gap duration of 250 ms (conditions WFC_250 and AFC_250). The geometric means for these conditions are shown in Fig. 2 for the three age groups (young, circles; middle aged, squares; elderly, triangles) for both the within-frequency-channel (filled symbols) and across-frequency-channel (open symbols) conditions.[2] The primary focus of this experiment was whether the young and middle-aged listeners differed in the core conditions of 0 and 35 ms standard gap durations. Accordingly, these data, combined across both subgroups, were submitted to a repeated-measures analysis of variance (ANOVA) with one between-subjects factor (age: young, middle aged) and two within-subjects factors (standard gap duration: 0 ms, 35 ms; frequency channel: within, across). All statistical analyses were performed on logarithmic transforms of the data in order to maintain homogeneity of variance across conditions. The analysis indicated significant main effects of age ($F_{1,39}=6.863$; $p=0.012$), standard gap duration ($F_{1,39}=91.59$; $p<0.001$), and frequency channel

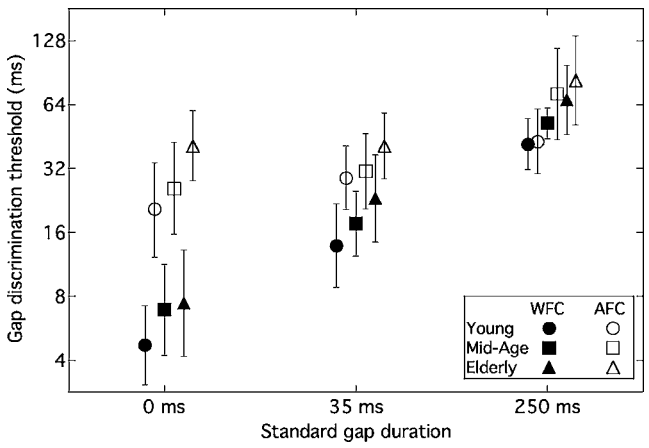


FIG. 2. Group mean GDD thresholds plotted as a function of standard gap duration for three age groups (young, circles; middle-aged, squares; elderly, triangles). Filled symbols indicate within-frequency-channel conditions; unfilled symbols indicate across-frequency-channel conditions. Error bars indicate ±1 standard deviation.

($F_{1.39}$=263.306; $p<0.001$). The only significant interaction was between standard gap duration and frequency channel ($F_{1.39}$=44.367; $p<0.001$). Means comparisons using Bonferroni correction indicated that this interaction was due to the difference between within- and across-frequency-channel conditions being greater for the standard 0 ms gap than the 35 ms gap, although both differences were significant (t_{40} =14.328 and 9.427, respectively; $p<0.01$). Overall, this pattern of results indicates that the middle-aged listeners performed more poorly than the young listeners across the four core conditions.

Two subsidiary questions were anticipated in association with these core conditions: (1) Does the age effect extend to longer standard gap durations?; and (2) How do young and middle-aged listeners perform relative to elderly listeners? To address these questions, Subgroup 1 was also tested using a standard gap duration of 250 ms; in addition, this subgroup included a cohort of eight elderly listeners. A repeated-measures ANOVA on the data from Subgroup 1 alone was therefore undertaken that included one between- subjects factor (age: young, middle aged, elderly) and two within-subjects factors (standard gap duration: 0 ms, 35 ms, 250 ms; frequency channel: within, across). The results indicated significant main effects of age ($F_{2.26}$=11.618; $p<0.001$), standard gap duration ($F_{2.52}$=184.031; $p<0.001$), and frequency channel ($F_{1.26}$=135.968; $p<0.001$). Again, the only significant interaction was between standard gap duration and frequency channel ($F_{2.52}$=58.485; $p<0.001$). Post-hoc analysis using the Tukey HSD test indicated that the middle-aged and elderly listeners had significantly higher GDD thresholds than the young listeners, but that the two older age groups did not themselves differ significantly. Simple contrasts showed that, for all age groups, performance declined significantly as standard gap duration increased both for within-frequency-channel and across-frequency-channel conditions. In summary, this analysis showed that the age effect extended to a standard gap duration of 250 ms; in addition, it showed that middle-aged listeners performed more like elderly listeners than like young listeners on this GDD task. Based on the work of Lister *et al.* (2002), this latter result was not expected. In that study, the elderly listeners performed substantially worse on a GDD task than either the young or middle-aged listeners. The reason for this discrepancy is not clear, although it is possible that the criterion of relatively normal hearing (≤20 dB HL across the octave frequencies 250–4000 Hz) in the small group of elderly listeners tested here may have resulted in the selection of a cohort with particularly resilient auditory function.

It was hypothesized that potential age effects would be exacerbated by increasing the complexity of the task. For Subgroup 1, task complexity was increased by embedding the 35-ms gap markers into a 3-toneburst train. For Subgroup 2, task complexity was increased by requiring listeners to make two concurrent temporal judgments. Dealing first with Subgroup 1, Fig. 3 shows the group mean GDD thresholds for the 35-ms standard gap presented both in isolation and in the complex context. It is evident that for all age groups and conditions, performance declined when task complexity was

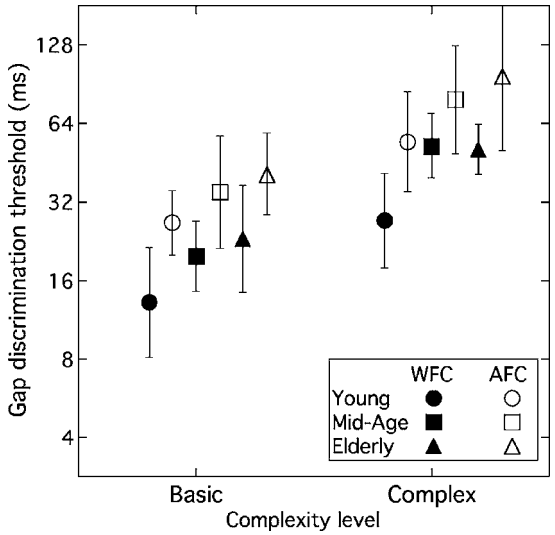


FIG. 3. Group mean GDD thresholds for a standard gap duration of 35 ms presented in isolation (basic) and as part of a 3-toneburst sequence (complex) for three age groups (young, circles; middle-aged, squares; elderly, triangles). Data are shown for both within-frequency-channel (filled symbols) and across-frequency-channel (unfilled symbols) conditions. Error bars indicate ±1 standard deviation.

increased. This was confirmed using a repeated-measures ANOVA that included one between-subjects factor (age: young, middle aged, elderly) and two within-subjects factors (complexity: absent, present; frequency channel: within, across). The analysis indicated significant main effects of age ($F_{2.26}$=10.999; $p<0.001$), complexity ($F_{1.26}$=158.962, $p<0.001$), and frequency channel ($F_{1.26}$=48.643; $p<0.001$). None of the interaction terms were significant. Post-hoc analysis using the Tukey HSD indicated that the performance of the middle-aged and elderly listeners was significantly poorer than that of the young listeners but that the two older age groups did not themselves differ significantly.

It could be argued, however, that the pertinent issue is not whether the middle-aged listeners performed more poorly than the young listeners in the complex task, but whether their decline in performance relative to the basic task was proportionally greater than that for young listeners. That is, age differences tend to be exacerbated by increases in task complexity and therefore it was expected that the added complexity would compound the difficulty for the middle-aged listeners. To address this, a ratio was computed for each listener that related the difference in GDD thresholds between the basic and complex conditions to the basic condition threshold itself. This Weber-type ratio was computed for both the within-frequency-channel and across-frequency-channel conditions (i.e., [WFC_35_CPX1−WFC_35]/WFC_35 or [AFC_35_CPX1 −AFC_35]/AFC_35). An ANOVA on the resulting Weber ratios indicated no effect of age ($F_{2.26}$=0.564; p=0.576) or frequency channel ($F_{1.26}$=0.429; p=0.518). This indicates that the decline in GDD performance due to embedding the 35-ms standard gap into a 3-toneburst sequence was proportionally the same for all age groups, and did not depend on within- vs. across-frequency channel conditions. The finding that the relative decline in GDD performance for the embed-

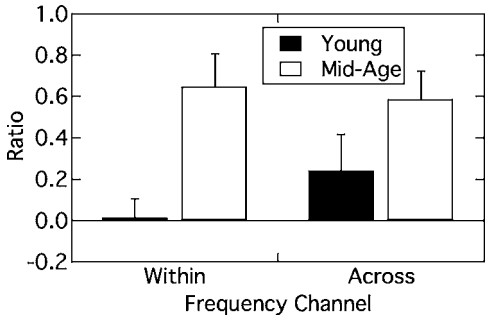


FIG. 4. Group mean Weber ratios encoding the relative change in GDD threshold as a function of task complexity for Subgroup 2. Ratios are plotted for young (filled bars) and middle-aged (open bars) listeners for within- and across-frequency channel conditions. Error bars indicate one standard error of the mean.

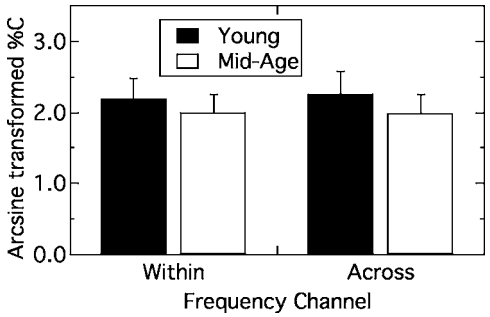


FIG. 5. Group mean arcsine-transformed percent correct scores for interval pattern identification for young (filled bars) and middle-aged (open bars) listeners for within- and across-frequency channel conditions. Error bars indicate one standard deviation.

ded standard gap did not depend on listener age was contrary to expectations. A number of studies comparing young and elderly listeners have shown that temporal processing deficits associated with advanced age are exacerbated by increased stimulus complexity, such as the use of stimulus trains (e.g., Fitzgibbons and Gordon-Salant, 1995).

For Subgroup 2, task complexity was increased by requiring the young and middle-aged listeners to make two concurrent temporal judgments. It was expected that GDD performance would decline under conditions where more than one temporal feature had to be monitored at the same time. Accordingly, GDD thresholds for the 35-ms standard gap in the basic task (WFC_35 or AFC_35) and the GDD thresholds in the respective complex task (WFC_35_CPX2 or AFC_35_CPX2) were compared using Weber ratios as described above. The group mean ratios are shown in Fig. 4. The ratio for one young listener in the within-frequency-channel comparison was excluded on the grounds of it being an outlier: the ratio in question was close to three standard deviations away from the group mean. It is apparent from Fig. 4 that the relative decline in performance with the added task complexity was greater for the middle-aged listeners than for the young listeners. This was confirmed using a repeated-measures ANOVA that included one between-subjects factor (age: young, middle aged) and one within-subjects factor (frequency channel: within, across). The effect of age was significant ($F_{1.23}=8.762$; $p=0.007$) but no other effects or interactions were significant. These results indicate that the middle-aged listeners experienced proportionally greater difficulty with the GDD task than the young listeners when the added complexity of making two concurrent temporal judgments was imposed.

To be included in Subgroup 2, all listeners had to demonstrate high accuracy ($\geqslant$90% correct) in making the interval pattern judgment in isolation. It was of interest to determine whether the young and middle-aged listeners differed in the degree to which this accuracy was compromised when the interval pattern judgment formed one of the two concurrent temporal judgments in the complex listening conditions. Figure 5 shows the average arcsine-transformed percent correct scores for interval pattern identification for the young and middle-aged listeners in both the within- and across-frequency channel conditions. A repeated-measures ANOVA on the arcsine-transformed data with one between-subjects factor (age: young, middle age) and one within-subjects factor (frequency channel: within, across) showed a significant effect of age ($F_{1.24}=4.645$; $p=0.041$), but not frequency channel ($F_{1.24}=0.846$; $p=0.367$). The interaction between these factors was not significant. Thus the younger listeners were not only better than the middle-aged listeners at discriminating increments in gap duration in the complex listening task, they were also more accurate in the concurrent temporal order judgment.

A final metric that was collected for Subgroup 2 was the time-to-respond for the GDD decision following each 3AFC trial. The group mean data are shown in Fig. 6 for the within-frequency-channel (left panel) and across-frequency-channel (right panel) conditions. Each panel depicts two entries for the standard gap duration of 35 ms: one for the basic task where only the GDD decision was required (T1), and one for the complex task where two concurrent temporal judgments were required (T2). It is apparent that response time was

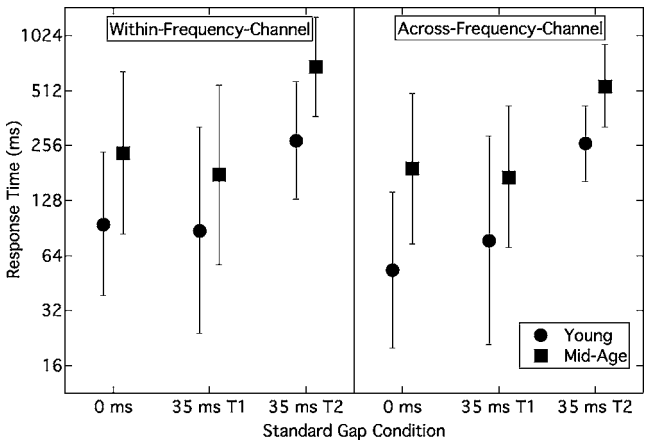


FIG. 6. Group mean response times for within-frequency-channel (left panel) and across-frequency-channel (right panel) conditions. Response times are plotted for young (circles) and middle-aged (squares) listeners for standard gap durations of 0 and 35-ms. The 35 ms gap was tested in the basic condition (T1) and in the complex condition where two concurrent temporal judgments were required (T2). Error bars indicate ±1 standard deviation.

quite variable across listeners, as evidenced by the large standard deviations. Nevertheless, it can be seen that the middle-aged listeners (squares) took consistently longer to respond than the young listeners (circles). This was confirmed with a repeated-measures ANOVA on the logarithmic transforms of the response data, with one between-subjects factor (age: young, middle aged) and two within-subjects factors (gap condition: 0 ms, 35 ms T1, 35 ms T2; frequency channel: within, across). The analysis indicated significant effects of age ($F_{1.24}=8.714$; $p=0.007$), gap condition ($F_{2.48}=38.516$; $p<0.001$), and frequency channel ($F_{1.24}=6.471$; $p=0.018$). None of the interaction terms were significant. Simple contrasts on the data for the different gap conditions indicated that response times did not differ between the basic 0-ms and 35-ms standard gap durations, but were significantly longer for the complex conditions. This pattern of results indicates that the middle-aged listeners responded more slowly in general than the younger listeners. It should be underscored that the listeners were not aware that response time was being measured, and were not instructed to respond as quickly as possible.

In summary, the purpose of experiment 1 was to test the hypothesis that temporal processing deficits are evident in middle age for listening tasks that involve brief stimuli, across-frequency-channel processing, and/or significant processing loads. The results support some aspects of this hypothesis:

(1) Middle-aged listeners performed more poorly than the young listeners in the GDD task. However, this difference was equally apparent for both within-frequency-channel and across-frequency-channel conditions.
(2) Increased task complexity compounded the performance differences between young and middle-aged listeners. However, this proportional decline occurred only when the added complexity consisted of making two concurrent temporal judgments and not when it consisted of embedding the gap markers into a 3-toneburst train.
(3) Middle-aged listeners were less accurate at temporal rhythm judgments in complex listening tasks, and were generally slower in responding than young listeners.

One of the limitations of experiment 1 was that marker duration was fixed at 20 ms. Because of this, overall stimulus duration (two markers+gap) covaried with gap duration. Thus, it is not possible to differentiate between sensitivity to the duration of the silent interval alone and sensitivity to overall stimulus duration. It has long been known that gap detection performance declines markedly when the durations of the gap markers are random (Penner, 1976). Randomizing gap marker duration forces the listener to monitor the duration of the silent interval alone and renders the cue of overall stimulus duration unreliable. In order to determine whether the results of experiment 1 held strictly for *gap* discrimination, listeners from Subgroup 2 undertook a supplementary experiment that used random-duration markers.

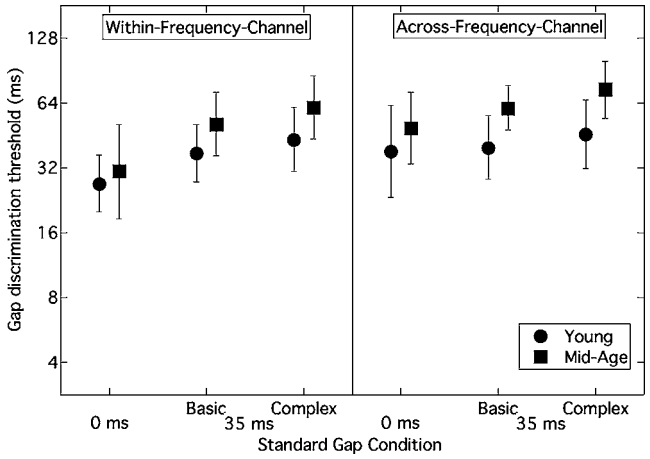


FIG. 7. Group mean GDD thresholds for the six conditions of experiment 2. Within-frequency-channel data is shown in the left panel; across-frequency-channel data in the right panel. The parameter is age groups (young, circles; middle-aged, squares). Error bars indicate ±1 standard deviation.

III. EXPERIMENT 2: GAP DURATION DISCRIMINATION FOR RANDOM-DURATION MARKERS

A. Method

1. Listeners

Ten young and 14 middle-aged listeners from Subgroup 2 participated. The young listeners ranged in age from 18–25 years (mean=21.4 years); the middle-aged listeners ranged in age from 40–54 years (mean=45.7 years).

2. Stimuli

The gap markers were tonebursts with nominal durations of 40 ms, but whose actual duration varied by 50% (40±20 ms, with random draws from a rectangular distribution) on a presentation-by-presentation basis. All other parameters, and the method of stimulus generation, were the same as in experiment 1.

3. Procedure

The same six conditions that Subgroup 2 underwent in experiment 1 were repeated here, except that marker duration was variable rather than constant. These six conditions included the four basic GDD conditions of WFC_0, WFC_35, AFC_0, AFC_35, and the two complex conditions of WFC_35_CPX2 and AFC_35_CPX2 where concurrent temporal judgments of GDD and the temporal order of the observation intervals were required. All other procedural details were the same as described for experiment 1.

B. Results and discussion

The group mean data for the six conditions are displayed in Fig. 7. As expected, performance became poorer and more variable in general, relative to the fixed marker durations of experiment 1 (cf. Fig. 2). A repeated-measures ANOVA on the six conditions was undertaken that included one between-subjects factor (age: young, middle aged) and two within-subjects factors (gap condition: 0 ms, 35- ms basic, 35-ms complex; frequency channel: within, across). As with

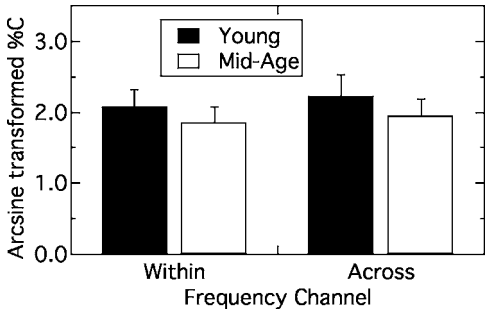


FIG. 8. Group mean arcsine-transformed percent correct scores for young (filled bars) and middle-aged (open bars) listeners for within- and across-frequency channel conditions for experiment 2. Error bars indicate one standard deviation.

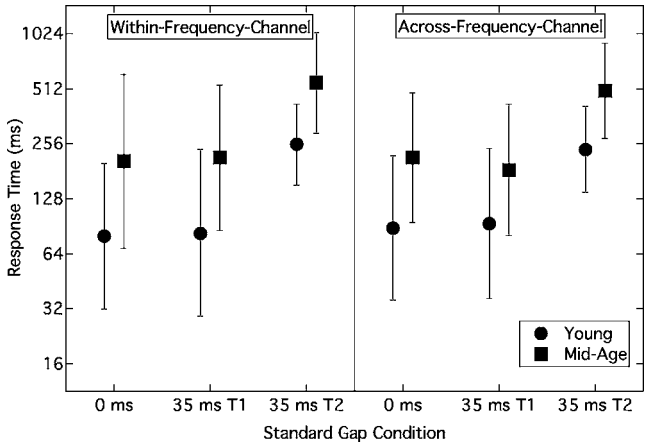


FIG. 9. Group mean response times for within-frequency-channel (left panel) and across-frequency-channel (right panel) conditions for experiment 2. Response times are plotted for young (circles) and middle-aged (squares) listeners for standard gap durations of 0 and 35 ms. The 35-ms gap was tested in the basic conditions (T1) and in the complex condition where two concurrent temporal judgments were required (T2). Error bars indicate ±1 standard deviation.

experiment 1, statistical analysis was performed on the logarithmic transforms of the data to maximize homogeneity of variance.

The analysis indicated significant main effects of age ($F_{1,22}=8.386$; $p=0.008$), gap condition ($F_{2,44}=28.15$; $p<0.001$), and frequency channel ($F_{1,22}=15.211$; $p=0.001$). The only significant interaction was between gap condition and frequency channel ($F_{2,44}=6.113$; $p=0.005$). Means comparisons using Bonferroni correction indicated that this interaction was due to the difference between within- and across-frequency-channel conditions being greater for the standard 0-ms gap than for either the basic or complex 35-ms gap. The difference for the 0-ms gap was significant ($t_{23}=4.612$; $p<0.001$) whereas that for the two 35-ms gaps failed to reach significance ($t_{23}=1.894$, $p=0.071$; $t_{23}=1.758$, $p=0.092$, for the basic and complex gaps, respectively). In terms of the four core conditions (WFC_0, WFC_35, AFC_0, AFC_35), this pattern of results is similar to that observed in experiment 1, and indicates that the middle-aged listeners performed more poorly than the young listeners across the four core conditions.

The repeated-measures ANOVA also addressed the effect of task complexity. The significant main effect of age noted above, with no interaction between age and gap condition, confirms that the middle-aged listeners performed more poorly than the young listeners on both the basic and complex 35-ms gap tasks. Means comparisons using Bonferroni correction indicated that all listeners exhibited elevated thresholds in the more challenging complex conditions, both for the within-frequency-channel conditions ($t_{23}=2.957$; $p=0.007$) and the across-frequency-channel conditions ($t_{23}=3.763$; $p=0.001$). There is no indication in Fig. 7, however, that the poorer GDD performance of the middle-aged listeners was compounded in the complex conditions. This was confirmed by a repeated-measures ANOVA on the Weber ratios that showed no significant effects of age group or frequency channel, indicating that the decline in performance in the complex conditions was equivalent for both age groups, irrespective of frequency channel.

Figure 8 summarizes the accuracy with which the temporal order judgment (short-long vs long-short) was made in the complex listening task in terms of the arcsine-transformed percent correct data. A repeated-measures ANOVA on the data showed significant effects of age ($F_{1,22}=6.155$; $p=0.021$) and frequency channel ($F_{1,22}=18.461$; $p<0.001$). The interaction between these factors was not significant. These results indicate that the young listeners were more accurate in their judgment of temporal order, in addition to being more sensitive to increments in gap duration. Also, irrespective of age, listeners were more accurate in temporal pattern identification in the across-frequency-channel condition, although no explanation is offered for this finding. Finally, the response time data are shown in Fig. 9. A repeated-measures ANOVA on the log-transformed response times showed significant effects of age ($F_{1,22}=6.754$; $p=0.016$) and gap condition ($F_{2,44}=37.042$; $p<0.001$), but not of frequency channel ($F_{1,21}=0.117$; $p=0.735$). None of the interaction terms were significant. Simple contrasts on the gap condition results indicated that response times did not differ between the 0-ms and 35-ms basic conditions, but were significantly longer for the complex conditions. The pattern of results for the response time data therefore indicates that the middle-aged listeners generally took longer to respond than the young listeners, but that all listeners took longer to respond when task complexity was increased.

In summary, the results of experiment 2 are essentially the same as those of experiment 1. Even though the use of random-duration markers made the task generally more challenging for all listeners, middle-aged listeners were still less sensitive to increments in gap duration than young listeners. In the complex task, where two concurrent temporal judgments were required, middle-aged listeners were both poorer at discriminating gap increments and less accurate in their temporal order judgments. Finally, middle-aged listeners were generally slower at responding than are young listeners. Overall, these results support the hypothesis that temporal processing deficits are evident in middle age. However, not all aspects of the hypothesis as originally proposed are supported. For example, age effects are not necessarily more pronounced for across-frequency than within-frequency con-

figurations, and performance deficits for middle-aged listeners are not necessarily exacerbated relative to young listeners in the more complex listening conditions.

Although the results of experiments 1 and 2 indicate that deficits in temporal processing are evident relatively early in the aging process, it is not clear that these deficits necessarily translate to actual performance differences in real-world listening tasks. Accordingly, a final experiment was undertaken that compared young and middle-aged listeners in a categorical perception task that depended on silent gap duration.

IV. EXPERIMENT 3: /s/-/st/ PERCEPTION AS A FUNCTION OF GAP DURATION

A number of studies have pointed out that silent intervals, or segments of low energy, in speech constitute an information-bearing feature of the acoustic wave form. For example, these gaps can provide cues for voice-onset time (Strouse *et al.*, 1998; Phillips, 1999), as well as the presence of stop constants that abut voiceless sibilants (Bailey and Summerfield, 1980; Best *et al.*, 1981; Dorman *et al.*, 1985; Nittrouer *et al.*, 1998). In light of this, there has been an interest in determining whether psychoacoustic measures of temporal resolution, such as gap detection, relate to speech performance measures for stimuli that involve silent intervals (Formby *et al.*, 1993; Nelson *et al.*, 1995; Lister and Tarver, 2004). One particular study of relevance here measured gap detection for noise-band markers that displayed some speechlike characteristics, and related this measure to the duration of the silent interval between the sibilant /s/ and the ensuing vowel /ei/ that forms the categorical boundary between the words say and stay (Nelson *et al.*, 1995). Although that study did not find a dependency of categorical perception on gap *detection* ability, it did demonstrate that both the gap detection task and the categorical perception task were able to differentiate among listener groups (in this case, categories of hearing loss). In light of this, the present experiment was undertaken to determine whether a difference between young and middle-aged listeners could be identified in a similar categorical perception task.

A. Method

1. Listeners

Twenty-four normal-hearing listeners participated: 12 young and 12 middle aged. The young listeners ranged in age from 19–27 years (mean=20.9 years) and the middle-aged listeners ranged from 41–55 years (mean=47.8 years). Three of the young and six of the middle-aged listeners also participated in either experiments 1 or 2.

2. Stimuli

The stimuli consisted of five sibilant-vowel words recorded by a male speaker. The five words were say, sigh, sow, see, and sue. Four tokens of each word were digitized at a sampling rate of 11.025 kHz (TDT PD1) after low-pass filtering at 4 kHz (Kemo VBF-8). Using a wave form editing program (Sound Edit™), the sibilant segment of each digitized token was excised and stored in one sound file, while the remaining vowel segment was stored in a separate sound file. For presentation of a particular word, a sibilant-segment sound file and a vowel-segment sound file were randomly selected from this library of eight sound files associated with the four tokens of that word, and digitally spliced together to form a single wave form. Silent gaps were introduced between the sibilant and vowel segments by introducing 0-voltage points between the two wave form segments. The imposed gap varied in duration from 5 ms to 100 ms in 5–ms steps. The stimuli were presented to the listeners through one earphone of a Sennheiser HD580 headset at a comfortable listening level.

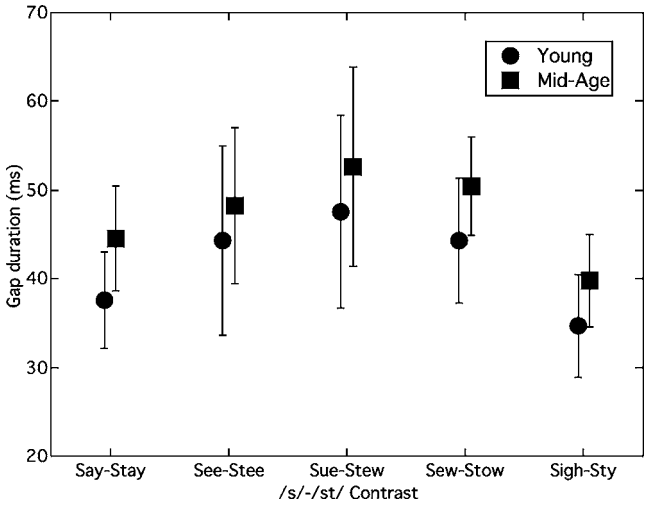


FIG. 10. Group mean gap durations at the /s/-/st/ categorical boundary for young (circles) and middle-aged (squares) listeners. Error bars indicate ±1 standard deviation.

3. Procedure

Each listener was presented with 525 stimuli in random order (5words × 21 gap durations × 5 replications). The task was to indicate via a response box whether the perceived word was of the form sibilant-vowel (i.e., say, sigh, sow, see, or sue) or of the form sibilant-stop-consonant-vowel (i.e., stay, sty, stow, stee, or stew). For each listener and word, the percent /st/ perception (sibilant-stop-consonant-vowel) was plotted as a function of gap duration. Each individual categorical function was then fit with a logit function, and the gap duration associated with the 50% point derived.

B. Results and discussion

The group mean data are displayed in Fig. 10 which shows the gap duration at the categorical boundary between /s/ and /st/ for each of the five contrasts tested. There is scant published data against which to compare these results. However, it is of interest that the gap duration associated with the categorical boundary between synthesized say-stay tokens for normal listeners is about 37 ms (Nelson *et al.*, 1995; high F1 stimulus). For the young normal-hearing listeners and natural say-stay stimuli used here, the boundary is about 38 ms. With reference to Fig. 10, it is apparent that the middle-aged listeners tended to require longer silent intervals between the sibilant segment and the ensuing vowel segment to perceive the word as containing a stop consonant. This

observation was confirmed using a repeated-measures ANOVA with one between-subjects factor (age: young, middle aged) and one within-subjects factor (five words). The analysis indicated a significant effect of age ($F_{1,22}=5.174$; $p=0.033$) and a significant effect of word ($F_{4,22}=17.094$; $p<0.001$), but no interaction between these two factors. This analysis confirms that the gap durations associated with the categorical boundaries for the /s/-/st/ contrasts tested here were longer for the middle-aged listeners than the young listeners. Interestingly, the mean gap durations at the categorical boundary measured here are longer than those found for discrimination thresholds for similar sibilant-vowel stimuli measured by Lister and Tarver (2004). This suggests that the silent interval at the categorical boundary was suprathreshold for all listeners but that the weighting applied to the gap duration in assigning the categorical boundary differed. In conclusion, the finding that an age effect is present in a task related more closely to speech perception suggests that the deficits in temporal processing measured in the GDD tasks of experiments 1 and 2 may have meaningful consequences in everyday listening environments.

V. SUMMARY AND CONCLUSION

The purpose of this study was to test the hypothesis that temporal processing deficits are evident in the pre-senescent (middle-aged) auditory system for listening tasks that involve brief stimuli, across-frequency-channel processing, and/or significant processing loads. Several aspects of this hypothesis were supported. Experiments 1 and 2 showed that GDD thresholds for middle-aged listeners were consistently higher than for young listeners at standard gap durations of 0 and 35 ms. This effect held both for fixed- and random-duration markers, as well as basic and complex temporal tasks. In addition, experiment 1 showed that, for complex listening conditions where two concurrent temporal judgments are required, the added processing load compounded the performance deficits of the middle-aged listeners relative to the young listeners. In contrast, other aspects of the hypothesis were not supported. The results of both experiments 1 and 2 indicated that the temporal processing deficits associated with the middle-aged listeners were not generally restricted to (or necessarily exacerbated by) across-frequency-channel processing. Conditions designed to increase task complexity, and therefore processing load, did not necessarily compound the differences between young and middle-aged listeners. For example, embedding the gap markers in a tonal sequence (experiment 1) or requiring concurrent temporal judgments for random-duration markers (experiment 2) did not exacerbate the performance deficits of middle-aged listeners relative to young listeners. In this vein, it is noteworthy that preliminary work from our laboratory that sought to increase GDD task complexity by incorporating a nontemporal task did not show a difference between young and middle-aged listeners (Grose *et al.*, 2004). In that report, task complexity was increased by requiring the listener to judge the frequency direction of the gap marker sequence concurrently with the GDD task. The general finding from experiments 1 and 2, however, is that temporal processing deficits are evident in the middle-aged auditory system under many conditions that employ brief stimuli.

The results of experiment 3 indicate that these temporal processing deficits associated with the pre-senescent auditory system may have meaningful consequences for the processing of more ecologically relevant stimuli. The finding of an age effect for the gap durations associated with categorical boundaries is noteworthy in light of the response time results of experiments 1 and 2. These latter measures indicated that middle-aged listeners had longer response times than young listeners. It could be argued that slower response times are indicative of a general slowing in decision-making mechanisms, and as such reflect higher-level processes such as memory, rather than temporal processing, *per se*. This important issue is considered carefully in Pichora-Fuller (2003). However, the categorical perception result of experiment 3 suggests that the age effects seen in the GDD tasks are indicative of performance deficits at the level of temporal processing. In conclusion, a large body of work has indicated that temporal processing declines with advanced age, and independently of hearing loss. This study extends this characterization to show that deficits in temporal processing are evident in the pre-senescent, or middle-aged, auditory system under certain conditions.

ACKNOWLEDGMENT

This work was supported by NIDCD R01-DC01507.

[1]Submitting the data for only the young and middle-aged listeners from Lister *et al.* (2002) to a repeated-measures analysis of variance reveals no main effect of age group, but a significant interaction between the factors of age group and marker frequency separation ($F_{5.50}=4.92$; $p=0.001$). Bonferroni-corrected paired-comparisons indicate that for the widest marker frequency separation (2 octaves) the young and middle-aged listeners differed significantly ($t_{10}=3.339$; $p=0.008$).

[2]The maximum gap increment allowed by the procedure was 250 ms. For the rare instances where a listener was unable to perform the task reliably below this ceiling, a threshold value of 250 ms was entered. This occurred for three listeners: (1) one elderly listener for condition AFC_250; (2) one elderly listener for condition AFC_35_CPX1; and (3) one middle-aged listener for conditions AFC_250 and AFC_35_CPX1.

Abel, S. M., Krever, E. M., and Alberti, P. W. (**1990**). "Auditory detection, discrimination and speech processing in ageing, noise sensitive and hearing-impaired listeners," Scand. Audiol. **19**, 43–54.

ANSI (**1996**). *American National Standards Specification for Audiometers (ANSI S3-1996)* (American National Standards Institute, New York).

Artieda, J., Pastor, M. A., Lacruz, F., and Obeso, J. A. (**1992**). "Temporal discrimination is abnormal in Parkinson's disease," Brain **115**, 199–210.

Bailey, P. J., and Summerfield, Q. (**1980**). "Information in speech: observations on the perception of [s]-stop clusters," J. Exp. Psychol. Hum. Percept. Perform. **6**, 536–563.

Best, C. T., Morrongiello, B., and Robson, R. (**1981**). "Perceptual equivalence of acoustic cues in speech and nonspeech perception," Percept. Psychophys. **29**, 191–211.

Dorman, M. F., Marton, K., Hannley, M. T., and Lindholm, J. M. (**1985**). "Phonetic identification by elderly normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 664–670.

Fitzgibbons, P. J., and Gordon-Salant, S. (**1994**). "Age effects on measures of auditory duration discrimination," J. Speech Hear. Res. **37**, 662–670.

Fitzgibbons, P. J., and Gordon-Salant, S. (**1995**). "Age effects on duration discrimination with simple and complex stimuli," J. Acoust. Soc. Am. **98**, 3140–3145.

Fitzgibbons, P. J., and Gordon-Salant, S. (**2004**). "Age effects on discrimi-

nation of timing in auditory sequences," J. Acoust. Soc. Am. **116**, 1126–1134.
Formby, C., Barker, C., Abbey, H., and Raney, J. J. (**1993**). "Detection of silent temporal gaps between narrow-band noise markers having second-formantlike properties of voiceless stop/vowel combinations," J. Acoust. Soc. Am. **93**, 1023–1027.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1999**). "Profile of auditory temporal processing in older listeners," J. Speech Lang. Hear. Res. **42**, 300–311.
Grose, J. H., Hall, J. W., III, and Buss, E. (**2001**). "Gap duration discrimination in listeners with cochlear hearing loss: effects of gap and marker duration, frequency separation, and mode of presentation," J. Assoc. Res. Otolaryngol. **2**, 388–398.
Grose, J. H., Hall, J. W., and Buss, E. (**2004**). *Age and Gap Discrimination: Effect of Processing Load* (American Auditory Society, Scottsdale, AZ).
He, N.-J., Horwitz, A. R., Dubno, J. R., and Mills, J. H. (**1999**). "Psychometric functions for gap detection in noise measures from young and aged subjects," J. Acoust. Soc. Am. **106**, 966–978.
Kaasinen, V., Vilkman, H., Hietala, J., Nagren, K., Helenius, H., Olsson, H., Farde, L., and Rinne, J. O. (**2000**). "Age-related dopamine D2/D3 receptor loss in extrastriatal regions of the human brain," Neurobiol. Aging **21**, 683–688.
Lister, J., Besing, J., and Koehnke, J. (**2002**). "Effects of age and frequency disparity on gap discrimination," J. Acoust. Soc. Am. **111**, 2793–2800.
Lister, J., and Tarver, K. (**2004**). "Effect of age on silent gap discrimination in synthetic speech stimuli," J. Speech Lang. Hear. Res. **47**, 257–268.
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.
Mozley, P. D., Acton, P. D., Barraclough, E. D., Plossl, K., Gur, R. C., Alavi, A., Mathur, A., Saffer, J., and Kung, H. F. (**1999**). "Effects of age on dopamine transporters in healthy humans," J. Nucl. Med. **40**, 1812–1817.
Muchnik, C., Hildesheimer, M., Rubinstein, M., Sadeh, M., Shegter, Y., and Shibolet, B. (**1985**). "Minimal time interval in auditory temporal resolution," J. Aud Res. **25**, 239–246.
Nelson, P. B., Nittrouer, S., and Norton, S. J. (**1995**). ""Say-stay" identification and psychoacoustic performance of hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 1830–1838.
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.
Ostroff, J. M., McDonald, K. L., Schneider, B. A., and Alain, C. (**2003**). "Aging and the processing of sound duration in human auditory cortex," Hear. Res. **181**, 1–7.
Penner, M. J. (**1976**). "The effect of marker variability on the discrimination of temporal intervals," Percept. Psychophys. **19**, 466–469.
Phillips, D. P. (**1999**). "Auditory gap detection, perceptual channels, and temporal resolution in speech perception," J. Am. Acad. Audiol **10**, 343–354.
Pichora-Fuller, M. K. (**2003**). "Processing speed and timing in aging adults: psychoacoustics, speech perception, and comprehension," Int. J. Audiol. **42**, S59–S67.
Rammsayer, T., and Classen, W. (**1997**). "Impaired temporal discrimination in Parkinson's disease: temporal processing of brief durations as an indicator of degeneration of dopaminergic neurons in the basal ganglia," Int. J. Neurosci. **91**, 45–55.
Schneider, B., Speranza, F., and Pichora-Fuller, M. K. (**1998**). "Age-related changes in temporal resolution: envelope and intensity effects," Can. J. Exp. Psychol. **52**, 184–191.
Schneider, B. A., and Hamstra, S. J. (**1999**). "Gap detection thresholds as a function of tonal duration for younger and older listeners," J. Acoust. Soc. Am. **106**, 371–380.
Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**, 980–991.
Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.
Snell, K. B., and Frisina, D. R. (**2000**). "Relationships among age-related differences in gap detection and word recognition," J. Acoust. Soc. Am. **107**, 1615–1626.
Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Age-related changes in within- and between-channel gap detection using sinusoidal stimuli

Antje Heinrich[a] and Bruce Schneider
Center for Research on Biological Communications Systems, University of Toronto, Mississauga, Ontario, L5L 1C6, Canada

(Received 20 July 2005; revised 10 January 2006; accepted 18 January 2006)

Pure tone gap stimuli with identical (within-channel) or dissimilar (between-channel) marker frequencies of 1 and 2 kHz were presented to young and old listeners in a two-interval forced choice gap detection task. To estimate the influence of extraneous duration cues on gap detection, thresholds in the between-channel conditions were obtained for two different sets of reference stimuli: reference stimuli that were matched to the overall duration of the gap stimulus, i.e., two markers plus the gap, and reference stimuli that were fixed at the combined duration of the two markers excluding the gap. Results from within-channel conditions were consistent with previous studies, i.e., there were small but highly reliable age differences, smaller gap thresholds at longer marker durations, and an interaction between the two variables. In between-channel conditions, however, age differences were not as clear cut. Rather, the effect of age varied as a function of duration cue and was more pronounced when stimuli were matched for overall duration than when the duration of the reference tone was fixed. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2173524]

PACS number(s): 43.66.Mk, 43.66.Ba [JHG] Pages: 2316–2326

I. INTRODUCTION

A. Within- and between-channel gap detection

In a two-interval forced-choice gap-detection task the participant is asked to listen to two successively presented sounds and identify the one that contains a small period of silence, the gap. The sound containing the gap consists of three parts: the leading marker (the portion of the sound before the gap), the gap, and the lagging marker (the portion of the sound after the gap). The no-gap reference stimulus consists of the leading and lagging markers with no gap between them. When the two markers consist of pure tones, the frequencies of the two markers can either be the same or different. Numerous studies have shown that gap detection thresholds are sensitive to the spectral similarities of the markers bounding the gap. In general, gap detection thresholds tend to increase and then asymptote, as the spectral difference between the leading and lagging markers becomes greater (Grose *et al.*, 2001; Formby *et al.*, 1998a; Formby and Forrest, 1991; Divenyi and Danner, 1977).

Several mechanisms have been suggested to account for this result (e.g., Formby and Forrest, 1991; Formby *et al.*, 1996; Phillips *et al.*, 1998, 1997; Fitzgibbons *et al.*, 1974). All of these accounts assume that, in the first stage of auditory processing, the signal is processed by a bank of auditory filters, with each filter defining a separate auditory channel. Hence, when the two markers are presented to the same ear and are identical or nearly identical with respect to frequency content, then both markers are processed by the same auditory channel, and the detection of a gap can be based on within-channel comparisons. On the other hand, when the two markers differ substantially with respect to their frequency content and/or are presented to opposite ears, the detection of a gap has to involve the integration and/or comparison of events across two or more auditory channels (between-channel comparisons).

B. Age-related changes in gap detection

Comparatively few studies have investigated the influence of age on gap detection. The results to date suggest that the effect of age on gap detection thresholds differs for within- and between-channel tasks. In within-channel tasks age differences in threshold are generally small and related to marker duration. Schneider and Hamstra (1999), for instance, tested young and old listeners in a gap detection task where both markers were 2-kHz pure tones. The duration of the markers varied in different conditions. They found that whereas there were considerable age differences when marker durations were short, these differences completely vanished for the longest durations. This result integrated previously conflicting evidence from Moore *et al.* (1992) and Schneider *et al.* (1994). Specifically, whereas Moore and colleagues did not find age differences in a within-channel gap detection task when marker durations were large, Schneider *et al.* (1994), using very short durations, found that thresholds for older listeners were generally larger and more variable than for young listeners.

A number of factors have been ruled out as having a direct influence on the age effect in within-channel gap detection. For instance, Schneider *et al.* (1998) demonstrated that age differences are probably not due to either differences in off-frequency processing or longer integration times of the temporal window in older listeners. Moreover, several stud-

[a]Present address: MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, UK. Electronic mail: antje.heinrich@mrc-cbu.cam.ac.uk

ies were unable to find any association between audiometric and gap thresholds, thus indicating that peripheral hearing loss has no direct influence on gap detection thresholds (e.g., Moore *et al.*, 1989; Schneider *et al.*, 1994; Schneider and Hamstra, 1999; Moore *et al.*, 1992; but also see Lutman, 1991, for a contrary position). One mechanism that has been suggested to contribute to the age effect is neural adaptation. Schneider and Hamstra (1999) argue that the fact that age differences are closely linked to marker duration suggests that there are differences in neural adaptation and recovery in response to markers of different lengths.

Lister and colleagues were among the first to examine the effects of age and hearing status in between-channel gap detection tasks (Lister *et al.*, 2000, 2002; Roberts and Lister, 2004). In a first study they tested listeners with and without hearing loss in a gap duration discrimination task in which the two stimuli could be distinguished by the duration of the gap rather than its presence. The rationale for inserting a small (1 ms) gap in the reference stimulus was to ensure that both intervals contained similar gating transients. The stimuli were narrow-band noises. In a *posthoc* analysis Lister *et al.* (2000) tested the effect of age on gap threshold and found differences in threshold for young and old adults. Specifically, they found a steeper increase in threshold for older than for younger listeners as frequency disparity increased. Moreover, consistent with within-channel results, hearing status did not significantly influence gap thresholds. Rather, younger listeners, including those with a severe hearing impairment, always had smaller thresholds compared to even minimally impaired older adults. Hence, the study suggested that in between-channel tasks, like in within-channel tasks, reduced temporal processing in older adults is independent of peripheral hearing loss. In a second study, Lister *et al.* (2002) investigated the matter in greater detail and tested three groups of young, middle-aged, and older adults, respectively. The study showed again that thresholds increase more sharply with increasing frequency disparity for older listeners, a result that was confirmed by Roberts and Lister (2004). Hence, the effect of age is more apparent for frequency-disparate stimuli. Lister *et al.* (2002) speculate that the effect is caused by the fact that perceptual channels are more sharply tuned in older adults, presumably permitting them to use between-channel processing more effectively than younger adults for smaller frequency disparities.

Others have suggested that the increased age effect for between- compared to within-channel tasks can be linked to increased task complexity (Pichora-Fuller, 2003; Pichora-Fuller *et al.*, 2006), an interpretation in concordance with results of a study by Snell (1997) that used more complex low-pass noise stimuli embedded into modulated and unmodulated noise maskers. The results of the Snell study showed a robust age effect for all conditions, even when the noise stimuli were presented in quiet and even though pure-tone thresholds at the test frequencies were comparable for young and old listeners. Moreover, adding a noise floor to the gap stimuli further increased age differences.

In summary, age differences in gap detection in within-channel tasks depend on marker duration and may point to differences in neural adaptation. Age differences are amplified in between-channel conditions and appear to be related to the difficulty in integrating or comparing information across channels. Finally, age differences appear to increase as the spectral and temporal complexity of the markers is increased. One objective of the study was to determine if age differences in between-channel gap detection can be minimized when stimulus complexity is reduced to a minimum. Hence, the present study determined gap detection thresholds for tonal markers (1 and 2 kHz) for both within- and between-channel conditions in young and old adults.

C. Extraneous temporal and spectral cues to gap detection

Extraneous cues, only indirectly related to the size of the gap, can affect between-channel gap detection. For instance, when the marker durations of the leading and lagging markers in both the gap and no-gap stimuli are held constant, the overall duration of the gap stimulus is longer than that of the no-gap reference stimulus by the amount equal to the size of the gap. Thus, the listener could be using overall duration to discriminate between gap and no-gap stimuli. Alternatively, when the overall durations of gap and no-gap stimuli are kept constant, the duration of each marker in the reference stimulus is longer than the duration of each marker in the gap stimulus. Hence, the listener could be performing the task based on the duration of the markers rather than on the perception of a gap. Differences in marker duration between gap and no-gap stimuli, however, are only effective for stimuli with dissimilar frequencies in leading and lagging marker because when the frequency content is identical in both markers, the listener cannot tell when the leading marker in the reference stimulus ends and the lagging marker begins. Hence, for within-channel tasks, equating overall duration removes any possibility that the task is based on marker duration.

A number of studies (Formby and Forrest, 1991; Formby and Muir, 1989; Penner, 1977) have demonstrated that listeners do indeed use extraneous duration cues to aid gap detection which in turn can lead to an underestimation of the true gap detection threshold. However, these studies also show that not all listeners use these duration cues to the same extent and that different listeners may use different cues or no extraneous cues at all. Support for the notion that duration cues, especially overall duration cues, could possibly modulate gap detection accuracy also comes from a different line of research. Bergeson *et al.* (2001) investigated duration discrimination for pure tones and found that for young adults the Weber fraction for a 20-ms tone duration is 0.6, decreasing to 0.3 at 200 ms. Thus, given an overall marker duration of 40 ms in the reference stimulus, duration differences could become noticeable when the target exceeds 64 ms in overall duration; this equals a gap size of 24 ms for younger adults. For older adults the Weber fraction is 1.3, and duration differences would only be noticeable at a much higher level (92 ms). Gap thresholds of over 20 ms are not unreasonable for young adults, and gap durations larger than 52 ms have been observed in older adults, especially for between-channel conditions.

Whereas Formby and Forrest (1991) as well as Penner (1977) aimed to minimize duration cues in their studies by randomizing the duration of each marker independently, we chose to estimate their influence instead. Therefore, we included two sets of between-channel conditions that were specifically constructed to estimate the effects of overall and relative marker duration cues. In the first set of between-channel conditions, the overall duration of the reference stimulus matched that of the target. Hence, the individual durations of the two markers in the no-gap stimulus were longer than those in the gap stimulus. In a second set of conditions individual marker duration was kept constant between gap and no-gap stimuli, so that the overall duration of the no-gap stimulus was shorter than that of the gap stimulus.

Another possible cue to the presence of a gap is the spectral splatter that arises from the termination of the first marker and the onset of the second marker. If the amount of spectral splatter varies with gap duration, then spectral splatter could be used to determine the presence or absence of a gap. To minimize the effects of spectral splatter we constructed each marker by multiplying a pure tone by an envelope consisting of a sum of Gaussian envelopes, with the separation between the means of each successive Gaussian envelope equal to one standard deviation unit. This has been shown to minimize spectral cues in within-channel gap detection for short duration stimuli (Schneider and Hamstra, 1999, Schneider *et al.*, 1994). In the Appendix, we show that the likelihood that the detection of a gap in between-channel conditions is based on spectral splatter is negligible. Hence, we can be reasonably certain that the gap detection thresholds observed here are based on temporal and not spectral cues.

In the present study we compared gap detection performance for young and old listeners in within- and between-channel tasks using two different marker durations (10 and 20 ms). For within-channel tasks, Schneider and Hamstra (1999) found that the size of the age difference decreased as marker duration increased. For between-channel tasks, several studies suggested that age differences increase for more complex stimuli. Our first objective was to measure the extent of age differences for between-channel stimuli when stimulus complexity was reduced to a minimum. A second objective was to see if age differences also decreased as a function of marker duration for between-channel comparisons. Lastly, we sought to estimate the influence of extraneous duration cues on gap detection performance for young and old listeners in between-channel conditions.

II. METHODS

A. Participants

The final sample consisted of 24 young (mean age: 21.3 years; s.d.: 1.81; 14 females) and 24 older listeners (mean age: 71.5 years; s.d.: 4.85; 12 females). Three young listeners had to be replaced because they were not available for all experimental sessions, and three older listeners had to be replaced because they were not able to perceive the gap in between-channel conditions even with the longest duration. All listeners (with the exception of one in the older age group) had hearing thresholds at 1 and 2 kHz in the test (left) ear that were ⩽25 dB HL. One older listener had a hearing threshold of 35 dB HL at 2 kHz. Thresholds at the two test frequencies were significantly higher for older than they were for younger adults, and the extent of this age difference increased with increasing frequency. Audiometric thresholds for a range of frequencies in the test (left) ear are displayed in Table I. The older adults could be characterized as having at most a slight hearing loss at the test frequencies (Clark, 1981) and as being in the early stages of presbyacusis.

TABLE I. Mean and standard deviation (SD) of hearing thresholds in dB HL (ANSI, 1989) for nine pure-tone frequencies in the left (test) ear for young and old listeners. The superscript letters denote significant differences in hearing level between young and old listeners.

	Young		Old	
Frequency in kHz	Mean	SD	Mean	SD
0.25	9.79	4.54	14.17[a]	6.70
0.5	4.38	3.40	10.00[b]	6.76
1	3.13	6.05	7.92[a]	6.58
1.5	3.33	6.54	10.21[b]	8.91
2	3.33	5.45	10.00[b]	9.33
3	0.83	6.86	15.42[b]	8.84
4	3.13	5.86	23.33[b]	14.27
6	11.25	9.12	34.79[b]	17.78
8	6.04	7.07	41.67[b]	18.69

[a] $p<0.05$.
[b] $p<0.01$.

B. Stimuli

The stimuli were digitally generated with a sampling rate of 20 kHz and converted to analog form using a 16-bit Tucker Davis Technology (TDT) digital-to-analog converter. Various techniques were used in an effort to minimize the spectral splatter at the on- and off-set of the stimuli. Firstly, amplitude envelopes for gap and no-gap sounds were constructed by summing a series of Gaussian functions, spaced one standard deviation apart, with a standard deviation of 1 ms (Fig. 1). The sum of the envelopes formed a flat top with ogival rise and decay times. Secondly, all of the gap durations tested preserved the phase of the sinusoidal stimuli. Gaps were created by placing the mean of the first Gaussian in the second marker, an integer number of ms after the mean of the last Gaussian in the first marker. Hence, the smallest possible gap occurred when the first Gaussian in the second marker followed 2 ms after the last Gaussian in the first marker. However, pilot testing had revealed that this gap was perceivable by most of the young listeners in within-channel conditions. In order to create smaller energy dips for within-channel conditions, we filled the smallest gap with a reduced-amplitude Gaussian instead of completely omitting it. Now, the gap was detected as a more or less shallow dip in the stimulus envelope. Originally, gap durations were measured as the distance between the $\frac{1}{2}$-power point at the end of the first marker to $\frac{1}{2}$-power point at the beginning of the second marker. In the case of filled-in Gaussians, however, the amplitude envelope did not decay to zero, but, instead,

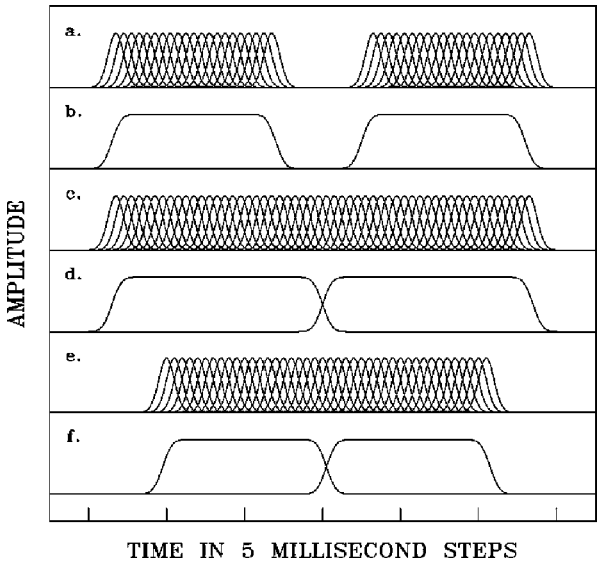


FIG. 1. (a) The Gaussian envelopes used to create two 20-ms markers (each marker consists of 21 Gaussians) with a 13-ms gap between the two markers. The Gaussian envelopes are spaced 1 ms apart, and there is a 13-ms gap between the last Gaussian in the leading marker and the first Gaussian in the lagging marker. (b) The two envelopes formed by summing the Gaussians in (a). (c) The Gaussians used to form the two marker envelopes of the no-gap stimulus that is equivalent in total duration to the 13-ms gap stimulus shown in (a) and (b). There are 27 Gaussians in the leading marker and 27 Gaussians in the lagging marker, with 1 ms separating the last Gaussian in the leading marker and the first Gaussian in the lagging marker. (d) The two envelopes formed by summing the Gaussians in (c). (e) The Gaussians used to form the envelopes of the no-gap stimulus whose marker durations are the same as in (a). There are 21 Gaussians in the first marker and 21 Gaussians in the lagging marker with 1 ms separating the last Gaussian in the leading marker and the first Gaussian in the lagging marker. (f) The two envelopes formed by summing the Gaussians in (e).

the sum of the two markers began to affect the $\frac{1}{2}$-power points, making these points unusable as points of reference. To avoid this problem when analyzing data, we used the area difference between the gap and the no-gap amplitude envelope functions to index gap duration. All analyses were conducted on the area difference between gap and no-gap envelope. However, in describing the data we converted the area measures back into equivalent gap durations.[1]

The leading and lagging marker envelopes were then multiplied by tones of 1 and 2 kHz. In the within-channel conditions, both leading and lagging marker were multiplied by the same 1- or 2-kHz tone, respectively. In the between-channel conditions, leading and lagging markers were multiplied by different frequencies, either 1 or 2 kHz. The resulting product was normalized in such a way that all stimuli, gap and no-gap, contained the same amount of energy (Fig. 1). The marker duration for the gap stimuli was kept constant within each condition at 10 or 20 ms, respectively. Two different sets of between-channel conditions were tested: a set where the overall duration of the no-gap stimulus matched that of the gap stimulus, and a set of conditions where the duration of the no-gap stimulus was fixed at the combined marker duration of leading and lagging markers, i.e., at either 20 or 40 ms. Stimuli matched for overall duration were also used in all within-channel conditions. Note that under conditions of matched duration, the reference stimuli were constructed in such a way that they matched gap stimuli in overall duration for odd gap durations (e.g., 9 ms) and exceeded them in overall duration by 1 ms in cases of even gap sizes (e.g., 8 ms). The sound pressure level of the pure tones before they were multiplied by the marker envelopes was 70 dB SPL. After the pure tones were multiplied by the amplitude envelopes, they were rescaled so that the total energy in a stimulus was equated across all conditions. Hence, there were no energy differences between gap and no-gap stimuli. Stimuli were presented to the left ear over TDH-49 earphones in a single-wall sound-attenuating booth.

C. Psychoacoustic procedure

A three-down, one-up tracking procedure was used in a 2AFC paradigm to determine the 79.4% accuracy point on the psychometric function (Levitt, 1971). At the beginning of each run, the gap size was set to its maximum value, 31 ms in within-channel conditions and 61 ms in between-channel conditions. In each trial, which was initiated by a button press, a gap and a no-gap stimulus were presented, randomly assigned to one of two 500-ms-long stimulus intervals, separated by 100 ms. The task was to indicate which of the two intervals contained the gap stimulus. After three consecutive correct responses the gap size of the stimulus was decreased. After each incorrect response the gap size was increased. The size of the increment or decrement in gap was halved after each reversal until a minimum step size of 1 ms was reached. Each stimulus began 100 ms into the interval.

Participants indicated which interval they thought contained the gap by pressing one of two buttons. Lights on the button box indicated the beginning and the length of each interval and also provided immediate feedback to the subject about the accuracy of their response. Each run was ended after 12 reversals. The gap-detection threshold for the condition was defined as the average of the last eight reversals. Participants cycled through all conditions a total of four times. The gap threshold for each condition was defined as the geometric mean of the thresholds of the four runs. All listeners came in for three successive sessions of 1.5 h each; thresholds for within-channel conditions were obtained in session 1, for between-channel conditions with the reference stimulus matched in total duration in session 2, and for between-channel conditions with fixed marker durations in session 3. Testing within each session was counterbalanced for frequency content of the leading marker and marker duration between listeners. The testing was self-paced and initiated by a button press. Listeners were not given practice trials prior to data collection. Table II details all conditions tested in the study.

III. RESULTS

A. Within-channel gap detection

Figure 2 displays mean gap thresholds in milliseconds for young and old listeners in all four within-channel conditions. The thresholds are generally very small: between 0.9 and 1.6 ms for young listeners and 1.2 and 2.5 ms for older listeners. Moreover, the data show that the shorter marker duration (10 ms) generally led to longer gap thresholds.

TABLE II. Stimulus conditions used in the study. Markers varied in frequency similarity and frequency order. Each condition was run a total of four times.

Session no.	Marker duration (ms)	Frequency of leading and lagging markers
(1) Within-channel	10	1 kHz, 1 kHz 2 kHz, 2 kHz
	20	1 kHz, 1 kHz 2 kHz, 2 kHz
(2) Between-channel, matched reference stimulus	10	1 kHz, 2 kHz 2 kHz, 1 kHz
	20	1 kHz, 2 kHz 2 kHz, 1 kHz
(3) Between-channel, fixed reference stimulus	10	1 kHz, 2 kHz 2 kHz, 1 kHz
	20	1 kHz, 2 kHz 2 kHz, 1 kHz

Lastly, thresholds appear to be higher for 2-kHz than for 1-kHz markers and older listeners generally exhibited larger thresholds than young listeners.

A 2 marker duration (10 ms, 20 ms) by 2 marker frequency (1 kHz, 2 kHz) by 2 age group (young, old) mixed measures ANOVA with marker duration and marker frequency as within-subject variables and age as a between-subjects variable supported the graphical results.[2] The ANOVA revealed main effects of marker duration [$F(1,46)=74.25, p<0.0001$] (20-ms markers led to smaller thresholds than 10-ms markers), frequency [$F(1,46)=70.78, p<0.0001$] (2 kHz stimuli resulted in larger overall thresholds), and age [$F(1,46)=12.28, p<0.001$] (young listeners generally achieved smaller thresholds). Moreover, the interaction of marker duration and age reached significance [$F(1,46)=4.52, p<0.05$]. This interaction indicated that the age difference was more pronounced for the 10-ms compared to the 20-ms marker duration, even though the age difference reaches significance in both instances [10-ms duration: $t(46)=-3.62, p<0.001$; 20-ms duration: $t(46)=-3.02, p<0.01$]. None of the other interactions were significant.

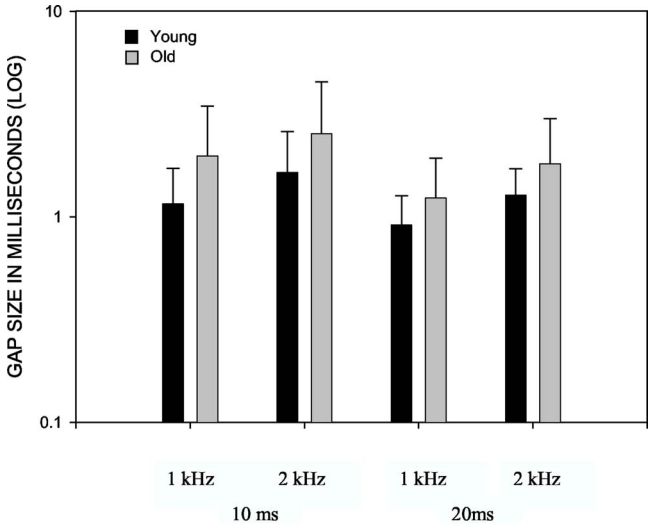


FIG. 2. Mean gap detection thresholds for young (black) and old (gray) listeners for 10- and 20-ms markers of identical frequency (within channel). The stimulus frequency of both markers was either 1 or 2 kHz. Error bars depict 1 standard deviation above the mean.

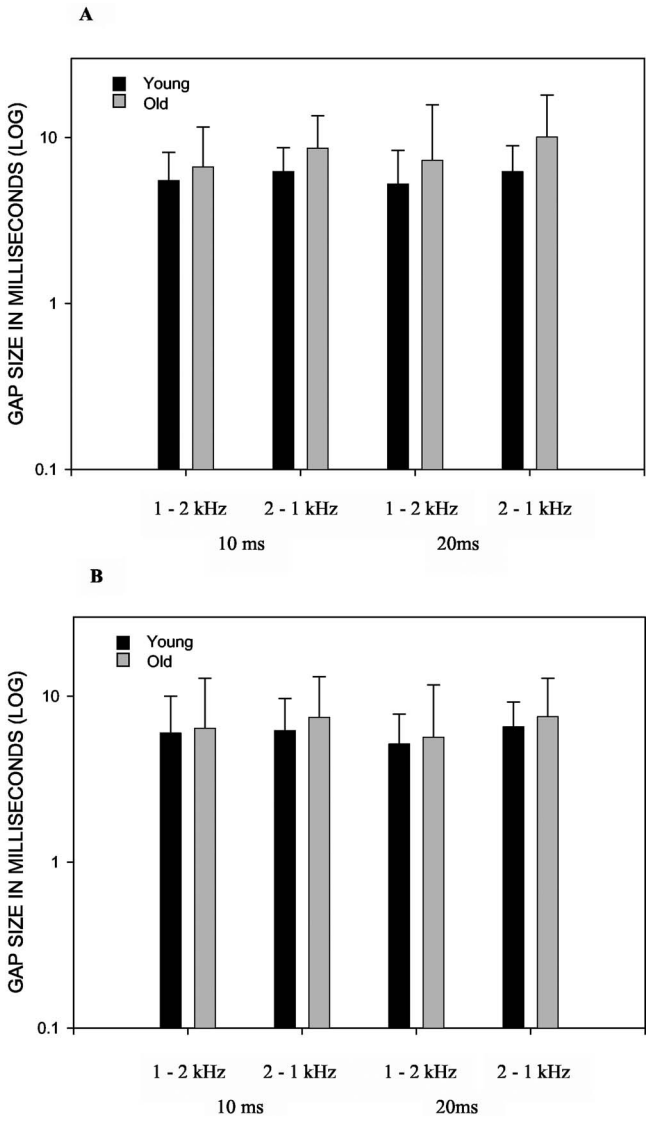


FIG. 3. (A) Mean gap detection thresholds for young (black) and old (gray) listeners for 10 and 20-ms markers of dissimilar frequency (between channel). Stimulus frequency either ascended (1 kHz in the first marker and 2 kHz in the second marker) or descended. Panel (A) presents the thresholds for the condition in which the no-gap stimulus matched the gap stimulus in total duration. Panel (B) presents the thresholds for the condition in which the duration of the markers in the no-gap stimulus was the same as the duration of markers in the gap stimulus. Note that in condition A the duration of the no-gap stimulus co-varied with that of the gap stimulus while in condition B, the no-gap stimulus was constant throughout the session. Error bars depict 1 standard deviation above the mean.

B. Between-channel gap detection

Figure 3 depicts gap detection thresholds for young and old listeners in between-channel conditions when the no-gap reference tone was (A) adjusted in each trial to match the overall duration of the gap stimulus or (B) fixed at the combined marker duration of the two markers in the gap stimulus. The data show age differences in both plots but these

differences appear to be more pronounced in the matched condition than in the fixed condition. Moreover, there is a consistent advantage for gap thresholds when the marker frequencies before and after the gap are ascending from 1 to 2 kHz compared to when they are descending. Lastly, thresholds are in general smaller in fixed marker conditions than in adjusted marker conditions.

A mixed factor ANOVA with marker duration (10 ms, 20 ms), frequency order (ascending from 1 kHz to 2 kHz or descending), and duration cue (matched or fixed) as within-subject variables and age as a between-subject variable revealed a significant main effect of frequency order [$F(1,46) = 27.92, p < 0.0001$] but no main effect of age, marker duration, or duration cue. The significant effect of frequency order confirms that thresholds are smaller for ascending than for descending frequency order. Moreover, there were three significant interaction effects, two of them involving the duration cue: between duration cue and age group [$F(1,46) = 4.97, p < 0.05$] and duration cue and marker duration [$F(1,46) = 5.02; p < 0.05$], and between marker duration and frequency order [$F(1,46) = 4.87, p < 0.05$]. No other interactions were significant.

The interaction between duration cue and age reflects the fact that age differences in threshold were more pronounced when stimuli were adjusted for overall duration than when the overall duration of the reference tone was fixed. In fact, a significant age effect was only present when reference stimuli were matched in overall duration to the gap stimuli [$t(46) = -2.70, p = 0.01$], but not when the overall duration of the reference stimuli was fixed [$t(46) = -0.92, p > 0.05$].

The interaction between duration cue and marker duration was caused by the fact that when the reference stimulus changed with gap size, gap thresholds were larger for 20-ms than for 10-ms markers. On the other hand, when the reference stimuli were fixed in overall duration, thresholds were smaller for 20-ms than for 10-ms markers. However, neither of these threshold changes were significant [$t(47) = -1.17, p = 0.25$ when the reference stimulus changed with gap size; $t(47) = 1.31, p = 0.20$, when the reference stimulus was fixed].

Lastly, the interaction between marker duration and frequency order was caused by the fact that, for an ascending frequency order, thresholds decreased in size when going from 10- to 20-ms duration markers. In contrast, for the descending frequency order thresholds were larger when the markers were 20 ms in duration compared to 10 ms. Yet, neither of these threshold changes at each marker duration reached significance [ascending frequency order: $t(47) = 1.26, p = 0.21$; descending frequency order: $t(47) = -1.53, p = 0.13$].

C. Correlation between pure-tone thresholds and within-channel gap detection thresholds

Lastly, in order to rule out that the age differences in the within-channel conditions were caused by age-related pure-tone threshold elevations, we computed Pearson product-moment correlations between listeners' gap detection thresholds for a particular frequency condition and their respective hearing level at that frequency. For instance, the correlation between individual gap detection thresholds for the 1-kHz within-channel condition with 10-ms markers and hearing levels at 1 kHz were −0.01 for young adults and −0.28 for old adults, respectively. Correlations were equally small for the other three within-channel conditions, ranging from 0.07 to 0.16 in young listeners and from −0.27 to −0.36 in old listeners. None of the correlations reached significance. Hence, there is no evidence that within-channel gap detection thresholds vary with hearing level.

To determine whether between-channel performance was related to hearing levels at each of the two frequencies, we correlated gap thresholds for each of eight conditions (see Table II) with audiometric thresholds at each of the two test frequencies. This yielded 16 correlation coefficients for younger adults and 16 correlation coefficients for older adults. Of the 32 correlations tested only one reached the alpha=0.05 level of significance. Under the null hypothesis that hearing level and gap detection thresholds were independent of each other we would expect 5% of the 32 correlation coefficients (or 1.6 correlation coefficients) to reach significance. Moreover, in the one case that reached significance (which occurred in older adults), the sign of the correlation indicated that gap detection thresholds improved as hearing levels worsened. Because of this, and because only one correlation coefficient turned out to be significant, there is no evidence to suggest that between-channel gap detection performance is correlated with hearing levels at either of the two frequencies.

IV. DISCUSSION

A. Within-channel gap detection

Figure 2 and its associated ANOVA indicated that (1) older adults had larger gap detection thresholds than younger adults; (2) gap detection thresholds decreased for both younger and older adults as marker duration increased from 10 to 20 ms; (3) the size of the age difference decreased as marker duration increased from 10 to 20 ms; and (4) gap detection thresholds were lower for the 1-kHz markers than they were for the 2-kHz markers for both younger and older adults. The first three results are consistent with the findings from other within-channel gap detection studies that have used 1- or 2-kHz markers (Moore *et al.*, 1992; Schneider *et al.*, 1994, 1998; Schneider and Hamstra, 1999; Strouse *et al.*, 1998). However, an effect of marker frequency on performance has not been consistently shown. Whereas some studies (Shailer and Moore, 1987; Moore and Glasberg, 1988; Formby and Forrest, 1991) found no appreciable difference in gap threshold for 1- and 2-kHz marker frequencies, others (Moore *et al.*, 1992) seemed to show that for both young and older normal listeners gap detection thresholds increased by about 1 ms when marker frequencies increased from 1 to 2 kHz. The results from the present study are in agreement with the latter results and show that an increase in marker frequency from 1 to 2 kHz led to larger thresholds.

Hence, there does appear to be a significant age difference in within-channel gap detection for short duration tonal markers that decreases with increasing marker duration and is uncorrelated with hearing loss.

B. Between-channel gap detection

Frequency order: A number of interesting results arose from the between-channel conditions. First, frequency order was the only significant main effect, whereas all other variables were only involved in interaction effects. The main effect of frequency order indicates that going from a 1-kHz marker to a 2-kHz marker resulted in smaller gap detection thresholds than when the 2-kHz marker preceded the 1-kHz marker. This result is in agreement with several studies that showed a similar asymmetry effect when the lagging marker has a higher frequency than the leading marker (Formby *et al.*, 1996; Lister *et al.*, 2000, 2002). However, other studies have not found an asymmetry in thresholds with respect to whether tonal frequency increased or decreased from the first to the second marker (Fitzgibbons *et al.*, 1974; Formy *et al.*, 1998b).

One possible explanation for this frequency order effect is harmonic distortion. If the 1-kHz leading marker has distortion products at 2 kHz, then the participant could be basing her or his judgment on events occurring in the 2-kHz channel, thereby turning a between-channel task into a within-channel task. To check whether the earphone tranducer was producing distortion products at 1 kHz, we recorded the sound pressure level produced by the earphones in the ear canal of a dummy head (Kemar), using the Bruel & Kjaer Pulse platform. Those measurements showed that any harmonic distortion produced by the earphone was more than 70 dB down from the signal intensity. Hence, headphone distortion products cannot account for the observed effect.

However, it is possible that the presence of aural harmonics in the inner ear could have turned the between-channel task into a within-channel one. Aural harmonics or subjective tones refer to the phenomenon that, in the course of the sound transmission from the air to the basilar membrane, the ear sometimes introduces new frequencies into the original sound if the intensity of the pure tone is high enough (e.g., Wegel and Lane, 1924; Fletcher, 1930). For instance, Steinberg (1935) claims that any tone below 1 kHz will produce noticeable harmonics when its intensity level exceeds 50 dB. In fact, for high-intensity sounds, the intensity of the harmonics may exceed the magnitude of the fundamental (Steinberg, 1935). Lawrence and Yantis (1956) extended the demonstration of aural harmonics to frequencies of up to 5 kHz. Specifically, they determined that aural harmonics first occurred for a 1-kHz tone at 50 dB SPL. At 70 dB SPL the second aural harmonic had an intensity of approximately 55 dB SPL. In the present case, the stimulus is a 1-kHz tone whose level was approximately 70 dB SPL before it was multiplied by the marker envelopes. Hence, according to these measurements, we can assume that in addition to the original 1-kHz frequency, sound energy was being generated by the ear for frequencies of 2 kHz and possibly even higher harmonics. The second harmonic is of specific interest to us because if energy was present not only at 1 but also at 2 kHz, then the task may have acquired a within-channel component. Hence, rather than basing their evaluation of the temporal gap solely on the temporal characteristics of the 1- and 2-kHz tone presentation, the listener may have also used within-channel cues provided by the second harmonic. When the first marker is a 1-kHz tone it will generate an aural harmonic at 2 kHz. Hence, if the listener attends to the 2-kHz channel, she or he should hear a weak 2-kHz tone (the aural harmonic of the 1-kHz leading marker) when the leading marker is played, followed by a stronger 2-kHz tone when the lagging marker (a 2-kHz tone) is played. This would turn a between-channel task into a within-channel task. Divenyi and Danner (1977) found that the intensity of the leading marker could be as much as 25–45 dB lower than that of the second marker before it began to have a substantial effect on gap discrimination (identifying which of two gaps is larger). Hence, provided that the aural harmonic produced by the leading 1-kHz tone in this experiment is sufficiently intense, participants could possibly base their judgments on events occurring in the 2-kHz channel.

Although aural harmonics generated by a lagging 1-kHz tone could also turn a between-channel task into a within-channel task, the effectiveness of monitoring the 2-kHz channel is likely to be diminished because of forward masking effects. If the first marker is a 2-kHz tone, the listener is presented with a strong 2-kHz tone followed by the weaker, 2-kHz, aural harmonic generated by the lagging 1-kHz tone. It has long been known that the presence of an intense masker elevates the threshold for a same-frequency probe stimulus following the masker (e.g., Duifhuis, 1973; McFadden and Yama, 1983; Nizami, 2003). Hence, in the ascending frequency case, the 2-kHz tone is present as a weak-intensity aural harmonic in the leading marker and a 2-kHz tone in the lagging marker. In the descending frequency case, the stronger 2-kHz tone is presented first and the weaker aural harmonic second. The data of Plomp (1964) and Penner (1977) show that it is much more difficult for observers to perceive a gap when the intensity of the first marker is greater than the intensity of the second marker than when the two markers have equal intensities. Thus the frequency-order asymmetry may result from a combination of the presence of aural harmonics at 2 kHz coupled with forward-masking effects.

Marker duration: The data did not provide evidence for a general effect of marker duration for durations of 10 and 20 ms. Although no studies have investigated the effects of marker duration on gap detection threshold for between-channel stimuli, at least one study has investigated the effect for within-channel stimuli. Schneider and Hamstra (1999) showed that age differences in within-channel gap detection thresholds declined as a function of marker duration, finally disappearing at the longest marker duration tested (500 ms).

Marker duration and frequency order: The interaction between marker duration and frequency order was caused by the fact that in the case of an ascending frequency order longer marker durations led to smaller gap thresholds whereas in the descending frequency order conditions the opposite was true. This result is consistent with the notion

that, due to the presence of aural harmonics, the ascending frequency conditions contained a within-channel component and therefore changes in marker duration in these conditions had similar effects as in pure within-channel conditions.

Marker duration and duration cue: Previous within-channel studies (e.g., Schneider and Hamstra, 1999) have indicated that gap thresholds decrease with increasing marker duration, a result that was replicated in our within-channel conditions. In the between-channel case, gap thresholds also decreased with increasing marker duration when the overall duration of the reference stimulus was fixed. Thus, the effects of marker duration on gap detection thresholds are the same for both within- and between-channel conditions when the overall duration of the no-gap or reference stimulus in the between-channel conditions is fixed. However, when the duration of the reference stimulus was matched to that of the gap stimulus in the between-channel case, gap detection thresholds increased as the duration of the markers defining the gap changed from 10 to 20 ms. This deviation from the expected pattern may reflect the fact that it is more difficult to determine the presence of a gap when the duration of the markers in the no-gap or reference stimulus are changing with the size of the gap.

Age: The only significant age effect in between-channel conditions occurred when the stimuli were matched for overall duration. This result is in stark contrast to the clear age differences that have been found in at least four between-channel gap detection studies (Lister *et al.*, 2000, 2002; Pichora-Fuller *et al.*, 2006, Roberts and Lister, 2004). We propose that the overall size of the gap thresholds as well as the size of the age effect is a function of the complexity of the stimuli. In this context stimulus complexity can be seen as a continuum where pure-tone markers represent the simplest form of gap markers that lead to the smallest overall thresholds and the smallest age differences. Markers consisting of harmonic complexes, noises, and speech sounds represent more complex sounds leading to generally higher thresholds and more pronounced age differences (Lister *et al.*, 2000, 2002; Haubert and Pichora-Fuller, 1999).

Duration cue and age: For older adults, gap detection thresholds were lower in between-channel conditions when marker duration was fixed than they were when gap and no-gap stimuli were matched in overall duration. In contrast, in younger adults, the two types of marker duration cues had no effect on gap detection thresholds.

Because the testing sequence was always such that conditions with matched overall durations were tested prior to conditions with fixed marker durations, we tried to assess whether practice effects may have been responsible for the effect of duration cue on older adults. To assess the effects of practice on performance, we looked for evidence of improvements in performance as a function of the number of three-down, one-up tracking runs in each between-channel condition in an omnibus ANOVA with duration cue, marker duration, frequency order, and trial number as within-subject variables. There was no indication of a practice effect for young [$F(3,69)=1.71$, $p=0.1$] or old adults [$F(3,69)<1$, $p=0.51$], confirming that in neither age group did gap detection thresholds decrease with practice. Moreover, practice also did not interact with duration cue for young [$F(3,69)<1$, $p=0.98$] or old adults [$F(3,69)<1$, $p=0.98$]. Hence, it is unlikely that learning was responsible for the effect that the two types of duration cues had on older adults.

Because the gap detection thresholds of the older listeners were more variable in all between-channel conditions, it is possible that some older adults used duration cues while others did not, or that older listeners used duration cues in some runs but not in others. This interpretation would also be in agreement with Formby and Muir (1989) who found that different listeners used different cues at different times.

Another possibility is that the presence of two sets of discriminable cues (marker duration cues and gap cues) at the larger gap sizes was more confusing to older than to younger adults. Because our adaptive procedure started with gap sizes of 61 ms, participants, when encountering these larger gap sizes, could base their decision on the difference in gap size (0 vs. 61 ms), the difference in overall duration for the condition in which marker duration was fixed (20 vs. 81 ms for the 10-ms markers), or on marker duration when gap and no-gap stimuli were matched in overall duration (10 vs. 40 ms for the 10-ms markers). Because age differences are more likely to appear when either the stimuli or the task is more complex, it is possible that some older adults performed less well than younger adults because of the additional complexity of the task. Moreover, one could argue that stimulus complexity was greater when the gap and no-gap stimuli were equated for overall duration than when the marker duration was fixed. In the former condition both the total duration of the no-gap stimulus and the duration of each of its markers covaried with gap duration. In the latter, the no-gap stimulus was unchanged from trial to trial. It is possible that older adults found it more difficult than younger adults to make the discrimination when the duration of the no-gap stimulus varied from trial to trial, but not when the no-gap reference stimulus remained unchanged over an adaptive run. It is important to note that such trial-by-trial changes in total duration are unlikely to affect performance in the within-channel task because the overall durations are so short that trial-by-trial variation in overall duration would not be noticeable near threshold levels.

Finally, there is always the possibility that temporal acuity was particularly well preserved in our sample of older adults and that this is the reason why we did not find any substantial age differences in between-channel tasks. However, the fact that the older adults in our sample performed worse than the younger adults in within-channel gap detection tasks, and that the pattern of age differences in within-channel gap detection was consistent with previous studies, suggests that our group of older adults was not atypical with respect to temporal acuity. Hence, we conclude that age differences in between-channel tasks increase with stimulus complexity.

V. SUMMARY

Within-channel, but not between-channel, pure-tone gap detection thresholds were consistently higher for older than for younger adults. The failure to find substantial age differ-

ences in between-channel gap detection in the present experiment stands in sharp contrast to a number of studies in which substantial age differences were reported (e.g., Lister *et al.*, 2000; 2002; Pichora-Fuller *et al.*, 2006; Roberts and Lister, 2004). However, those studies reporting substantial age effects have all used spectrally complex markers (noise bands and/or speech stimuli) to define a gap whereas our markers were pure tones. This suggests that age differences in gap detection are modulated by stimulus complexity, with greater complexity leading to greater age differences. The fact that age differences were larger in the between-channel condition of the present study when the duration of the no-gap stimulus varied over trials than when the no-gap stimulus was invariant over trials also suggests that age differences may be magnified by task complexity. Finally, the fact that between-channel performance was better when the first and second markers were 1 and 2 kHz, respectively, than when the same markers appeared in the reverse order is consistent with the notion that both younger and older adults can use the aural harmonics produced in the ear by the first marker to aid gap detection.

APPENDIX: POWER SPECTRA FOR GAP AND NO-GAP STIMULI WITH EQUAL DURATION MARKERS

A stimulus with a gap consists of two markers, a leading marker and a lagging marker with a gap between them. Consider the case where the leading and lagging markers have equal duration, with each marker consisting of a sinusoid multiplied by a temporal window. Let each temporal window consist of a sum of $n+1$ Gaussians spaced 1 standard deviation (σ seconds) apart. Furthermore, assume that the first window is multiplied by a sinusoid whose frequency is r_1/σ Hz, and that the second window is multiplied by a sinusoid whose frequency is r_2/σ Hz, where r_1 and r_2 are both integers greater than 0. (In the present experiment $r_1=1$, $r_2=2$, $\sigma=0.001$ s.) Finally, a gap between the two markers is produced by centering the first Gaussian in the envelope of the second marker, $(m+1)\sigma$ seconds after the center of the last Gaussian in the first marker, where m is an integer. Specifically, if $g_1(t)$ is the first marker, and $g_2(t)$ is the second marker, then

$$g_1(t)=\cos\left(\frac{2\pi r_1 t}{\sigma}\right)\left[\sum_{k=0}^{n} e^{-(t-k\sigma)^2/2\sigma^2}\right], \tag{A1}$$

$$g_2(t)=\cos\left(\frac{2\pi r_2 t}{\sigma}\right)\left[\sum_{k=m+n+1}^{m+2n+1} e^{-(t-k\sigma)^2/2\sigma^2}\right]. \tag{A2}$$

Note that if $m=0$, there is no gap between the first and second markers, that is, the sum of the envelopes of the two markers is flat during the switchover from frequency r_1/σ, to frequency r_2/σ. If m is an integer greater than zero, there is a gap between the two markers equal to $(m+1)\sigma$.

Figures 1(a) and 1(b) show how a 13-ms gap is inserted between two 20-ms markers ($n=20$, $m=12$). The standard deviation of the Gaussians shown in Fig. 1(a) is 1 ms, and the means of the Gaussians are spaced 1 ms apart. Therefore the envelope for the first marker is obtained by summing the first 21 Gaussians in Fig. 1(a). If the center of the first Gaussian is at T_0 ms, the center of the 21st Gaussian is at $T_0+n\times\sigma=T_0+20$ ms. Hence the duration of the first marker is 20 ms. The location of the first Gaussian in the second marker is set at $T_0+n\sigma+(m+1)\sigma=T_0+20+13$ ms. Therefore the duration of the gap is 13 ms. The 21 Gaussians defining the envelope of the second marker begin at $T_0+n\sigma+(m+1)\sigma=T_0+20+13$ ms and end at $T_0+2n\sigma+(m+1)\sigma=T_0+40+13$ ms. Hence the duration of the second marker is also 20 ms. The envelopes of the two markers are obtained by summing the Gaussians associated with each marker. The results of this summation process are shown in Fig. 1(b) (the envelopes have been normalized by setting the maximum amplitude to 1.0).

Figures 1(c) and 1(d) show how the envelopes of the no-gap comparison stimulus are constructed for the condition in which the duration of the no-gap comparison stimulus is equal to the total duration of the gap stimulus. The total duration of the gap stimulus is $2n\sigma+(m+1)\sigma$ ms=53 ms. The equal duration comparison stimulus is obtained by summing 54 Gaussians. The first 27 Gaussians define the envelope for the first marker. The second 27 Gaussians define the envelope of the second marker. The two summed envelopes are shown in Fig. 1(d).

Finally, Figs. 1(e) and 1(f) show how the envelope of the no-gap comparison is constructed for the condition in which the duration of each of the markers in the no-gap comparison stimulus is equal to the duration of each of the markers in the gap stimulus. The first 21 Gaussians are used to determine the envelope of the first marker, the second 21 Gaussians the envelope of the second marker. The two marker envelopes are shown in Fig. 1(f). For both types of comparison stimuli the sum of the two marker envelopes is flat during the transition period between the first and second markers.

To obtain the energy density function for $g_1(t)+g_2(t)$ we note that the Fourier transform of $g_1(t)$ is

$$\int_{-\infty}^{\infty}\cos(2\pi ft)\cos\left(\frac{2\pi r_1 t}{\sigma}\right)\left[\sum_{k=0}^{n} e^{-(t-k\sigma)^2/2\sigma^2}\right]dt - j\int_{-\infty}^{\infty}\sin(2\pi ft)\cos\left(\frac{2\pi r_1 t}{\sigma}\right)\left[\sum_{k=0}^{n} e^{-(t-k\sigma)^2/2\sigma^2}\right]dt \tag{A3}$$

and the Fourier transform of $g_2(t)$ is

$$\int_{-\infty}^{\infty}\cos(2\pi ft)\cos\left(\frac{2\pi r_2 t}{\sigma}\right)\left[\sum_{k=m+n+1}^{m+2n+1} e^{-(t-k\sigma)^2/2\sigma^2}\right]dt - j\int_{-\infty}^{\infty}\sin(2\pi ft)\cos\left(\frac{2\pi r_2 t}{\sigma}\right)\left[\sum_{k=m+n+1}^{m+2n+1} e^{-(t-k\sigma)^2/2\sigma^2}\right]dt. \tag{A4}$$

Hence the Fourier transform of $g_1(t)+g_2(t)$ is the sum of the quantities in (A3) and (A4). It can be shown[3] that the spectral density function of $g_1(t)+g_2(t)$ [the square of this Fourier transform of $g_1(t)+g_2(t)$] is

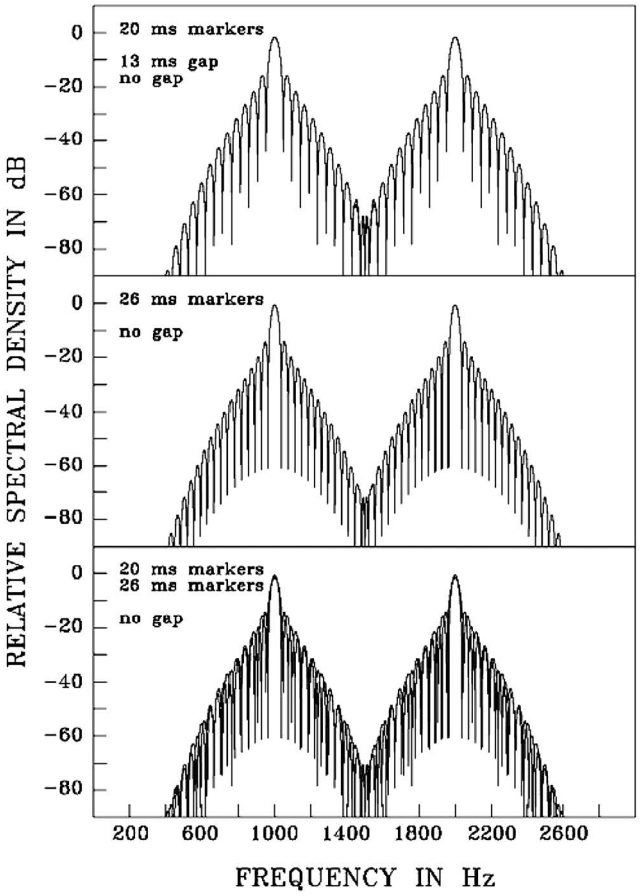


FIG. 4. Top panel: relative spectral densities of two 20-ms markers (one at 1 kHz, the other at 2 kHz) separated by a gap of 13 ms [see Figs. 1(a) and 1(b)], and the same two markers without a separation between them [the no-gap stimulus with equivalent marker durations, Figs. 1(e) and 1(f)]. Middle panel: the spectral density of the no gap stimulus whose total duration is equivalent to that of the stimulus with a 13-ms gap. Each marker duration is 26 ms long with a 1-ms separation between the end of the first marker and the start of the second [Figs. 1(c) and 1(d)]. Bottom panel: spectral densities of the two no-gap reference stimuli, one whose marker durations are equivalent to those in the gap stimulus (20 ms) and one whose marker durations are 26 ms each so that the total duration of the two markers is equivalent to that of the stimulus with the gap.

$$\frac{1}{2}\sigma^2\pi\frac{\sin^2((1+n)\pi f\sigma)}{\sin^2(\pi f\sigma)} \times[A^2+B^2+2AB\cos(2(m+n+1)\pi f\sigma)], \tag{A5}$$

where

$$A=[e^{-2\pi^2\sigma^2(f-r_1/\sigma)^2}+e^{-2\pi^2\sigma^2(f+r_1/\sigma)^2}],$$

$$B=[e^{-2\pi^2\sigma^2(f-r_2/\sigma)^2}+e^{-2\pi^2\sigma^2(f+r_2/\sigma)^2}]. \tag{A6}$$

Figure 4 plots the energy density functions (in decibels) for the gap and the two no-gap stimuli when $\sigma=0.001$ s, $r_1=1$, and $r_2=2$, so the two frequencies in the cross-channel condition become 1 and 2 kHz, respectively, $n=20$, so that the two marker durations are each 20 ms long and $m=12$, thereby producing a gap of 13 ms. The top panel of Fig. 4 plots the energy density functions for the stimulus with a 13-ms gap between two 20-ms markers [Fig. 1(b)], and for the no-gap comparison stimulus whose marker durations are 20 ms [Fig. 1(f)]. Note that except for the small region near 1500 Hz these two energy density functions are essentially equal. The middle panel of Fig. 4 plots the energy density function for the no-gap comparison stimulus whose total duration is equal to that of the gapped stimulus. Note that the overall form of this spectral density function is quite close to that of the no-gap comparison stimulus whose marker durations are equal to those in the gap stimulus. This is illustrated in the bottom panel of Fig. 4 where the energy density functions of both comparison stimuli are shown. Therefore, Fig. 4 shows that it is unlikely that the discrimination between the gap stimulus and either of the two no-gap comparison stimuli is based on spectral cues.

[1]Consider a sum of Gaussians. When m Gaussians are removed from that sum it creates a dip in the envelope. We can show that when m Gaussians are removed from the middle of the envelope, the time difference between when the envelope decays to 0.7 of its peak value and when it returns to 0.7 of its peak amplitude is approximately equal to $(m+1)\sigma$. Since the 0.7 position on the amplitude envelope is approximately the 0.5 position on the squared envelope, the time difference between the half power points on the envelope is approximately equal to $(m+1)\sigma$ for m greater than or equal to 1. The difference in area, A, between the full envelope and the envelope missing m Gaussians is equal to $m\sqrt{2\pi}\sigma$. Note that the time gap, t_g (measured in seconds), created by the m missing Gaussians is $t_g=(m+1)\sigma$, for m greater than or equal to 1. Therefore area is linearly related to time gap by $A=(t_g-\sigma)\sqrt{2\pi}$ for $t_g\geqslant 2\sigma$, t_g measured in seconds. When there are no missing Gaussians, the gap duration is defined as 0, and the area is 0. When there is one missing Gaussian the gap duration is 2σ and the area is $\sqrt{2\pi}\sigma$. If we now assume that the area is linearly related to gap duration over this range, area$=(\sqrt{2\pi}/2)t_g$ measured in seconds, and $0<t_g<2\sigma$. All analyses were performed on areas, and means were then converted back to equivalent gap durations using these linear relationships.

[2]Note that Fig. 2 shows gap thresholds in time units (ms) whereas area units were used for all statistical analyses. These area units were log transformed to keep the variances relatively constant over experimental conditions. The log-transformed mean area units were then translated back into time units.

[3]For a proof please contact Bruce A. Schneider at bschneid@utm.utoronto.ca

American National Standards Institute (**1989**). "Specification for audiometers (ANSI S3.6–1989)," New York: ANSI.

Bergeson, T. R., Schneider, B. A., and Hamstra, S. J. (**2001**). "Duration discrimination in younger and older adults," Can. Acoust. **29**, 3–9.

Clark, J. G. (**1981**). "Uses and abuses of hearing loss classification," ASHA **23**, 493–500.

Divenyi, P. L., and Danner, W. F. (**1977**). "Discrimination of time intervals marked by brief acoustic pulses of various intensities and spectra," Percept. Psychophys. **21**, 125–142.

Duifhuis, H. (**1973**). "Consequences of peripheral frequency selectivity for nonsimultaneous masking," J. Acoust. Soc. Am. **54**, 1471–1488.

Fitzgibbons, P. J., Pollatsek, A., and Thomas, I. B. (**1974**). "Detection of temporal gaps within and between perceptual tonal groups," Percept. Psychophys. **16**, 522–528.

Fletcher, H. (**1930**). "A space-time pattern theory of hearing," J. Acoust. Soc. Am. **1**, 311–343.

Formby, C., and Forrest, T. G. (**1991**). "Detection of silent temporal gaps in sinusoidal markers," J. Acoust. Soc. Am. **89**, 830–837.

Formby, C., and Muir, K. (**1989**). "Effects of randomizing signal level and duration on temporal gap detection," Audiology **28**, 250–257.

Formby, C., Sherlock, L. P., and Forrest, T. G. (**1996**). "An asymmetric roex filter model for describing detection of silent temporal gaps in sinusoidal markers," Aud. Neurosci. **3**, 1–20.

Formby, C., Sherlock, L. P., and Li, S. (**1998a**). "Temporal gap detection measured with multiple sinusoidal markers: Effects of marker number, frequency, and temporal position," J. Acoust. Soc. Am. **104**, 984–998.

Formby, C., Gerber, M. J., Sherlock, L. P., and Magder, L. S. (**1998b**). "Evidence for an across-frequency, between-channel process in asymptotic monaural temporal gap detection," J. Acoust. Soc. Am. **103**, 3554–3560.

Grose, J. H., Hall, J. W., III, Buss, E., and Hatch, D. (**2001**). "Gap detection

for similar and dissimilar gap markers," J. Acoust. Soc. Am. **109**, 1587–1595.
Haubert, N., and Pichora-Fuller, M. K. (**1999**). "The perception of spoken language by elderly listeners: Contribution of auditory temporal processes," Can. Acoust. **27**, 96–97.
Lawrence, M., and Yantis, P. A. (**1956**). "Onset and growth of aural harmonics in the overloaded ear," J. Acoust. Soc. Am. **28**, 852–858.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Lister, J., Besing, J., and Koehnke, J. (**2002**). "Effects of age and frequency disparity on gap duration discrimination," J. Acoust. Soc. Am. **111**, 2793–2800.
Lister, J., Koehnke, J., and Besing, J. (**2000**). "Binaural gap duration discrimination in listeners with impaired hearing and normal hearing," Ear Hear. **21**, 141–150.
Lutman, M. E. (**1991**). "Degradations in frequency and temporal resolution with age and their impact on speech identification," Acta Oto-Laryngol. **476**, 120–126.
McFadden, D., and Yama, M. F. (**1983**). "Upward shifts in the masking pattern with increasing masker intensity," J. Acoust. Soc. Am. **74**, 1185–1189.
Moore, B. C. J., and Glasberg, B. R. (**1988**). "Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears," J. Acoust. Soc. Am. **83**, 1093–1101.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1992**). "Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss," J. Acoust. Soc. Am. **92**, 1923–1932.
Moore, B. C. J., Glasberg, B. R., Donaldson, E., McPherson, T., and Plack, C. J. (**1989**). "Detection of temporal gaps in sinusoids by normally hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **85**, 1266–1275.
Nizami, L. (**2003**). "Afferent response parameters derived from postmasker probe-detection thresholds: 'the decay of sensation' revisited," Hear. Res. **175**, 14–35.
Penner, M. J. (**1977**). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," J. Acoust. Soc. Am. **61**, 552–557.
Phillips, D. P., Hall, S. E., Harrington, I. A., and Taylor, T. L. (**1998**). "'Central' auditory gap detection: A spatial case," J. Acoust. Soc. Am. **103**, 2064–2068.
Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). "Detection of silent intervals between noises activating different perceptual channels: Some properties of 'central' auditory gap detection," J. Acoust. Soc. Am. **101**, 3694–3705.
Pichora-Fuller, M. K. (**2003**). "Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension," Int. J. Audiol. **42**, S59–S67.
Pichora-Fuller, K., Schneider, B. A., Benson, N., Hamstra, S., and Storzer, E. (**2006**). "Effect of age on detection of gaps in speech and non-speech markers varying in duration and spectral symmetry," J. Acoust. Soc. Am. **119**, 1143–1155.
Plomp, R. (**1964**). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**, 277–282.
Roberts, R. A., and Lister, J. L. (**2004**). "Effects of age and hearing loss on gap detection and the precedence effect: Broadband stimuli," J. Speech Lang. Hear. Res. **47**, 965–978.
Schneider, B. A., and Hamstra, S. J. (**1999**). "Gap detection thresholds as a function of tonal duration for younger and older listeners," J. Acoust. Soc. Am. **106**, 371–380.
Schneider, B. A., Speranza, F., and Pichora-Fuller, M. K. (**1998**). "Age-related changes in temporal resolution: Envelope and intensity effects," Can. J. Exp. Psychol. **52**, 184–190.
Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**, 980–991.
Shailer, M. J., and Moore, B. C. J. (**1987**). "Gap detection and the auditory filter—Phase effects using sinusoidal stimuli," J. Acoust. Soc. Am. **81**, 1110–1117.
Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.
Steinberg, J. C. (**1935**). "Discussion from the point of the physicist," Ann. Otol. Rhinol. Laryngol. **44**, 819–822.
Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.
Wegel, R. L., and Lane, C. E. (**1924**). "The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear," Phys. Rev. **23**, 266–285.

Monaural speech segregation using synthetic speech signals

Douglas S. Brungart, Nandini Iyer,[a] and Brian D. Simpson
Air Force Research Laboratory, Wright-Patterson AFB, Ohio 45433-7901

(Received 7 March 2005; revised 6 January 2006; accepted 9 January 2006)

When listening to natural speech, listeners are fairly adept at using cues such as pitch, vocal tract length, prosody, and level differences to extract a target speech signal from an interfering speech masker. However, little is known about the cues that listeners might use to segregate synthetic speech signals that retain the intelligibility characteristics of speech but lack many of the features that listeners normally use to segregate competing talkers. In this experiment, intelligibility was measured in a diotic listening task that required the segregation of two simultaneously presented synthetic sentences. Three types of synthetic signals were created: (1) sine-wave speech (SWS); (2) modulated noise-band speech (MNB); and (3) modulated sine-band speech (MSB). The listeners performed worse for all three types of synthetic signals than they did with natural speech signals, particularly at low signal-to-noise ratio (SNR) values. Of the three synthetic signals, the results indicate that SWS signals preserve more of the voice characteristics used for speech segregation than MNB and MSB signals. These findings have implications for cochlear implant users, who rely on signals very similar to MNB speech and thus are likely to have difficulty understanding speech in cocktail-party listening environments. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2170030]

PACS number(s): 43.66.Pn, 43.66.Rq, 43.71.Gv [RAL] Pages: 2327–2333

I. INTRODUCTION

In the study of speech perception, it is often insightful to conduct listening experiments with abstract "speech-like" auditory stimuli that reproduce the intelligible qualities of natural speech with only a few of the acoustic features normally present in the human voice. Several studies have measured intelligibility of these synthetic signals in order to understand the perceptual salience and relevance of different acoustic features present in natural speech (Remez *et al.*, 1981; Shannon *et al.*, 1995; Smith, Delgutte, and Oxenham, 2002). While it is evident that synthetic speech is intelligible even when acoustic features are radically reduced (e.g., sine-wave speech), relatively little is known about the robustness of this intelligibility when the synthetic speech is presented in the presence of a second competing speech signal. In order to gain a better understanding of the cues that enable listeners to segregate speech signals in natural environments, it is helpful to study the extent to which speech segregation can be achieved with the same residual cues that result in good intelligibility in isolated synthetic speech signals.

Three synthetic signals have been most commonly studied in the literature: The first is sine-wave speech (SWS), which is composed of a small number of sinusoids that are amplitude and frequency modulated to track the first few formants (usually 3–4) of a normal speech signal (Remez *et al.*, 1994). In contrast to commercial speech synthesizers, which attempt to realistically replicate the acoustic elements in a natural speech utterance, SWS is designed to produce a minimal representation of speech that discards all acoustic features except the time-varying vocal tract resonances. SWS signals lack the broadband formant peaks and harmonic fine structure that are believed to form the basis of speech perception. Yet, despite the unnatural characteristics of SWS, listeners are still able to extract phonetic and syntactic information from these highly impoverished sentences.

Two more types of abstract speech-like stimuli can be created by artificially reproducing the frequency-dependent envelope modulations that occur in normal speech (Shannon *et al.*, 1995). The first type, which we refer to as modulated noise-band (MNB) speech, is constructed by using a filter bank to divide the original speech signal into frequency subbands, extracting the envelope from each sub-band, and using these envelopes to modulate the corresponding bands of a filter-bank-processed broadband noise source. The resulting speech signal captures the resonance information associated with the changing shape of the vocal tract, but lacks the periodic fine structure normally present in the voiced segments of natural speech. MNB signals are acoustically and perceptually similar to natural whispered speech, so their study has some ecological relevance in realistic listening environments. Additionally, MNB signals reproduce the cues that are provided through cochlear implants and have been extensively used in cochlear implant simulation studies; thus, their study is clinically relevant.

Another type of modulated-band synthetic speech is modulated sine-band (MSB) speech, which is constructed in the same way as MNB speech except that the individual sub-band envelopes are used to modulate logarithmically spaced sinusoids located at the center frequencies of each sub-band. Although the nonharmonic tonal structure of MSB speech results in a much less natural-sounding voice than the breathy whisper usually associated with MNB speech, both MNB and MSB speech signals have been shown to produce similar levels of intelligibility, ranging from near-chance performance for signals with only a single sub-band to near-

[a]Electronic mail: nandini.iyer@wpafb.af.mil

perfect performance for signals containing five or more subbands (Dorman, Loizou, and Rainey, 1997; Brungart *et al.*, 2005). While research has shown these two signals to be virtually identical in terms of intelligibility, recent findings point to better talker and sex identification with MSB speech (Gonzalez and Oliver, 2005). This difference in performance, which was attributed to improved modulation detection and retention of some periodicity information in the speech signals for MSB signals, may have some impact on the ability to segregate voices on the basis of talker characteristics, especially those based on the sexes of the talkers.

Listening in multitalker environments has been extensively studied with natural speech signals, where the intelligibility of a target talker is often degraded by the presence of one or more competing talkers (Brungart, 2001; Brungart *et al.*, 2001). Target extraction in such situations has been widely researched and has been commonly called the "cocktail party" problem (Cherry, 1953). Presumably, difficulties in target extraction occur due to the existence of two types of masking: energetic and informational. Energetic masking occurs at the auditory periphery due to the presence of overlapping frequency bands of target and interfering signals. As a consequence of the overlap, the representation of the target signal is somewhat degraded at higher auditory centers. With informational masking, the degradation results from the inability of the listener to separate a target signal from a masker due to factors such as target-masker similarity and target and/or masker uncertainty (Arbogast, Mason, and Kidd, 2002; Brungart and Simpson, 2002; Kidd *et al.*, 2003; Durlach *et al.*, 2003). In multitalker listening situations, informational masking effects are generally most prevalent when the target and masking voices are qualitatively similar, and least prevalent when the competing talkers are easily distinguishable due to differences in their apparent spatial locations (Cherry, 1953; Freyman, Balakrishnan, and Helfer, 2001), their vocal characteristics (Assmann and Summerfield, 1990; Brungart *et al.*, 2001) and, in some cases, their relative levels (Egan, Carterette, and Thwing, 1954; Carhart, Tillman, and Greetis, 1969; Brungart, 2001). Because abstract synthetic speech signals lack many of the acoustic cues that could ordinarily be used to distinguish between the different competing voices in an acoustic mixture, most notably voicing information, there is reason to believe that the effects of informational masking might be greater for multitalker stimuli consisting of synthetic SWS, MNB, or MSB utterances than they are for those for natural speech. Furthermore, because these synthetic speech signals lack much of the redundant acoustic information present in natural speech, they might also preclude the use of other strategies that listeners adopt to segregate natural voices, such as the selective focusing of attention on the quieter of the two talkers in an acoustic mixture (Brungart, 2001). Thus, there is reason to expect listeners to perform much worse in multitalker listening experiments with abstract speech-like stimuli than in those with natural speech stimuli.

To this point, relatively little effort has been made to compare multitalker listening performance across different types of synthetic speech signals. There has, however, been a handful of studies that have directly or indirectly measured cocktail-party listening with a *single* type of synthetic speech [MNB signals—Qin and Oxenham (2003) and Stickney *et al.* (2004)], motivated primarily by the assumption that the MNB speech is perceived by normal-hearing listeners in approximately the same way that normal speech is perceived by the users of cochlear implant devices (Wilson *et al.*, 1991). Both of these studies found that multitalker MNB speech signals produced substantially lower overall intelligibility scores than multitalker natural speech signals. Both also reported that natural speech interferers produced less overall masking than speech-shaped noise interferers for a natural speech target, but that MNB speech interferers produced *as much or more* masking than noise interferers for an MNB speech target. This result strongly suggests that MNB speech produces more informational masking than normal speech, presumably because of the loss of distinct acoustic features in the competing voices. There has also been a handful of studies examining talker identification and target segregation with SWS. Cues to sex and talker identification, for instance, are remarkably preserved in SWS replicas of natural speech tokens (Fellowes, Remez, and Rubin, 1997; Sheffert *et al.*, 2002). This suggests that SWS signals might be relatively easy to segregate on the basis of target and masker talker characteristics. However, another study examined the perception of multitalker speech signals consisting of SWS stimuli (Barker and Cooke, 1999), and found that listeners were much less able to segregate two male SWS voices than normal voices. Again, this presumably occurred due to the loss of distinctive voice characteristics in the SWS voices. To this point, no studies have examined segregation of SWS signals as a function of target-masker voice similarity.

The main aim of the present study is to more fully examine the effects that MNB, MSB, and SWS processing have on a listener's ability to segregate the competing talkers in an acoustic mixture. Evidence from previous research suggests that the acoustic cues for sex and talker identification are retained in synthetic signals even in the absence of fine spectral details (Fellowes *et al.*, 1997; Sheffert *et al.*, 2002; Qin and Oxenham, 2003; Stickney *et al.*, 2004); however, there is no direct evidence to suggest if these cues are better retained in one type of synthetic signal over the other. Of particular interest is the extent to which the distinctive voice characteristics (such as pitch, vocal tract size, speaking style) that help listeners differentiate between the competing voices in a multitalker stimulus are preserved in SWS, MNB, and MSB stimuli. Previous research investigating segregation of multiple talkers has used the Coordinate Response Measure (CRM) (Bolia *et al.*, 2000), a call-sign-based speech intelligibility test (Brungart, 2001; Brungart *et al.*, 2001). Intelligibility in the CRM task has been assessed for a target phrase in the presence of syntactically similar masking phrases, with the acoustic similarity of the target and masking voices as a variable. In general, the results indicate that target intelligibility is best when the target was qualitatively different from the masker (e.g., in conditions when the target and masker were different sexes) and worst when they were qualitatively similar (e.g., when the target talker was also the masking talker). A recent study by Darwin, Brungart, and Simpson (2003) investigated vocal characteristics that could explain

the target segregation based on voice differences. Three voice characteristics could account for most of the results in Brungart (2001). The first characteristic, fundamental frequency, presumably would be absent from the voiceless SWS, MNB, and MSB synthetic speech stimuli. However, the second two characteristics, vocal tract length and speaking style (for example, rate of speech), should be preserved to some extent in the synthetic speech. The next section describes the procedures used to conduct an experiment examining how well these residual differences can be used to segregate competing SWS, MNB, and MSB voices.

II. METHODS

A. Listeners

Eight normal-hearing listeners (three males, five females), with ages ranging from 21 to 55 years, participated in the study. All had audiometric thresholds of 20 dB HL or less at octave frequencies between 250 and 8000 Hz. The listeners were all native speakers of American English and were paid for their participation in the study. The listeners had participated in similar listening experiments with similar signals and were well-practiced in the experimental task.

B. Signals

The speech materials used for the experiment were synthesized from recordings of the Coordinate Response Measure (CRM) corpus (Bolia *et al.*, 2000), which contains phrases spoken by four male and four female talkers. Each phrase has the following structure: "Ready [call sign], go to [color], [number] now," with the possibility of eight different call signs (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, and Tiger), four colors (red, white, blue, and green) and eight different numbers (1 to 8). Thus, 256 different phrases are available for each talker, with a total of 2048 phrases in the whole corpus. Each phrase in the corpus was digitally preprocessed in order to produce three additional kinds of speech signals: (1) sine-wave speech [SWS]; (2) modulated sine-band speech [MSB]; and (3) modulated noise-band speech [MNB]. The processing procedures, which were identical to those used in a recent study of dichotic synthetic speech perception (Brungart *et al.*, 2005), are briefly summarized here. The SWS signals were generated by using the MATLAB scripts provided on the Internet by Ellis (2003). The signals in the CRM corpus were downsampled to 8 kHz in order to extract the first four formant frequencies; the formant frequencies and magnitudes were estimated in 2.6-ms windows using Linear Predictive Coding (LPC). The resulting signals were then resynthesized and resampled to 50 kHz. The MSB signals were created using a procedure developed by Arbogast *et al.* (2002). This procedure first downsampled the CRM phrases to 20 kHz and divided them into sub-bands with 15 logarithmically spaced bandpass filters with center frequencies from 215 to 4891 Hz (fourth-order, 1/3-octave Butterworth filters). The amplitude envelopes of each band were then extracted by half-wave rectifying and low-pass filtering the outputs of these bandpass filters. The resulting envelopes were then used to modulate a pure tone with the same frequency as the center frequency of the corresponding bandpass filter. The 15 bands of the envelope-modulated pure tones were then summed together to produce the MSB stimuli. The MNB signals were generated using an identical procedure, but with bandpass-filtered noises as carriers instead of pure tones.

Each stimulus presentation consisted of two simultaneous phrases from the modified CRM corpus. The target phrases were designated by the presence of the call sign "Baron," whereas the masker phrases contained different randomly selected call signs. The masker and target phrases also contained different color and number coordinates. Each target sentence, generated in one of three ways described above, was paired with a masker sentence generated in the same way. In addition, the similarity of the target and masking voices was varied in three ways, similar to previous experiments with multitalker speech intelligibility tasks (Brungart, 2001; Brungart *et al.*, 2001). The masker and target sentences were paired so that the masking phrase was spoken by the same talker as the target phrase (TT condition), a different same-sex talker (TS condition), or a different-sex talker (TD condition). For comparison, a normal speech control condition was also included that paired natural target and masker sentences from the original unprocessed CRM corpus. In all cases, the overall rms level of target signal was varied relative to the rms level of the masker to produce five different signal-to-noise ratios (SNRs): −8, −4, 0, +4, and +8 dB.

C. Procedure

On each trial, the signals were generated by a sound card (Soundblaster Audigy) and presented diotically over headphones (Beyerdynamic DT 990) to listeners seated in front of a computer screen in a quiet room. In a single-interval, forced-choice identification task, listeners were required to identify the color and number spoken by the target talker (cued by the call sign "Baron"). A four-row, eight-column display on the computer screen, corresponding to the four colors and eight numbers, allowed listeners to respond by using a computer mouse. Feedback was provided at the end of each trial, and percent-correct feedback was provided at the end of every block. There were three experimental variables: Type of speech signals (normal speech signals from the CRM corpus, SWS, MSB, and MNB), talker configuration (TT, TS, or TD), and SNR ($-8, -4, 0, +4, +8$ dB). For the MSB, MNB, and normal speech signals, the type of speech signal and the talker configuration were constant within each 80-trial block, and the SNR varied from trial to trial. For the SWS speech signals, the TT configuration was fixed within a block of trials, but the TS and TD configurations were randomly presented within the same blocks of trials. The order of presentation of the blocks was randomized across the listeners. The data collected from each listener represented the average from three blocks of trials, and the overall data reported in the study represent the average of 48 trials per subject at each SNR in each condition.

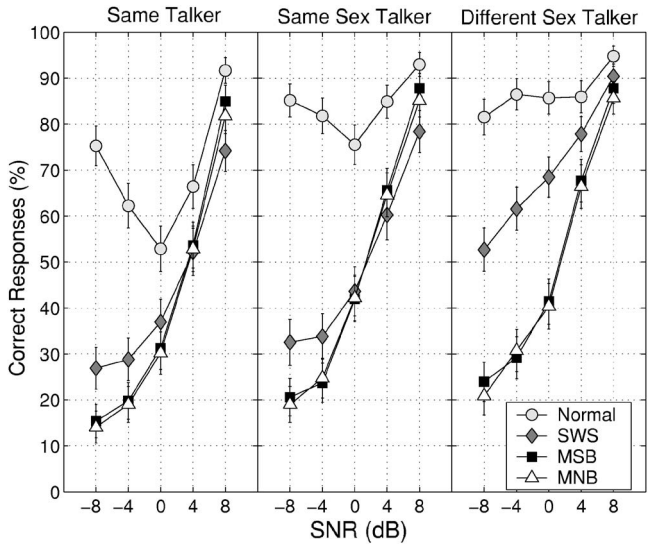


FIG. 1. Color and number identification performance in a diotic listening experiment with one target talker and one masking talker. The curves are plotted as a function of the signal-to-noise ratio (SNR) between the target and masking talker and talker configuration (columns). The four curves show performance for the normal speech signal and three different types of synthetic signals (see the legend). The error bars represent the 95% confidence interval for each data point.

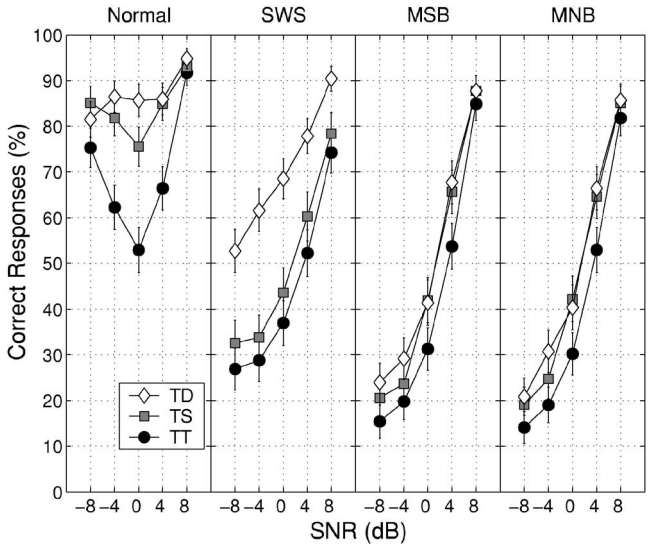


FIG. 2. Percentage-correct color and number identification in a diotic listening experiment with one target talker and one masking talker. The curves are plotted as a function of the SNR between the target and masking talker and the types of signals (columns). The three curves show performance for a different masking talker configuration (see legend). The error bars represent the 95% confidence interval for each data point.

III. RESULTS

Figure 1 shows the percentage of correct color and number responses as a function of the SNR for each of the three types of talker configuration. The circles show average percentage-correct color and number identification in the normal speech conditions, whereas the diamonds represent the performance with SWS signals. The squares and the triangles represent the performance with the MSB and MNB signals, respectively. Although this study did not explicitly measure target intelligibility in the absence of maskers (i.e., at an infinite SNR value), previous results have shown that all three kinds of synthetic speech produce near 100% intelligibility when presented in isolation (Brungart *et al.*, 2005). As expected, the results in Fig. 1 show that performance was consistently best with the normal speech. This findings was verified by submitting the arcsine-transformed percentage-correct data for each subject to a three-factor within-subject analysis of variance (ANOVA), with talker configuration (TT, TS, and TD), SNR (−8 to +8 dB), and type of processing (normal speech, SWS, MNB, and MSB) serving as the three factors. The analysis indicated that all three types of synthetic speech signals resulted in significantly poorer performance than normal speech [$F_{(3,21)}=185.8$, $p<0.01$].

The performance advantage of normal speech over the three types of synthetic speech was particularly large at negative SNR values, where the normal speech exhibited a leveling off in performance in the TD condition and an actual increase in performance in the TS and TD conditions due to the listener's ability to use a level cue to help distinguish the quieter target talker from the more intense masking voice (Brungart, 2001; Brungart *et al.*, 2001). These nonmonotonicities were conspicuously absent from the performance curves for the synthetic speech signals, indicating that listeners were not able to selectively attend to the quieter talker in the stimulus when synthetic speech signals were used instead of natural speech.

Relative performance in the three synthetic speech conditions depended on the SNR value of the stimulus. At positive SNR values, performance was slightly better in the MNB and MSB conditions than in the SWS condition. At negative SNR values, the SWS condition was somewhat better than the MNB and MSB conditions. This interaction between processing type and SNR was probably the result of the greater spectral overlap between the target and masking signals in the MNB and MSB conditions. This spectral overlap would cause the MSB and MNB speech signals to more effectively mask out the interfering speech signal at positive SNR values (thus reducing the informational masking caused by the interfering speech and improving performance). Similarly, the increased overlap would cause the interfering speech to more effectively mask out the target speech at negative SNR values (thus increasing energetic masking and decreasing performance). The spectral overlap produced by the different types of synthetic speech signals is discussed in more detail in Sec. IV. No difference in performance was found between the MSB and MNB conditions at any SNR value, suggesting that the slightly higher talker and sex identification that has recently been reported for MSB speech (Gonzalez and Oliver, 2005) does not result in a significant improvement in speech segregation.

Figure 2 replots the data from Fig. 1 in order to make it easier to compare the effects of target-masker similarity for each type of synthetic speech. The circles show performance for the conditions in which the target and masking talkers were the same (TT condition). The squares represent performance for conditions when the target and masking talkers were the same sex (TS), whereas the triangles show performance when target and masking talkers were different sexes

(TD). Of particular interest is the finding that talker configuration was important as a segregation cue for SWS signals, but was only a marginally effective segregation cue for MNB and MSB signals. This significant interaction between talker configuration and type of processing [$F_{(6,42)}=6.0$, $p<0.01$] was further probed with a Tukey test (at $p<0.05$), which confirmed that there were significant differences in performance across the TT, TS, and TD conditions for SWS and normal speech, but not for MNB and MSB speech. Thus, it seems that the distinct speaker-dependent voice characteristics that listeners use to segregate competing speech signals, and particularly those that listeners use to distinguish male and female voices, are more faithfully preserved in SWS speech than in MSB or MNB speech. In the case of MNB speech, it is worthwhile to note that this result is consistent with the results of Qin and Oxenham (2003) as well as Stickney *et al.* (2004), despite differences in the methodologies and speech materials used in those studies. Note that the former study showed that sex differences in the voices could not be used to improve the segregation of speech signals even when the MNB signals contained 24 bands, suggesting that the similar performance obtained in the TS and TD configurations was not limited by the 15-band resolution used in this MSB and MNB stimuli in this study.

IV. DISCUSSION AND CONCLUSIONS

In this experiment, listeners exposed to natural speech signals were able to take advantage of talker-dependent differences in voice characteristics to substantially improve their performance in a two-talker target identification task. They were also able to perform reasonably well at SNR values down to −8 dB, presumably because they were able to focus their attention on the quieter of two talkers in the stimulus. When the normal speech signals were replaced with SWS speech, performance was much worse overall, and listeners seemed to lose the ability to attend to the quieter talker in the stimulus. However, they were still able to substantially benefit from differences in the voice characteristics of the talkers, particularly when the target and masking talkers were different in sex. In the SWS condition, listeners showed a significant improvement in percentage-correct scores when a different-sex masking talker was used, and even a marginal improvement with a same-sex masker compared to a same-talker masker. This finding reflects the fact that the amplitude-modulated sine waves tracked the formant-frequency variations of the target and masker signals, and retained the pitch differences that listeners were able to utilize in the target identification task. Although sine-wave speech does not have a *pitch* in the traditional sense, an earlier study by Fellowes *et al.* (1997) has shown that sine-wave speech retains enough information to distinguish male and female speakers, and our findings indicate this information is also sufficient to allow listeners to distinguish between the male and female talkers in an SWS stimulus.

Given the nature of the signal processing for the three types of synthetic speech, the improved segregation obtained with SWS signals was not surprising. The difference between the MSB/MNB and SWS signals is that the formant information in the former two types is smeared by a 1/3-octave filter, whereas the formant values in the SWS speech are preserved exactly. Thus, sex differences in vocal tract length (on the order of 40%) are somewhat less distinct in the MSB and MNB speech than in SWS speech. In the MNB and MSB conditions, the listeners performed slightly better than with SWS speech at positive SNR values, but performed substantially worse than with SWS at negative SNR values. However, in contrast to the SWS condition, the listeners in the MNB and MSB conditions received, at best, only a marginal benefit from the differences in voice characteristics between different same-sex talkers and no apparent additional benefit from the much larger differences in voice characteristics between different-sex talkers. This seems to indicate that the coarse spectral differences between the masker and target signals provided by MSB and MNB speech are not adequate to code pitch differences between target and masking voices. This finding is particularly relevant for listeners using cochlear implants, because it suggests that these listeners might not have access to the primary segregation cues that allow normal listeners to separate a target speech signal from masking speech.

Perhaps the most compelling outcome of this study is the finding that listeners who could quite easily use level differences to selectively attend to the quieter of two same-sex or same-talker natural speech signals were consistently unable to use level cues to segregate the quieter of two synthetic voices. One possible explanation for this result is that the synthetically generated target and masking speech signals had greater spectral overlap than the natural speech signals. This increased spectral overlap might have led to an increase in energetic masking at low SNR values that overwhelmed any decrease in informational masking that could be obtained by an apparent level difference between the two talkers. In order to test this hypothesis, an acoustic analysis was conducted to compare the amount of spectral overlap that occurred between the target and masking voices for each of the four types of speech. First, the MATLAB {*specgram* command was used to calculate the amount of energy present in each 3.2 ms by 20-Hz time-frequency (*T-F*) region of the target and masking speech signals. Then, the local SNR value was calculated for each *T-F* region to determine which regions had a local SNR value of at least +3 dB, presumably indicating regions where the target signal would not be significantly distorted by the presence of the masker. These regions were then used to determine the percentage of the total energy in the target speech signal that was contained in *T-F* regions with local SNR values greater than 3 dB. The results of this analysis are shown in Fig. 3. These results clearly show that the MSB and MNB conditions did produce more spectral overlap than the normal speech conditions at negative SNRs, which can probably be attributed to the spectral smearing that the 1/3-octave band filters used to produce these signals caused in the frequency-domain features of the target and masking signals. Thus, it seems that energetic masking effects can, in part, account for the relatively poor performance that occurred at negative SNR values with the MSB and MNB speech. However, it is also apparent that the SWS condition produced significantly less spectral overlap

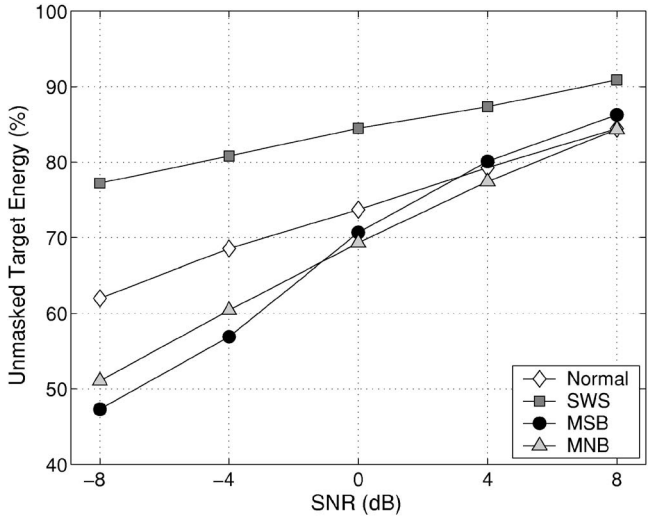


FIG. 3. Spectral overlap between the target and masking speech signals as a function of SNR for the target utterance "Ready Baron go to Blue Two Now," and the masker utterance "Ready Arrow go to Green Seven Now" for the same talker (Talker 1 in the CRM corpus). The ordinate shows the percentage of the total energy present in the target speech signal that was contained in time-frequency regions where the local SNR value was greater than 3 dB. See the text for details.

than the normal speech condition (which is perhaps not surprising considering that the SWS speech signals only contained energy at four discrete frequencies at any given point in time). Thus, spectral overlap cannot explain why the listeners were unable to use level cues to segregate the SWS speech at negative SNR values. It is also worth noting that the lack of voicing information may have significantly contributed to the poor performance that occurred at low SNR values in all three synthetic conditions. Previous research using synthetic as well as natural signals has verified that differences in fundamental frequency of voicing can help listeners listen to one or the other voice (Brokx and Nooteboom, 1982; Culling and Darwin, 1993). Indeed, Darwin (1981) has posited that fundamental frequency played two important roles in segregating sound sources: (1) to group consecutive segments into the speech from one talker and (2) to group the harmonics from one talker to the exclusion of other talkers. Results from de Cheveigne (1993) have suggested another role of fundamental frequency in segregation of concurrent harmonic sounds: to group together components of the masker so as to cancel them effectively. The lack of harmonic voicing information in SWS, MSB, and MNB synthetic speech signals may prevent listeners from exploiting these harmonic cancellation cues. This might also help explain why listeners have a great deal of difficulty extracting information from the quieter of two synthetic voices.

Although the listeners were able to extract some useful pitch information from the SWS, performance in all cases was much worse with synthetic speech than with natural speech. This was particularly true at low SNR values, where listeners are believed to rely on level difference cues for the segregation of natural speech signals. Further research will be needed to determine if it is possible to generate a sparse synthetic representation of speech that can be segregated on the basis of apparent level differences. If this kind of representation could be appropriately coded for direct electrical stimulation of the inner ear, it could lead to improved segregation performance for the users of cochlear implant devices.

Arbogast, T., Mason, C., and Kidd, G. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different undamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.

Barker, J., and Cooke, M. P. (**1999**). "Is the sine-wave speech cocktail party worth attending?" Speech Commun. **27**, 159–174.

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

Brokx, J., and Nooteboom, S. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.

Brungart, D. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D., and Simpson, B. (**2002**). "Within-channel and across-channel interference in the cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

Brungart, D., Simpson, B., Ericson, M., and Scott, K. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Brungart, D., Simpson, B., Darwin, C., Arbogast, T., and Kidd, G. (**2005**). "Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**, 292–304.

Carhart, R., Tillman, T., and Greetis, E. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

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

Culling, J., and Darwin, C. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by Fo," J. Acoust. Soc. Am. **93**, 3454–3467.

Darwin, C. (**1981**). "Perceptual grouping of speech components differing in fundamental frequency and onset-time," Q. J. Exp. Psychol. **33**, 185–207.

Darwin, C., Brungart, D., and Simpson, B. (**2003**). "Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers," J. Acoust. Soc. Am. **114**, 2913–2922.

de Cheveigne, A. (**1993**). "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model of auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290.

Dorman, M., Loizou, P., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Durlach, N., Mason, C., Kidd, G., Arbogast, T., Colburn, H., and Shinn-Cunningham, B. (**2003**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Egan, J., Carterette, E., and Thwing, E. (**1954**). "Factors affecting multichannel listening," J. Acoust. Soc. Am. **26**, 774–782.

Ellis, D. (**2003**). *Sinewave speech analysis/synthesis in* MATLAB (http://www.ee.columbia.edu/dpwe/resources/matlab/sws/).

Fellowes, J. M., Remez, R. E., and Rubin, P. E. (**1997**). "Perceiving the sex and identity of a talker without natural vocal timbre," Percept. Psychophys. **59**, 839–849.

Freyman, R., Balakrishnan, U., and Helfer, K. (**2001**). "Spatial release from informational masking in speech recognition," J. Acoust. Soc. Am. **109**, 2112–2122.

Gonzalez, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. **118**, 461–470.

Kidd, G. J., Mason, C., Arbogast, T., Brungart, D., and Simpson, B. (**2003**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Remez, R., Rubin, P., Pisoni, D., and Carrell, T. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–950.
Remez, R. E., Rubin, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (**1994**). "On the perceptual organization of speech," Psychol. Rev. **101**, 129–156.
Shannon, R. V., Zeng, F., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., and Remez, R. E. (**2002**). "Learning to recognize talkers from natural, sinewave, and reversed speech samples," J. Exp. Psychol. Hum. Percept. Perform. **28**, 1447–1469.
Smith, Z., Delgutte, B., and Oxenham, A. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.
Stickney, P., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
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.

A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties

Alicia Beckford Wassink[a)]
Department of Linguistics, University of Washington, Seattle Washington 98195

(Received 14 March 2005; revised 9 December 2005; accepted 2 January 2006)

A geometrical method for computing overlap between vowel distributions, the spectral overlap assessment metric (SOAM), is applied to an investigation of spectral (F1, F2) and temporal (duration) relations in three different types of systems: one claimed to exhibit primary quality (American English), one primary quantity (Jamaican Creole), and one about which no claims have been made (Jamaican English). Shapes, orientations, and proximities of pairs of vowel distributions involved in phonological oppositions are modeled using best-fit ellipses (in F1 × F2 space) and ellipsoids (F1 × F2 × duration). Overlap fractions computed for each pair suggest that spectral and temporal features interact differently in the three varieties and oppositions. Under a two-dimensional analysis, two of three American English oppositions show no overlap; the third shows partial overlap. All Jamaican Creole oppositions exhibit complete overlap when F1 and F2 alone are modeled, but no or partial overlap with incorporation of a factor for duration. Jamaican English three-dimensional overlap fractions resemble two-dimensional results for American English. A multidimensional analysis tool such as SOAM appears to provide a more objective basis for simultaneously investigating spectral and temporal relations within vowel systems. Normalization methods and the SOAM method are described in an extended appendix. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2168414]

PACS number(s): 43.70.Jt [AL] Pages: 2334–2350

I. INTRODUCTION

This paper deals with one classification problem resulting from systematic sources of variability affecting acoustic vowel features—the classification of languages as length- versus quality-contrasting. A broad range of linguistic and nonlinguistic effects would seem to render impossible the identification of numerical cutoffs (durational ranges and concomitant formant averages) which might allow for direct categorization of vowel systems based upon whether temporal or spectral features play a primary or secondary role. In this paper, spectral/temporal relations are investigated in three linguistic varieties that have been claimed to differ with regard to the role of these features in phonological contrast. We consider how the *overlap fraction* yielded by the spectral overlap assessment metric, or SOAM (described below), may help to address the question of whether or how vowel quantity contributes to the distinction (in production) between phonologically tense and lax vowels in three linguistic varieties: Jamaican Creole (JC hereafter), Jamaican English (JE), and American English (Pacific Northwest variety; PNWEng hereafter).

Broad variation exists in the possible distribution and combination of acoustic features (e.g., vowel formants and segmental durations) resulting from language-specific coarticulatory effects (e.g., vowel-to-vowel coarticulation: Macken, 1980; Strange and Bohn, 1998; Perkell and Matthies, 1992), nonlinguistic influences such as physiological differences (e.g., vocal tract length differences between sexes, adults and children: Nordström and Lindblom, 1975), and sociolinguistic factors (e.g., dialectal variation and linguistic change: Labov, 1994; Hagiwara, 1997). This variation, though many of its causes yield systematic effects, has not yet been integrated into a unified theory of influences on acoustic vowel features. Researchers still lack acoustic characterizations of many vowel systems. Also lacking are models of vowel systems that enable the examination of the combined influences of sets of effects on vowel features. This paper particularly addresses the latter issue. We apply SOAM to a cross-linguistic investigation of three vowel systems. SOAM allows the simultaneous comparison of spectral and temporal vowel features. Statistical properties of normalized estimates of the first two acoustic vowel formants (F1, F2) and segmental duration provide the basis for ellipse and ellipsoidal models of vowels in phonological oppositions. The shapes of the two vowel distributions and their proximity in two- and three-dimensional space (i.e., F1 × F2 and F1 × F2 × duration) are calculated. Graphical representation then allows for the visual evaluation of algorithm results.

II. BACKGROUND: SPECTRAL/TEMPORAL INTERACTIONS

A good deal of research attention has been devoted to the typological classification of vowel systems. Increasingly, acoustic phonetic analyses are undertaken to support phonological classifications (i.e., into types of contrast) based upon spectral and temporal vowel features. *Primary quality* has been used to classify languages in which spectral contrasts,

[a)]Electronic mail: wassink@u.washington.edu

 0001-4966/2006/119(4)/2334/17/$22.50

typically carried in the first two formants (F1 and F2), provide the basis for phonological contrast. The designations *primary* and *secondary quantity* are used to classify languages within which temporal features (typically segmental duration) play a crucial role. In the former, phonological contrast in a vowel system is based upon "robust" differences in the segmental durations of vowels in phonological oppositions. In the latter, phonological contrast is based upon robust differences in the quality of vowels in phonological oppositions while systematic differences in vowel length are also observed. One concern has been to understand what relations between vowel quality and quantity features must obtain in order to establish a basis for distinguishing between the two temporal designations. Certain inconsistencies occur in the literature regarding the definitions given to these three classifications. Namely, systematic differences in vowel length may result from several sources, some language-general and others language-specific. However, lengthening due to different sources is often not distinguished. For example, systematic differences in vowel length may result from (1) a language-specific phonetic tendency to "enhance" quality (i.e., phonemic) distinctions as in the tense/lax contrast in English (Lehiste, 1970) or (2) language-general (phonetic) processes, such as preconsonantal "lengthening," which supports the perception of consonant voicing. As an example of the former, the English tense/lax ratio is underlyingly 1.2:1—i.e., the average duration of short vowels is 83% of that of the long vowels. This represents a language-specific tendency related to phonological length. This ratio increases to 1.5:1 (67%) when the vowel precedes a voiced obstruent—a language-general tendency unrelated to phonological contrast (Lehiste, 1970). Kluender *et al.* (1988) suggest that true length contrasts (i.e., in primary quantity languages) may be so auditorily distinctive "that the contrast-enhancing effects of vowel-length variation are simply unnecessary." Thus, difficulty arises when classifications are used that do not distinguish between the phonetic outcomes of phonological length and processes of phonetic lengthening unrelated to phonological length. Additionally, classifications are often made without reference to spectral differences, e.g., whether a systematic difference in vowel duration is accompanied by a concomitant absence of quality overlap.[1] Crucially, as has been pointed out by Lehiste (1970, p. 35), a linguistic rationale (either auditory or acoustic) is lacking for determining how large a temporal contrast must be before it may be considered *primary*. Furthermore, it is even less clear what type of spectral distinction must obtain simultaneously with a given durational contrast. These two problems may be illustrated with examples related to perspectives on the role of *overlap in spectral features* and *overlap in temporal features*.

In the spectral domain, the lack of experimental research that uses a multidimensional approach in investigating the nature of spectral and temporal distributional patterns has impeded scholarly understanding of prototypical versus "ambiguous" vowel qualities. Hillenbrand *et al.* (1995) carry out a replication of the most widely cited study of the acoustics and perception of vowels (Peterson and Barney, 1952; P&B hereafter). They analyze spectral (f0, F1–F3) and temporal (segmental duration) information for speakers of a Midwestern dialect of American English and find differences between mean segmental durations for males' as compared to both children's and females' productions (males' durations being shortest). These results are statistically significant in *post hoc* comparisons, but the authors are unable to provide an explanation for this effect (Hillenbrand *et al.*, 1995, p. 3102). Their intention was to obtain, for General American English, a set of data comparable to that original study but more precisely controlled in terms of several parameters, including regular interval measures to trace the trajectory of the vowel, inclusion of measures for segmental duration, and controls for dialect of both speaker and listener dialect.

One key difficulty that reduces comparability between P&B and Hillenbrand *et al.* lies in the fact that the latter collected data from speakers of a Midwestern dialect. The authors were not unaware of dialect differences, which is clear in the fact that they attempted to screen for certain differences, excluding those who merged /a, ɔ/, or whose speech evidenced a "departure from general American English" (Hillenbrand *et al.*, 1995, p. 3100). However, although it is true that this dialect region is well known for not (yet) participating in the widespread merger of /a, ɔ/, it is not clear how meaningful it is to say that they also omitted speakers native to the area who were not speakers of "General American English."[2] There are differences between the productions of the final Midwestern sample and the P&B sample that reflect further spectral differences for which the authors cannot account. For example, sociophonetic descriptions of the vowel system for this region indicate marked raising of /æ/ that distinguishes the Midwest from other regions (Hagiwara, 1997; Labov, 1994), with women and children leading in this change. (Also, since the merger of /a, ɔ/ is widespread among dialects, the lack of merger already places this region outside the norm with respect to a "general American" feature). The Midwestern speakers' systems do indeed show raising of /æ/. For example, for the Midwestern women (the group with the *greatest* differences in mean F1, F2), Hillenbrand *et al.* report the following category means for /æ/ and /ɛ/: in F1—669 and 731 Hz (a difference of 62 Hz) and in F2—2349 and 2058 Hz (a difference of 291 Hz).[3] These differences in F1 and F2 are smaller than the mean differences reported for these formants in the earlier Peterson and Barney study (1952, p. 183), suggesting greater proximity between the /æ/ and /ɛ/ vowel categories of Midwestern speakers.[4]

This proximity appears to be relevant to the perception study results as well. Hillenbrand *et al.* report that listeners showed greatest confusion for /æ, ɛ, ɔ/. This is possibly the result of spectral overlap between some tokens of /æ/-/ɛ/ (which is perhaps most marked in males' tokens, as will be discussed for this complex situation, below). For /ɔ/, it seems likely that some speakers did lower or merge the quality of this vowel toward [a], so that while /a/ tokens remain unambiguous, lowered tokens of /ɔ/ were confusible with /a/. However, spectral overlap in the presence of systematic and auditorally salient temporal differences could still cue a distinction. In fact, there is evidence that temporal features aid in distinction.

First, Hillenbrand *et al.* report that a quadratic discriminant analysis was most successful when duration was included in the set of parameters used to classify tokens. Second, reexamination of production results suggests that temporal features indeed play a crucial role. Close examination of their Fig. 9, which shows plots of F3-F2 (mels) by F0-F1 (mels) of formant patterns sampled at 20% and 80% of the duration of the vowel, is quite telling. This figure shows longer curves for /æ/ than for /ɛ/ (for all groups) and for women than men.[5] Longer curves indicate greater spectral change over time, a temporal feature of vowel production. Specifically, women's /æ/ exhibited greatest change in the dimension of vowel height.

The duration difference between groups that Hillenbrand *et al.* found problematic may thus be understood as follows: females' longer durations are consistent with their greater advancement in the change by which /æ/ is produced as [iɛ]. This vowel is (for women and children) diphthongal, while still relatively phonetically static for males. Males' /æ/ and /ɛ/ are spectrally and temporally most extensively overlapping of the three Midwestern groups. Plots of midpoint values alone (as the authors acknowledge) mask crucial temporal distinctions. A representation of spectral and temporal features together is most helpful. Thus, knowledge of the phonetic extent of dialect differences is important in designing a replication and in interpreting study results. However, the more crucial point for the present study is that this difference suggests systematic dialectal differences in the relations between spectral and temporal cues to a phonological contrast.

Secondly, cross-linguistic comparisons of vowel length contrasts provide an illustration of how the lack of experimental research that uses a multidimensional approach has impeded scholarly understanding of the role of durational differences in phonological contrast. Japanese, Luganda, Icelandic, and Thai have been characterized as unambiguous examples of primary quantity languages on the basis of the typical durational differences between phonologically long and short vowels (conventionally expressed as a ratio of the duration of the phonologically long member of a pair of vowels to the phonologically short one). That is, vowel durations for contrasts in these languages typically fall above 1.6:1, while English and German are characterized as having secondary quantity (Lehiste, 1970). This length ratio is often accompanied by an impressionistic determination about spectral contrast (full overlap, partial overlap, or no overlap), which is unquantified in terms of phonetic characteristics of the vowels. Measures of F1 and F2 for the distributions are not typically presented. The conventional means of presentation is summarized in Table I.

Both of these problems demonstrate the need for a mathematical procedure that provides a means of simultaneously representing and evaluating the roles of spectral and temporal acoustic phonetic vowel features. Such a measure would have several advantages for between-language and within-language comparisons. Among these, it would facilitate categorization of primary and secondary quantity languages and provide a means of determining, within a language, whether a systematic difference in vowel duration is accompanied by a concomitant absence of quality overlap. An experiment was carried out in which spectral/temporal relations were tested for long-short vowel pairs in three language varieties: a Pacific Northwest variety of American English (referred to hereafter as PNWEng), Jamaican Creole (JC), and Jamaican English (JE). One longstanding question in the phonology of Jamaican varieties has been whether spectral contrasts or temporal ones are the basis for phonological vowel contrast (Wassink, 2001). Jamaican English (the term used in this paper to refer the Jamaican "acrolect") and Jamaican Creole (the Jamaican "basilect") represent distant ends of the post-Creole linguistic continuum in Jamaica. These two varieties are of interest because they differ systematically in the distribution of spectral (F1 and F2) and temporal (duration) features in their vowel inventories (Wassink, 1999, 2001), such that Jamaican Creole has been characterized as retaining more of the phonetic features of its West African adstrates (which utilized phonemic vowel length), and Jamaican English of its English adstrates (several of which utilized secondary quantity). Thus, researchers are interested in understanding the relative contributions of the phonologically different input languages to the Jamaican continuum. The experiment that follows reports the results of the within-language characterizations, followed by a between language-comparison which allows some observations to be made about the application of the terms primary quantity, primary quality, and secondary quantity to these varieties.

TABLE I. Languages reported to utilize vowel length (from Crothers, 1978).

Language	Long: short ratio	Spectral overlap	Primary/secondary quantity
Japanese	2.5:1	Full	Primary
Thai	1.9:1	Full	Primary
Icelandic	2.0:1	Partial	Primary
Luganda	2.5:1	Partial	Primary
German	1.5:1	Partial	Secondary
English	1.2:1	Partial	Secondary
(arbitrary cutoff)	1.6:1		

III. EXPERIMENT

A. Speakers

The Pacific Northwest sample included six female speakers born and raised in Washington state. The Jamaican English sample included nine speakers (five male and four female) native to the Kingston Corporate Area (St. Andrew Parish), and the Jamaican Creole sample included ten speakers (five male and five female) native to the rural parish of St. Thomas. None reported a knowledge of hearing difficulties.

B. Materials

Target data for this study include six vowels representing three phonological oppositions, namely /i:/ *beat* ~ /ɪ/ *bit*, /a:/ *bought*~ /a/ *bat*, /u:/ *boot*~ /u/ *book*. Here, the symbols /i:,i,a:,a,u:,u/ are used to provide a consistent, convenient means of representing the word classes indicated and

to signal the phonological specification of the lengths of the vowels in each and not the phonetic quality of vowels in any of the language varieties under study. The phonetic qualities differ, as will be discussed below, and phonetic qualities will be represented using square brackets "[ ]". Vowels from these six classes were elicited for all speakers of the three linguistic varieties. Vowels representing all other monophthongal qualities were also elicited, but were used solely for purposes of normalization (see below). Tokens were collected in word-list elicitation tasks in language-appropriate carrier frames: "Write ___ please" for Jamaican English, "Unu rait ___ pon i" for Jamaican Creole, and "Say ___ again" for PNWEng. The carrier frames and token numbers vary for PNWEng and the two Jamaican varieties because data were originally elicited for two different investigations. For all language varieties, vowels were elicited in real-word monosyllables of the shape /b,d,k_p,b,t,d,k,g/, as well as /h_d/ for PNWEng.[6] Three hundred ninety-one tokens were analyzed for PNWEng (6 speakers × 6 vowels × 5 repetitions × a subset of the 10 contexts), 2362 tokens for Jamaican English (9 speakers × 6 vowels × ≤5 repetitions × 10 contexts), and 2437 tokens for Jamaican Creole (10 speakers × 6 vowels × ≤5 repetitions × 10 contexts).

All Jamaican English and Jamaican Creole tokens were collected in a field research setting, which facilitated the collection of data for two linguistic varieties not previously subjected to experimental phonetic study. However, environmental conditions had to be carefully controlled to ensure a high-fidelity signal for acoustic analysis. To this end, recordings were made with a close-talking AIWA lapelle microphone and Sony TCD-D8 digital audio tape recorder. Recordings of the PNWEng speakers were made in the Linguistic Phonetics Laboratory at the University of Washington, using a TASCAM DAP-11 and Nakamichi (CM-100) cardioid microphone. The speech signal was recorded at a 44.1 kHz sampling rate and digitally transferred for analysis in Praat software. Tokens were downsampled to 11.025 kHz for analysis.

The first three vowel formants (F1, F2, F3) were measured at the temporal midpoint for each vowel token. Vowel onset was determined conservatively from the waveform as the beginning of the first clear pitch pulse associated with modal voicing following the release of the preceding consonant. Vowel offset was defined as the end of the final clear pitch pulse before the closure of the following consonant. Vowel duration was automatically calculated by the software (Praat) as the difference between vowel onset and offset. F1, F2, and F3 were obtained using superimposed FFT and LPC spectra from a 25.6-ms window. FFT analyses used a 59-Hz filter; LPC spectra used 14 predictor coefficients. In an initial pass, formant measures were taken using an automatic peak-picking routine based on the LPC spectra. Then, all automatically calculated frequencies were checked in the combined FFT/LPC display. Questionable values were corrected or confirmed using a narrow-band spectrogram. To check the reliability of the measures, approximately 20% of the dataset was independently analyzed by a second phonetician. F1, F2, and F3 were obtained using Praat's autocorrelation-based procedure and the values generated by this procedure were manually inspected using Praat's formant trace function. Intermeasurer agreement was assessed at 92%.

Individual F1 and F2 values for all monophthongs, including (/i:, ɪ, ɛ, æ, (ɔ,) a:, u:, ʊ/[7] were normalized using an adaptation of Nearey's (1977) logarithmic uniform scaling technique (Wassink, 1999). This procedure is detailed in Appendix A. This method was preferred because it suppresses variation due to sex-related physiological differences, but appears to retain variation such as may be introduced by differences in linguistic quality (Adank *et al.*, 2004; Hindle, 1978). This vowel-extrinsic procedure achieves within-speaker normalization using information distributed over the formant frequencies for all monophthongs in that speaker's system.

C. Data analysis procedures: The spectral overlap assessment metric (SOAM)

Mathematical determination of the relationships between two vowel distributions is accomplished using one of two algorithms: for spectral features alone (SOAM 2-D, two-dimensional model) or between spectral and temporal vowel features (SOAM 3-D, three-dimensional model). Both algorithms take as input normalized values for individual vowel tokens. The two- and three-dimensional algorithms are summarized in Appendix B.

IV. RESULTS

In this section, the results of the cross-linguistic investigation of JE, JC, and PNWEng vowel features are presented. First, we examine the results yielded by a conventional F1 × F2 representation, using a Plot Formants-style representation (Ladefoged, 1996). Next, the results of overlap calculations generated in SOAM 2-D (F1 × F2) and SOAM 3-D (F1 × F2 × duration) are presented.

Table II presents the mean and standard deviation information for F1 and F2 (in Hz) for each vowel for each variety. Figures 1(a)–1(c) show conventional inverted F1 × F2 scatterplots of the normalized formant values calculated using these means. In Fig. 1(d), the raw data (in Hz) for PNWEng are presented, encircled by best-fit ellipses generated in PlotFormants (Ladefoged, 1996). The main benefit of the conventional method, i.e., of reporting mean and deviation information along with a supplemental inverted F1 × F2 plot of vowel data, is that it enables one to generally locate vowel qualities in particular regions of acoustic vowel space. For example, comparison of the scatter plots reveals a different overall shape for the Jamaican vowel spaces relative to that of PNWEng. While the high front pair /i:, i/ lie in roughly the same locations, the low vowels /a:, a/ are central in Jamaican varieties (the apex of a V-shaped vowel space), while the PNWEng low vowels occupy different regions entirely in that space: /a/ *bat* is phonetically [æ], a low front vowel, while /a:/ *bought* has a low back place of articulation.[8] Similarly, the high back pair /u:, u/ occupy different regions. This, however, is most easily determined from inspection of the mean F2 data in Table II. Mean F2 values for PNWEng /u:, u/ (1653.96 and 1480.17 Hz, respectively) are higher than

TABLE II. Mean and standard deviations for F1, F2, and duration for Jamaican English, Jamaican Creole, and PNWEng /i:, i, a:, a, u:, u/ (n =number of tokens).

	Jamaican English						
	n=2362	F1 (Hz)	Standard deviation	F2 (Hz)	Standard deviation	Duration (s)	Standard deviation
i:	493	305.04	*53.36*	2440.43	*237.29*	0.1271	*0.0308*
i	525	372.52	*62.05*	2219.33	*203.78*	0.0751	*0.0291*
a:	130	691.78	*87.24*	1179.93	*135.61*	0.1803	*0.0232*
a	527	755.97	*87.07*	1525.81	*197.90*	0.1200	*0.0297*
u:	588	338.31	58.46	986.00	*396.64*	0.1303	*0.0365*
u	99	365.99	*81.32*	858.42	*190.48*	0.0867	*0.0542*
	Jamaican Creole						
	n=2437	F1 (Hz)	Standard deviation	F2 (Hz)	Standard deviation	Duration (s)	Standard deviation
i:	511	309.77	*77.40*	2463.66	*251.38*	0.1386	*0.0317*
i	551	379.36	*63.08*	2277.18	*206.47*	0.0748	*0.0228*
a:	136	688.37	*95.52*	1231.58	*162.71*	0.1868	*0.0355*
a	565	728.56	*110.32*	1493.20	*198.78*	0.1174	*0.0278*
u:	567	365.94	*58.96*	834.33	*200.98*	0.2294	*1.9129*
u	107	421.12	*56.63*	892.17	*186.15*	0.0624	*0.0190*
	American English (PNW variety)						
	n=391	F1 (Hz)	Standard deviation	F2 (Hz)	Standard deviation	Duration (s)	Standard deviation
i:	56	382.75	*53.50*	2614.63	*365.08*	0.0843	*0.0231*
i	64	465.07	*50.76*	2177.36	*169.13*	0.1003	*0.0401*
a:	101	818.59	*128.91*	1366.01	*117.99*	0.1047	*0.0397*
a	66	837.61	*146.76*	1893.62	*168.13*	0.1015	*0.0343*
u:	68	400.30	*48.74*	1653.96	*345.66*	0.1399	*0.0425*
u	36	531.72	*62.54*	1480.17	*285.53*	0.1629	*0.0481*

the corresponding Jamaican English values (986.00 and 858.42 Hz, respectively) and the corresponding Jamaican Creole values (834.33 and 892.17 Hz, respectively). Thus, PNWEng high back vowels lie further to the front of the acoustic vowel space, consistent with reports by Hillenbrand *et al.* (1995) for the American Midwest. While it might seem reasonable to impute the difference in the overall dispersion of acoustic vowel space to the presence in the Jamaican samples of male speakers (whose high back vowels tend to have lower F1 and F2 frequencies than those of adult females), differences of the magnitude observed (for JC versus PNWEng /u:/=819.63 Hz; for /u/=588.00 Hz) are far greater than would be expected due to sex-related differences (Hindle, 1978). However, it is important to test this by examining the normalized values. Table III presents mean and standard deviation information for the log-mean normalized F1 and F2 values of all three language varieties. Log-mean normalized values for /u:/ ~ /u/ (averaged across all speakers within the language group) for Jamaican Creole are −0.116(log F1) and −0.262 (log F2); for PNWEng −0.289(log F1) and −0.157(log F2). These values are more difficult to interpret than values in raw Hz. They represent mean token deviations from the center of each language system's vowel space. Therefore, if both spaces are centered at (0,0), we see that the PNWEng /u:/ data tend to lie further front along the front-back dimension in that system: normalized F2 values tend to be smaller than for the Jamaican Creole /u:/. Normalized data provide the basis for the comparisons that follow, which examine, first, the extent of overlap of each vowel onto its acoustically adjacent partner in two and then in three dimensions.

Standard deviation information is of some utility in determining the proximity of vowel distributions. We may observe, for example, that the standard deviations calculated for Jamaican English /a:,a/ are sufficient to close the distance between the F1 means for these vowels, suggesting that these distributions are quite proximal, and possibly overlapping, at least in this dimension. We may independently observe that the standard deviations for this same pair of vowels indicate that some scatter lying near the edges (=2σ of the mean) of each distribution approximate data in the other distribution by up to ~135 Hz. Such comparisons are limited, however, because it is possible that two distributions that are proximal in one spectral dimension may yet be spectrally distinct, because they may show significant separation in the other dimension. The results of the spectral overlap assessment will therefore be enlightening.

Overlap in two dimensions. The leftmost columns listed for each language in Table IV provide the *overlap fractions* yielded by SOAM 2-D for each phonologically contrasting vowel pair. An overlap average, calculated across all pairs within a language variety, is provided in the final row, for convenience in comparing the three linguistic varieties. This value must be considered with some caution, however, because it gives only a global sense of systems that show internal variability. As the discussion above demonstrated for Midwestern American English *bag*, intrinsic and extrinsic factors (such as those operating in linguistic change) can

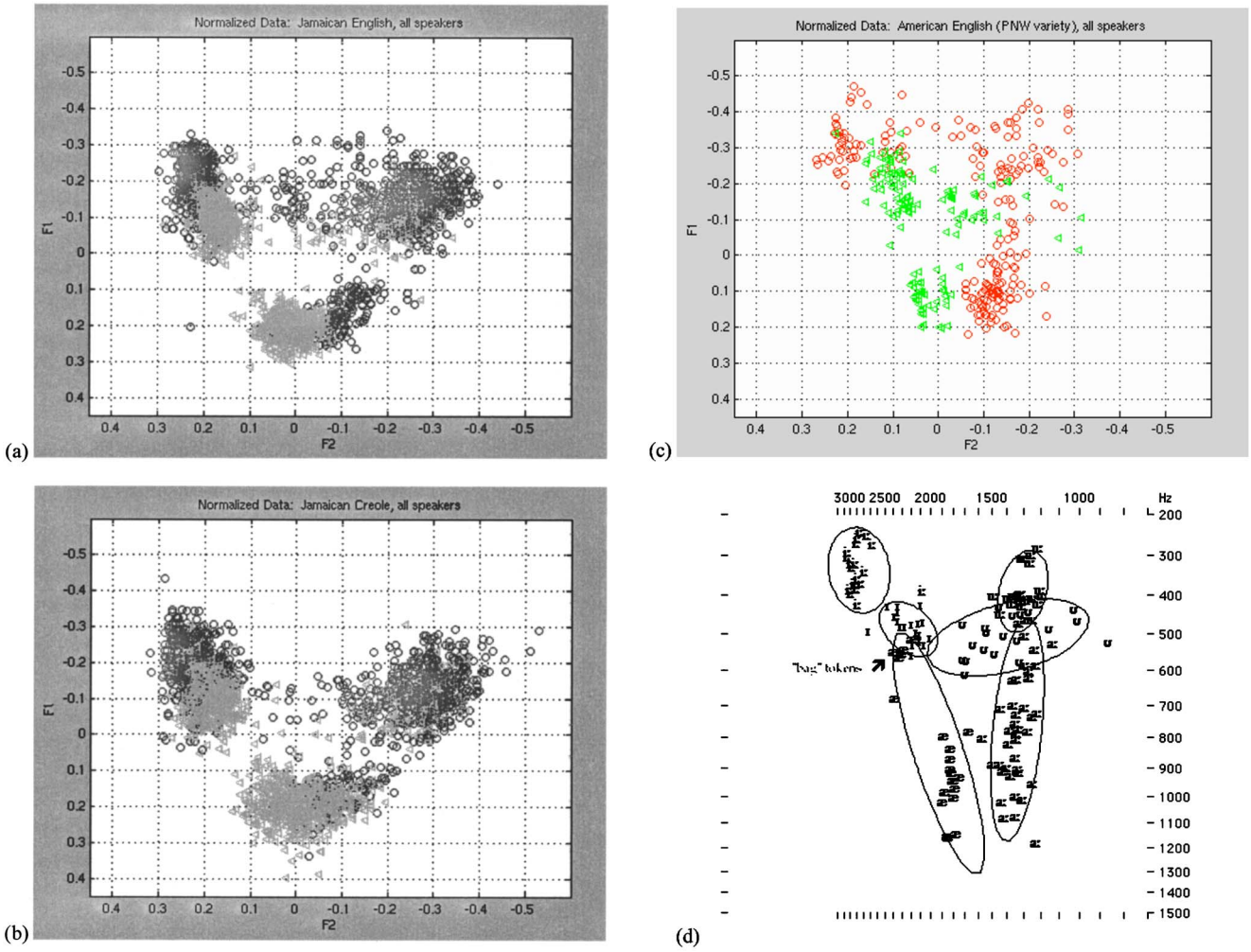


FIG. 1. (Color online) Scatterplots illustrating the log-mean normalized values for vowels in three tense-lax oppositions for three language varieties: (a) Jamaican English, (b) Jamaican Creole, and (c) PNWEng. Tense vowels /i:, a:, u:/ are represented with "∘," lax vowels /i, a, u/ with "◁". (d) is a second representation of the PNWEng data in (c), generated from raw F1, F2 values (in Hz) by Plot Formants software, "stats" function. Note vowel raising of several tokens of /a/ (phonetically [as]: specifically, those in the word "bag").

result in the displacement of one member of a pair. While it might be convenient to be able to accord classifications such as primary quantity or primary quality on the basis of the overall system, it cannot be expected that all pairs exist in the same phonetic relationships. For this reason, the discussion will treat each pair separately. For PNWEng, the vowel pair /i:∼i/ shows the greatest overlap, 46%. This result for the Pacific Northwest speakers is consistent with a visual inspection of Hillenbrand *et al.*'s findings for a group of Midwest American English speakers. The pair /u:∼u/ shows slightly less overlap, 34%, and /a:∼a/ the least, 15%. (Recall that /a/ represents the *cat* word class.) This vowel shows some raising in the Pacific Northwest vowel space as in the Midwestern one, however, raising is apparent for just one lexical item, namely "*bag*," which is widely realized by PNW speakers as [beg] or [beɪg]. When *bag* tokens are removed, the distribution mean for this class is still close enough to [æ] that the pattern observed above remains: that is, PNWEng /a:∼a/ show the least distributional crowding across all three varieties in a two-dimensional representation (reflecting linguistic changes that have taken place in varieties of American, but not Jamaican, English).

Overlap fraction values for Jamaican English appear to be similar to those for PNWEng. Percent overlap for the pair /i:∼i/ is 36% (10 percentage points less than PNWEng) and 23% for /a:∼a/ (3 points less). It is in the /u:∼u/ pair that Jamaican English values are most different from those of PNWEng: the overlap fraction for the former language variety is 75% (against 34% for PNWEng, a difference of 41 percentage points). Thus, whereas the high front and low pairs appear to be less coextensive in F1×F2 acoustic space, the high back pair shows greater overlap.

The Jamaican Creole oppositions received substantially higher overlap fraction values than either Jamaican English or PNWEng. All vowel pairs overlap in F1×F2 space by more than 55% of their total area: /i:∼i/—75%, /a:∼a/—55%, /u:∼u/—86%.

Overlap in three dimensions. As described below in Appendix B, SOAM 3-D includes a second least-squares fit calculation to transform the data in a manner that reflects their orientation in $x^* \times$ duration space. By thus incorporating duration into the model of our best-fit ellipsoids, we may recalculate overlap fractions and observe how segmental duration information influences the extent of protrusion of ad-

TABLE III. Log-mean normalized F1 and F2 values for three language varieties. (Tokens collected for the word bag have been omitted for the PNWEng speakers.)

Vowel	Jamaican English		Jamaican Creole		PNW English	
i:	$n=533$		$n=579$		$n=56$	
log F1 (*std dev*)	−0.198	*0.057*	−0.191	*0.076*	−0.311	*0.063*
log F2 (*std dev*)	0.214	*0.029*	0.218	*0.039*	0.176	*0.061*
i	$n=586$		$n=619$		$n=62$	
log F1 (*std dev*)	−0.096	*0.048*	−0.065	*0.106*	−0.209	*0.060*
log F2 (*std dev*)	0.169	*0.036*	0.150	*0.103*	0.102	*0.034*
a:	$n=132$		$n=136$		$n=97$	
log F1 (*stddev*)	0.163	*0.047*	0.166	*0.051*	0.085	*0.080*
log F2 (*std dev*)	−0.101	*0.048*	−0.083	*0.061*	−0.128	*0.036*
a	$n=559$		$n=601$		$n=47$	
log F1 (*std dev*)	0.205	*0.035*	0.191	*0.052*	0.117	*0.043*
log F2 (*std dev*)	0.009	*0.037*	−0.004	*0.058*	0.021	*0.029*
u:	$n=618$		$n=590$		$n=58$	
log F1 (*std dev*)	−0.148	*0.066*	−0.116	*0.066*	−0.289	*0.070*
log F2 (*std dev*)	−0.221	*0.109*	−0.262	*0.091*	−0.157	*0.084*
u	$n=99$		$n=107$		$n=36$	
log F1 (*std dev*)	−0.065	*0.050*	−0.073	*0.061*	−0.138	*0.058*
log F2 (*std dev*)	−0.209	*0.070*	−0.249	*0.076*	−0.102	*0.093*

jacent vowels. In general, we expect to find an inverse relationship between durational separation and overlap fraction. That is, as mean duration differences for a given vowel pair increase, the overlap fraction will decrease. The rightmost column presented below each language variety in Table IV provides the results for the three-dimensional overlap calculations (SOAM 3-D).

We may first observe that overall, overlap for all three language varieties decreases in the three-dimensional model. The overlap fraction decreases the least for PNWEng: by 19 percentage points for /i:∼i/, 26 for /a:∼a/, and 19 for /u:∼u/. The decrease, however, does indicate that incorporation of duration into the model pulls the members of each pair further apart. All oppositions now overlap by 27% or less (14% on average).

Overlap fractions for Jamaican English under the three-dimensional representation show a greater change than do those of PNWEng. Amount of overlap decreases are 27 percentage points for /i:∼i/, 21 for /a:∼a/, and 57 for the pair that showed greatest overlap in the two-dimensional model, /u:∼u/. These changes bring the Jamaican English oppositions into relationships where each pair overlaps, on average by only 10%.

Jamaican Creole overlap fractions show the greatest change from the two- to the three-dimensional model. Degree of overlap decreases are 58 percentage points for /i:∼i/, 32 for /a:∼a/, and 39 for /u:∼u/. These changes bring the Jamaican Creole oppositions into relationships where each pair protrudes, on average, by 29%.

Because the transformation of a two-dimensional to a three-dimensional overlap outcome is not linear, it is inappropriate to subtract the three-dimensional from the two-dimensional overlap fractions to determine the contribution of duration. At present, we must accomplish this impressionistically by manually rotating the three-dimensional graph and visually observing the extent of overlap from several angles. This is shown in Fig. 2, which provides three views of each /a:∼a/ model ellipsoid pair generated for each of the test language varieties. Unfilled circles represent datapoints contained within each model ellipsoid. For each language variety, proximity between pairs of model ellipsoids may be visually inspected from any number of angles. The top panel in each set of figures displays the default three-dimensional output of VOIS3D software. The second and third panels show rotations of the data as projected onto the F1×F2 and the F2×duration planes, respectively. It may be observed from the second panel in each set [Figs. 2(a), 2(b), 2(c)] that coordinates for /a:/ and /a/ coextend along a similar range of values in F1 for all three varieties: distributions for /a:/ are typically larger, while /a/ typically has a more compact distribution in F1. The PNWEng ellipsoids differ from the Jamaican ones in two respects. First, as may be observed in Fig. 2(c) model distributions do not overlap along the F2 dimension in PNWEng, consistent with the overlap percentages reported above. Second, it may be observed from Fig. 2(c) that duration values exhibit a similar range for both the /a:/ and /a/ word classes.

Comparison of the two- and three-dimensional overlap fractions supplied by SOAM 2-D and SOAM 3-D leads us to conclude that for the language varieties and dimensions investigated, duration is most critical for production contrasts in Jamaican Creole. That is, duration has a more notable reducing effect on overlap in Jamaican Creole than in the other varieties, and within this variety an effect of similar size obtains in all oppositions. Interestingly, although Jamaican English vowels already show little overlap in F1×F2 for two contrasts, overlap fractions decrease most substantially

TABLE IV. Overlap fractions calculated in two (by SOAM 2-D) and three (by SOAM 3-D) dimensions, F1×F2 and F1×F2×duration, respectively, for three language varieties.

	Jamaican English		Jamaican Creole		PNW English	
	F1×F2	F1×F2×Dur	F1×F2	F1×F2×Dur	F1×F2	F1×F2×Dur
i:∼i	36%	9%	75%	17%	46%	27%
a:∼a	23%	2%	55%	23%	15%	0%
u:∼u	75%	18%	86%	47%	34%	15%
Average	45%	10%	72%	29%	35%	14%

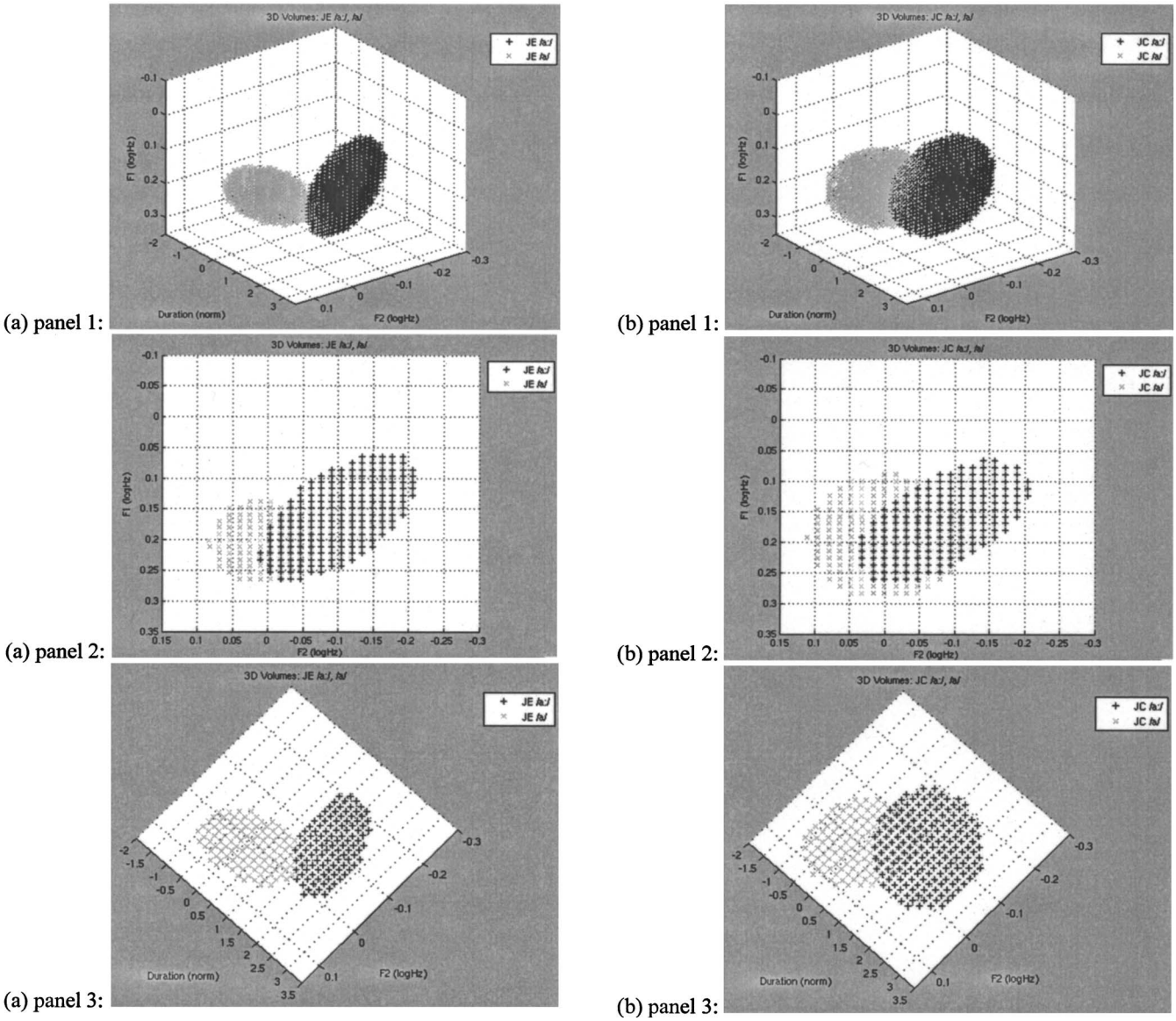


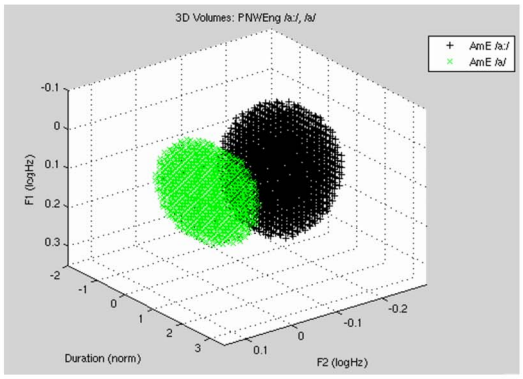
FIG. 2. Three-dimensional representations of /a: ~ a/ for (a) Jamaican English, (b) Jamaican Creole and (c) PNWEng data, generated using VOIS3D software. Panels 2 and 3 represent manual rotations of the 3-D volume graphs in panel 1 oriented in the F1 × F2 (panel 2) and F2 × duration (panel 3) planes.

for the high back pair (the magnitude of change in the overlap fraction from a two- to a three-dimensional model being most similar to Jamaican Creole). For PNWEng the overlap between volumes decreases, but less substantially—volumes that already shared only 26–47% of their volumes show an additional decrease of 14 percentage points. Thus, the most complex relation obtains for Jamaican English. It appears that duration contributes to a distinction already primarily carried along the spectral dimension (for high front and low central contrasts). In the high back pair, this enhancement is quite large. Listener identification might be tested to determine whether independent evidence may be found, in the perceptual domain, to suggest whether spectral or temporal features, or a complex relation between the two, serves as the primary basis for distinction. It appears to be important to investigate whether one parameter can be more crucial to contrast for one particular opposition than for others.

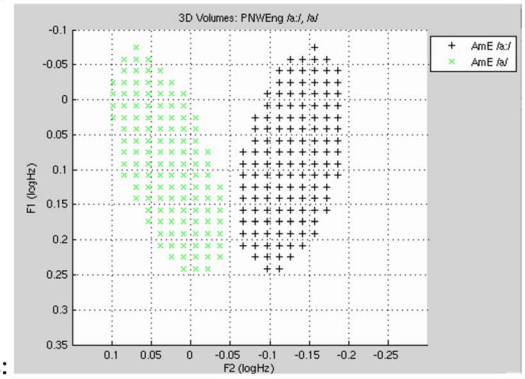
V. DISCUSSION

The findings presented above are illuminating in relation to the typological question of whether we might classify the three varieties investigated as showing primary quantity, secondary quantity, or primary quality. Ideally, overlap patterns will be investigated for a wide range of languages, and these findings are used to see how overlap patterns cluster cross-linguistically. This is preferred over the designation of arbitrary cutoffs. However, it may be instructive to assign a more principled (though still somewhat arbitrary) cutoff for purposes of informal comparison, as such data are not presently available. We may employ terms currently used in the literature for describing impressionistic relations in the quality dimension only, applying these instead to our multidimensional representations: complete overlap, partial overlap, and no overlap. We may provisionally define a fairly conservative cutoff for partial overlap as protrusion in less than half

(c) panel 1:



(c) panel 2:



(c) panel 3:

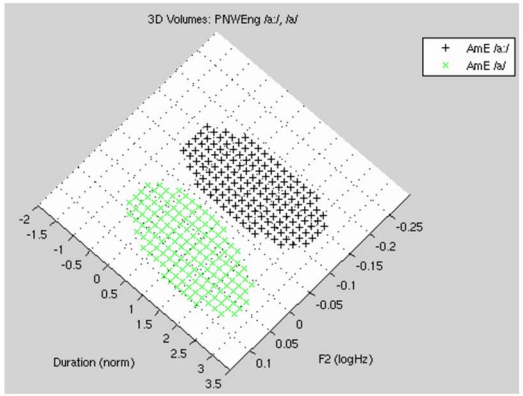


FIG. 2. (*Continued).*

of the volume of the shortest vowel in an acoustically adjacent pair (e,g., an overlap fraction of 20%–40%), one for no overlap as 0%–20%, and a third for complete overlap as ⩾40%. We may observe that once a factor is included for duration, all oppositions for all three language varieties must be classed as showing partial or no overlap, with the exception of Jamaican Creole /u:~u/, which is only slightly above the cutoff for partial overlap (47%). The Jamaican English oppositions and two of three PNWEng ones must be classed as showing no spectral overlap, while the third PNWEng opposition (/i:~i/) falls just above the cutoff for partial overlap. The Jamaican Creole oppositions would be classed as no overlap (/i:~i/), partial overlap (/a:~a/), and complete overlap (/u:~u/), respectively.

When we consider the magnitude of change within each language variety from a two- to a three-dimensional best-fit model, we find that in the initial two-dimensional analysis, two of the three PNWEng oppositions already show no overlap (low central and high back), and the third may be classified as just above the cutoff for partial overlap. All Jamaican Creole oppositions must be classed as exhibiting complete overlap when the model includes F1 and F2 alone. As mentioned, these relations are reduced to no or partial overlap with the incorporation of a factor for duration in /i:~i/ and /a:~a/, while /u:~u/ falls just above the cutoff for partial overlap. Jamaican English /i:~i/ and /a:~a/ are reduced from partial to no overlap when modeled in three dimensions, and /u:~u/ from complete to no overlap. Thus, once duration is accounted for, JE overlap fractions come to most closely resemble those of the two-dimensional results for PNWEng (10%, 14%; no overlap). The overall average for JC decreases from complete to partial overlap (29%).

VI. SUMMARY

We have considered an approach to vowel characterization that uses linear algebraic formulas and geometrical models to represent vowels as ellipses (with area) and ellipsoids (with volume). Two algorithms are used to calculate the area of overlap (for ellipses) or the shared regions within overall vowel volumes (for ellipsoids). The output of each algorithm is an overlap fraction. These models are statistical in that they are not directly based upon calculations of acoustic vowel features, but rather take as their input summary statistical information (formant and duration means and standard deviations) that is used to locate the center and define the spread of each vowel in acoustic vowel space. Two least-squares lines are used to orient each vowel relative to F1, F2, and duration axes. Best-fit ellipses (or ellipsoids) are then used for area (or volume) overlap calculations. This method provides a more objective basis for simultaneously investigating spectral and temporal relations within vowel systems.

These methods have been applied to an investigation of spectral/temporal relations in three linguistically different varieties. American English has traditionally been classified, on the basis of impressionistic evidence, as a secondary quantity language. This investigation has demonstrated that for a sample of speakers of a Pacific Northwestern variety of American English, the domains for three tense/lax vowel pairs may be considered to be partially overlapping when modeled in F1 and F2, and that incorporation of durational information in the model reveals systematic but small additional reductions in overlap. Previous research into Jamaican phonology has reported that Jamaican Creole distinguishes vowels only by duration (Akers, 1981). What has not been previously reported is the nature of vowel quality differences in Jamaican Creole in light of concomitant differences in vowel length. Differences between Jamaican Creole and Jamaican English have also gone undescribed. The present study suggests that spectral separation of vowels does occur for Jamaican English speakers (as revealed in the overlap fractions in the two-dimensional model), but these differences are not equivalent for all oppositions. For Jamaican Creole, measures of overlap in two parameters support the interpretation that complete spectral overlap obtains for all oppositions. For Jamaican English, it appears that duration

plays an enhancing role (for high front and low central contrasts) and serves as the primary basis for acoustic distinction in the high back pair.

The finding that vowel length may play a different role for different pairs of vowels within the same linguistic system suggests a more complex view of vowel system classifications than is typically taken in the linguistics literature. While it may be necessary from a phonological point of view to theorize that a particular abstract phonological feature such as vowel length operates at the level of the system (or subsystem) so that each contrast in the system bears a specification for that feature, it has been informative to see the differences between pairs of oppositions, revealed by the analysis above. While theories of phonology typically acknowledge that robust phonetic (defined broadly) differences may provide the basis of phonological contrast, it is well understood that phonetic surface features are complex and difficult to predict. Even when normalization algorithms have suppressed the effects of nonlinguistic factors (e.g., physiological differences between speakers), there are additional, systematic sources of variability that remain and must be explained. Because vowel-extrinsic (see Appendix A) normalization methods have been applied in the present analysis, the variability we observe in overlap fractions appears to be linguistic or sociolinguistic in nature. We have considered all oppositions separately (e.g., before duration is modeled, Jamaican Creole oppositions overlap by 75%, 55%, and 86%). We have also noted whether or not all three oppositions may be given the same classification (e.g., before duration is modeled, all Jamaican Creole oppositions fall above the arbitrary cutoff for complete spectral overlap). The latter perspective allows us to see that duration differences of a certain magnitude are maintained in all contrasts, which is useful in understanding forces possibly operating within an overall system. However, it should not be surprising that different vowel pairs stand in different relations (i.e., that different overlap fractions obtain for different pairs in the same system; e.g., compare Jamaican English /a:~a/, 23%, to /u:~u/, 75%). This finding is interpretable from the perspective of several phenomena well known to linguists, such as sound change. Historical phonological research has amply demonstrated that vowels within a system may begin a course of change one at a time (e.g., lowering of the nucleus of Middle English long-u, the vowel in *mouse*, "breaking" [u:] into diphthong [aʊ] in Early Modern English which was subsequently followed by raising of Middle English long-o). Such changes occur within the system in an orderly fashion, as so-called "chain shifts." These changes often disturb the symmetry of a vowel system, bringing one vowel (the one undergoing change in this case) into a closer spectral or temporal relation with a new neighbor but further from vowels in other parts of the system. Thus, surface phonetic features must be understood to reflect both diachronic (historical) and synchronic (sociolinguistic) variation, from one community to another and from one point in time to another.

There is a continued need within experimental phonetic research to critically consider the use of the ellipse and the ellipsoid to represent vowel distributions in acoustic space. This convention appears to exist out of convenience while having no basis in auditory or acoustic reality (Ladefoged, p.c.). We find that best-fit ellipses (using two standard deviations of the mean for calculation of ellipse vertices) superimposed upon the original values for vowel data typically fit the data closely, without appearing to leave large "gaps" within a distribution, however this has by no means been tested mathematically. It would, of course, be desirable to find optimal geometrical representations on a more principled, by-vowel basis. Such an approach might be illuminating with respect to different patterns in acoustic clustering between vowels, as well as possible auditory or acoustic motivations for this. Finally, further research would benefit from dynamic representations of time-varying vowel features. The present method, while enabling examination of spectral and temporal features in a multidimensional geometrical representation, is limited to use of spectral vowel measures from one temporal point. The next phase of this project involves seeking means of representing spectral features across the duration of the vowel.

Widespread continued use of the results of descriptive studies (such as those of Peterson and Barney, 1952; Hagiwara, 1997, etc.) underscores the ongoing need for research providing baselines for the features of vowel systems. However, as argued by Hagiwara (1997) and others, such studies have not tended to adequately account for dialectal variation, to say nothing of the effects of ongoing phonological change embedded within the speech community. Further studies are needed to provide current baselines, drawn on carefully designed samples, that reflect what is currently known about phonetic variation. We have argued here that descriptive applications such as the one undertaken in this article further require the transformation of data using a normalization process that minimally suppresses possible linguistic effects. Nearey's log-mean uniform scaling technique was used to achieve within-speaker normalization. The normalized values for each speaker constituted the input values (i.e., scatter) for between-vowel comparisons. Other sociolinguists have similarly found this method to be suitable for phonetic analysis of data in studies of the speech community (e.g., Labov, 1994; Adank *et al.*, 2004). For both sociolinguistic and phonetic research, integration of the methods, and increased attention to the results, of the other holds clear promise for the advancement of both disciplines.

ACKNOWLEDGMENTS

The author wishes to express gratitude to Jeremy Waltmunson for extension of the SOAM graphical user interface into the VOIS3D (Vowel Overlap Indication Software, in 3 Dimensions) software package, that enables visualization of SOAM output, and to Setsuko Shirai for assistance in researching normalization techniques. Development of the VOIS3D graphical user interface was supported by University of Washington Royalty Research Grant No. 65-2597.

APPENDIX A: NORMALIZATION METHODS USEFUL FOR OVERLAP CALCULATIONS

In the last 20 years, a good deal of discussion around the topic of normalization of the acoustic signal has been carried

in this journal (e.g., Lobanov, 1971; Nearey, 1977, 1989; Disner, 1980; Syrdal and Gopal, 1986; Miller, 1989). Other studies (e.g., Hindle, 1978; Syrdal, 1984; Deterding, 1990) have also demonstrated the need for normalization algorithms that allowed for the suppression of systematic, predictable effects of various types, such as the influence of sex-based differences on the fundamental frequency domain, due to physiological differences such as vocal tract length and circumference. Adank *et al.* (2004) evaluate the effectiveness of 11 vowel normalization procedures for application in language variation research. These procedures are classed into one of two types depending on the information they employ. Vowel-intrinsic normalization procedures make use of formant data for a single vowel token and yield normalized vowels for that token. Common methods include nonlinear transformations of the frequency scale (log, mel, bark). Vowel-extrinsic normalization procedures make use of (spectral) information about more than one vowel or about the speaker's vowel system to normalize the data for each vowel. For example, some vowel-extrinsic methods make use of the formant frequencies of the point vowels, while others make use of formant frequencies for all vowels in the system. Such methods often calculate a grand mean, representing the center of a speaker's vowel space.

The present study constitutes a language variation study of a different type than that presented in Adank *et al.*, however, we also find that vowel-extrinsic procedures are most appropriate. We are concerned here with modeling vowel distributions and assessing the relative contributions of spectral and temporal features to vowel distribution volume. This requires an approach to normalization and quantification that both suppresses the effects of unwanted sources of variability (physiological effects of speaker sex) and preserves extralinguistic dimensions of contrast (e.g., regional or cross-linguistic differences) as well as phonemic differences (i.e., vowel identity). This need may be reduced to two problems: a normalization problem and a multidimensional modeling problem. SOAM uses one of several "vowel-extrinsic" normalization methods to accomplish within-speaker normalization to address the first need. This process yields normalized formant and duration values for each speaker that may be used for between-speaker comparisons. Then, the multidimensional modeling problem is solved by using a geometric algorithm that represents normalized scatter for two vowel distributions as best-fit ellipsoids oriented at angles with respect to the F1, F2, and duration axes. The output of SOAM is an overlap fraction, a value that represents the volume of the region of overlap (the coordinates of normalized vowel space shared by both best-fit ellipsoids). Three vowel-extrinsic methods for spectral normalization and two methods for temporal normalization have to date been tested for use with SOAM. Since the primary purpose of the present study was to develop a method for modeling vowels in three dimensions and not to evaluate vowel normalization techniques, the discussion of normalization below takes as its point of departure the evaluations of normalization methods described in the literature. However, it will be necessary to describe limitations that were discovered while researching particular procedures, and the steps taken to address these. Formant-based methods tested for use in SOAM include log-mean normalization (Nearey, 1989), known-extremes normalization (Shirai, 2004), and a modified *Z*-score normalization. Duration-based methods include segmental duration *Z*-score and phrase duration *Z*-score. Methods are classed as "vowel-extrinsic" or "vowel-intrinsic," following Adank *et al.* (2004). The overlap method is described in Appendix B.

1. Methods for spectral normalization

I. *Log-mean normalization* (vowel-extrinsic).

This procedure, adapted by Wassink [1999; based on Nearey's (1977) Uniform Scaling technique; see also Disner, 1980], is appropriate for normalizing data for two vowels produced by a single speaker when the spectral locations (in an F1 × F2 acoustic space) of all vowels in the speaker's linguistic system are known. Nearey's method has become perhaps the most widely used method of normalization used in sociolinguistic studies of language variation because it minimally reduces linguistic and sociolinguistic variation while suppressing certain effects of physiological factors (Hindle, 1978; Labov, 2001).

In Nearey's normalization procedure Hertz frequency values for vowel formants are first log-transformed to better reflect the scale of sensitivity of the human ear to changes in frequency (Moore, 1989; Nearey, 1989). A second normalization step suppresses the effects of interspeaker variation. For example, males tend on average to have lower fundamental and lower formant frequencies than females, and their vowel spaces tend to be more compact overall than those of females. The log-transformed frequency G_{hijk} of a particular formant h (F1 or F2) occurring for a token i of a given vowel j for a particular speaker k is converted to a difference score F_{hijk}: the difference of that value from the mean of all values for that speaker for that formant, $\bar{G}_{hk}$ [Eq. (A1)], is

$$F_{hijk} = G_{hijk} - \bar{G}_{hk}. \tag{A1}$$

If the number of tokens is the same for all vowel categories in the speaker's system, the speaker grand mean $\bar{G}_{hk}$ (i.e., the F1 or F2 grand mean) is calculated over all realizations of that formant for that speaker, using Eq. (A2):

$$\bar{G}_{hk} = \frac{\left(\sum_{j=1}^{n} \sum_{i=1}^{n_j} G_{hijk}\right)}{nn_j}, \tag{A2}$$

where n is the number of vowels in the speaker's system and n_j is the number of data tokens for each vowel. This method, however, results in skewing of the grand mean for a given dimension away from the true center of speaker's vowel space when the number of vowels in any category is disproportionately represented in the sample. Therefore, it was necessary to adapt the method. If the number of tokens is different for different vowel categories, then $\bar{G}_{hk}$ is calculated as the mean of the category means for each formant for each vowel, as in Eq. (A3):

$$\bar{G}_{hk} = \frac{\sum_{j=1}^{n} \left(\sum_{i=1}^{n_j} G_{hijk}/n_j \right)}{n}, \quad \text{(A3)}$$

where n is the number of vowels in the speaker's system and n_j is the number of data tokens for vowel j.

II. *Known extremes normalization* (vowel extrinsic). This method was proposed by Shirai (2004) for normalization of spectral data for languages that have asymmetrical acoustic distributions, such as Japanese, where the so-called "point vowels" /i, a, u/ do not demarcate the edges of vowel space in F1 and F2. The "known extremes" method is also appropriate when locations are only known for a subset of the vowels in a speaker's system where these vowels demarcate the edges of the system. This method allows for the calculation of a grand mean for the system and, for this reason, may be used for cross-speaker comparisons (within one language). Unlike methods I and III, this method includes a formula for the determination of the vowel categories that represent the maxima and minima in F1 and F2. This method also differs from I and III in the method for calculating a grand mean for each speaker [Eq. (A4)]:

$$\bar{G}_{hk} = \left(\sum_{j=1}^{2} \left(\frac{\sum_{i=1}^{n_j} G_{hijk}}{n_j} \right) \right)/2, \quad \text{(A4)}$$

where $j=1$ is vowel at the lower extreme of the formant h for the speaker's vowel system, and $j=2$ is a vowel at the upper extreme of the formant h for the speaker's vowel system. Thus only two vowels are used in the calculation of the known extremes grand mean for a particular formant. Once the grand mean has been calculated in this way, the procedure for normalizing individual formant values follows as in Eq. (A1).

III. *Modified Z-score normalization* (vowel extrinsic). This method, introduced by Lobanov (1971), is appropriate when only a subset of the vowel locations is known (e.g., the language has five vowels, but we have data only for the two vowels whose overlap we desire to compare). Crucially, this method is only used for overlap calculations within a single speaker. If the number of tokens is the same for each vowel for which data are available, a formant grand mean $\bar{G}_{hk}$ (i.e., an F1 or F2 grand mean) for the subject is calculated using Eq. (A2) above, only that n is the number of vowels for which data are available. As with log mean normalization, because numbers of datapoints for each vowel category may be unequal, a vowel category mean is used to represent each vowel category rather than individual observed datapoints. In this case Eq. (A3) above may be used, again with n referring to the number of vowels for which data are available, rather than the total number of vowels in the speaker's system. Once again, the normalized formant values are calculated using Eq. (A1).

2. Methods for temporal normalization

IV. *Segmental duration Z-score normalization*. This procedure is similar to the one used for calculating Z scores for spectral data, except that the data are not log transformed. Here, segmental duration values are used. Calculate the mean duration for each vowel category (e.g., unique category means are calculated for /i/, /a/, /u/, etc.) A grand mean of duration, $\bar{D}_{o,k}$, for speaker k is calculated across all category durations in Eq. (A5):

$$\bar{D}_{o,k} = \frac{\sum_{j=1}^{n} \left(\sum_{i=1}^{n} D_{ijk}/n_j \right)}{n}, \quad \text{(A5)}$$

where D_{ijk} is the observed segment duration for token i of vowel j for speaker k. The normalized value of duration, δ_{ijk}, is equal to the observed value minus the category grand mean [Eq. (A6)]:

$$\delta_{ijk} = D_{ijk} - \bar{D}_{ok}. \quad \text{(A6)}$$

V. *Phrase-duration normalization*. Quite frequently, it is necessary to use phrase durations to normalize segmental durational values. For example, in experimental methods that rely upon use of reading lists, or other methods where rate of speech is not readily controlled, within- or between-speaker differences may arise in the durations due to fluctuation in speech or reading rate, rather than intrinsic differences in vowel durations. Ericsdotter and Ericsson (2001) and Simpson and Ericsdotter (2003) normalize duration of a vowel segment relative to the word or phonological phrase in which it was uttered. One adaptation of this idea is given in Eq. (A7),

$$D_{\text{norm}} = D_{\text{segment}}/D_{\text{phrase}}, \quad \text{(A7)}$$

where D_{norm} is the phrase-duration normalized segment duration, D_{segment} is the measured segment duration of a vowel token, and D_{phrase} is the duration of the carrier word or phrase for that token. This preprocessing constitutes an initial normalization step and is done before normalizing these durations to the system grand mean.

In summary, vowel-extrinsic procedures are always used for spectral and temporal normalization in vowel overlap calculations. This is because the majority of types of overlap comparisons that we might desire to compute appear to require scaling to a (system or language) grand mean. Furthermore, several vowel-extrinsic procedures incorporate nonlinear transformations of the frequency scale, for representing auditory sensitivity. For example, grand means (or uniform scaling) are desirable: for comparisons of overlap in two acoustically adjacent vowel categories in a single language where data are pooled across more than one speaker; for cross-linguistic comparisons, such as those reported above, in pairs of vowels present in the systems of both languages; for comparisons of the extent of overlap in vowel pairs for speakers of different (developmental) ages in studies of the development of phonemic contrast; and in investigations of the extent of loss of contrast (decrease in overlap) between two vowels under different speaking rates or registers. Even where we might wish to compute overlap between two vowels produced by a single speaker, we typically want to know how that speaker's overlap compares to that of other speakers (or to other vowels produced by the same speaker in different environmental conditions). Because calculation of the size of a vowel distribution (as compared to the spectral

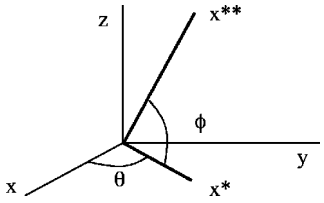


FIG. 3. Graphic of 3-D coordinate system, showing x, y, z, x^*, and x^{**} axes, angles θ, and ϕ.

location of a single formant) relies on distributional information (as described below), we need a method that will reflect distances between elements of the system.

APPENDIX B: AN ANALYTIC GEOMETRIC METHOD FOR QUANTIFYING VOWEL OVERLAP

1. Introduction

The spectral overlap assessment metric (SOAM) uses two algorithms for the calculation of the extent of overlap between pairs of vowels. Extent of overlap between two normalized vowel distributions is determined either by considering spectral features alone (two-dimensional overlap, SOAM 2-D) or by considering spectral and temporal features together (three-dimensional overlap, and SOAM 3-D). Each member of a pair of vowel distributions is modeled geometrically as an ellipse in F1 × F2 space or an ellipsoid in F1 × F2 × duration space. The ellipse or ellipsoid is best fit to the data scatter for the vowel using least-squares fitting to determine principal axes. Overlap between a pair of vowel distributions is calculated based on a fraction of uniformly distributed test points in the region of overlap relative to the number of test points in each vowel distribution.

Vowel distributions have been modeled in the past as ellipses in F1 × F2 space (Wakita, 1976; Ladefoged, 1996). In this study, a vowel distribution is modeled as a best-fit ellipse oriented at an angle with respect to the F1 × F2 axes or as an ellipsoid in F1 × F2 × duration space. Wassink (1999) described a method for assessing overlap between two vowel distributions modeled as ellipses based on the extent of protrusion of one ellipse into the other. Here, instead of a protrusion metric, area of overlap is calculated.

2. Calculation of vowel distribution overlap in three dimensions

a. Modeling two vowel distributions in three dimensions

In this case, spectral information is supplied for both vowels in each pairwise comparison for two dimensions (where F2 is the x axis and F1 is the y axis) while temporal information is supplied for the third (z axis). Prior to application of this method, data are normalized using one of the procedures described in Appendix A.

For convenience, plotting of data in three dimensions is accomplished using a polar coordinate system, as illustrated in Fig. 3. x, y, and z represent the axes associated with F2, F1, and duration dimensions, respectively.

Each of the vowel distributions may be modeled as an ellipsoid in F2 × F1 × duration space. Two vowel distributions may be defined as $\boldsymbol{v}_1=(\boldsymbol{x}_1,\boldsymbol{y}_1,z_1)$ and $\boldsymbol{v}_2=(\boldsymbol{x}_2,\boldsymbol{y}_2,\boldsymbol{z}_2)$ where $\boldsymbol{x}_j, \boldsymbol{y}_j$, and $\boldsymbol{z}_j$ are column vectors of F2, F1, and duration data, respectively, for vowel j, and (x_{ij}, y_{ij}, z_{ij}) are F2 × F1 × duration data triplets for a particular vowel token.

We begin by subtracting the mean values of F2, F1, and duration for vowel j, $\boldsymbol{x}_{o,j}=(x_{o,j}, y_{o,j}, z_{o,j})$, from each of the tokens in that distribution. Next, the orientation of the principal axis projected into the F2 × F1 plane is calculated through a least squares fit of a line to the F2 × F1 data. The angle, θ_j, between the $\boldsymbol{x}$ axis and the $\boldsymbol{x}_j^*$ axis is calculated from the slope of the least squares line [Eq. (B1)]

$$\theta_j = \tan^{-1}(k_{xy,j}). \tag{B1}$$

A rotation matrix $\boldsymbol{R}_{\theta j}$ is developed [Eq. (B2)] to rotate the data into a coordinate system aligned with $\boldsymbol{x}_j^*$:

$$\boldsymbol{R}_{\theta j} = \begin{bmatrix} \cos(-\theta_j) & -\sin(-\theta_j) & 0 \\ \sin(-\theta_j) & -\cos(-\theta_j) & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{B2}$$

The matrix containing translated values for each data point is multiplied by the above rotation matrix producing data in $(\boldsymbol{x}_j^*, \boldsymbol{y}_j^*, \boldsymbol{z}_j^*)$ coordinates [Eq. (B3)]:

$$\boldsymbol{v}_j^* = [\boldsymbol{x}_j - \boldsymbol{x}_{o,j}]\boldsymbol{R}_{\theta,j}. \tag{B3}$$

Once the rotation within the F2 × F1 plane has been carried out, a least squares line is fit to the $(\boldsymbol{x}_j^*, \boldsymbol{z}_j^*)$ data to get a second principal axis for the ellipsoid, oriented at an angle ϕ_j relative to the x_j^* axis [Eq. (B4)]. Finally, the $\boldsymbol{v}_j^*$ data are transformed to a coordinate system $\boldsymbol{v}_j^{**}=(\boldsymbol{x}_j^{**}, \boldsymbol{y}_j^{**}, \boldsymbol{z}_j^{**})$ aligned with this angle, using a second rotation matrix $\boldsymbol{R}_{\phi j}$ [Eqs. (B5) and (B6)]:

$$\phi_j = \tan^{-1}(k_{x^*z^*,j}), \tag{B4}$$

$$\boldsymbol{R}_{\phi j} = \begin{bmatrix} \cos(-\phi_j) & 0 & -\sin(-\phi_j) \\ 0 & 1 & 0 \\ \sin(-\phi_j) & 0 & \cos(-\phi_j) \end{bmatrix}, \tag{B5}$$

$$\boldsymbol{v}_j^{**} = \boldsymbol{v}_j^* \boldsymbol{R}_{\phi,j}, \tag{B6}$$

Figures 4(a) and 4(b) show the first vowel distribution prior to and following translation and rotation, while Figs. 5(a) and 5(b) show the second vowel distribution prior to and following translation and rotation.

Once the vowel distribution data are converted to a coordinate system aligned with the principal axes, it is simple to calculate the principal radii of the ellipsoid. The standard deviation of data along each axis (that is, in each of the coordinates) is calculated, and the principal radius along that axis is defined as m times the standard deviation in that direction. A value of $m=2.0$ is used. The formula for each ellipsoid is solved separately, using Eq. (B7):

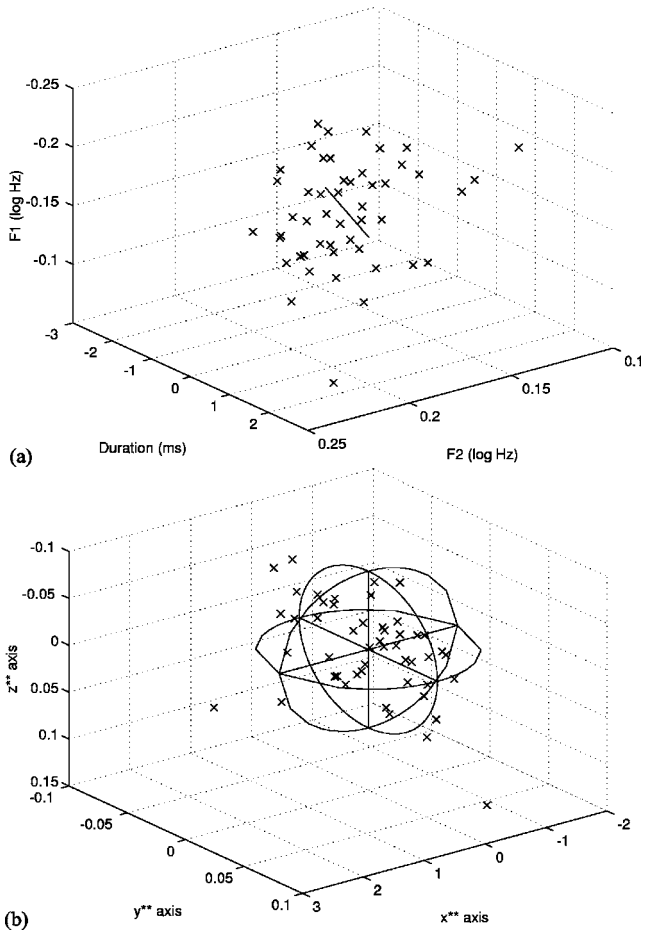


FIG. 4. Distribution of scatter for /i/ for speaker TX (a) before data translation and rotation and (b) after. The solid line in (a) represents the major axis defined by least-squares line.

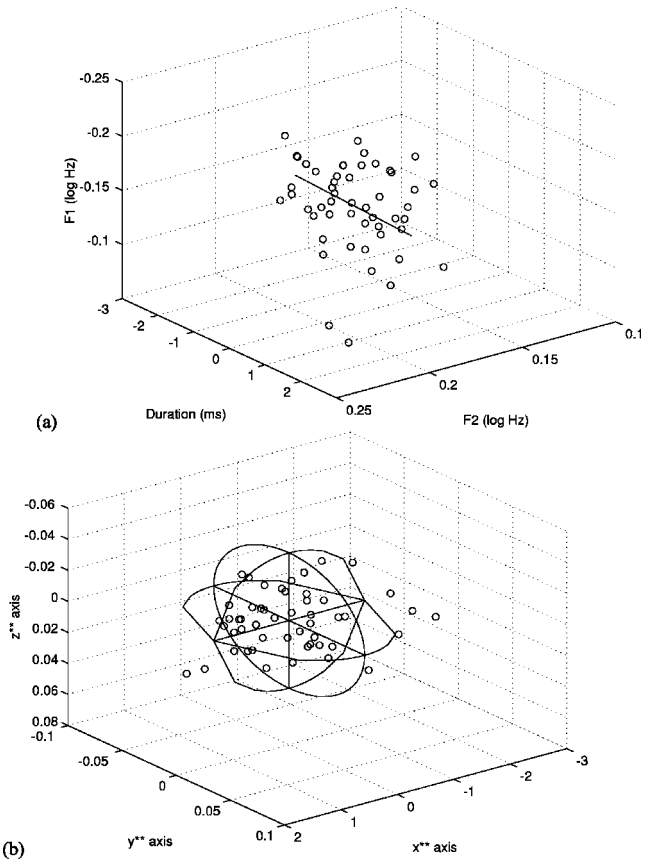


FIG. 5. Distribution of scatter for /i:/ for speaker TX (a) before data translation and rotation and (b) after. The solid line in (a) represents the location of the major axis.

$$\frac{x^2}{a_j^2}+\frac{y^2}{b_j^2}+\frac{z^2}{c_j^2}=1, \tag{B7}$$

where a_j, b_j, and c_j are the principal radii for ellipsoid j. The ellipsoid equation for vowel j may be converted to polar coordinates, using $x=r\cos\theta\cos\phi$; $y=r\sin\theta\cos\phi$; $z=r\sin\phi$ [Eq. (B8)]:

$$\frac{r^2\cos^2\theta\cos^2\phi}{a_j^2}+\frac{r^2\sin^2\theta\cos^2\phi}{b_j^2}+\frac{r^2\sin^2\phi}{c_j^2}=1. \tag{B8}$$

Equation (B8) can be solved for r and written as a function of $r_j(\theta,\phi)$ [Eq. (B9)]:

$$r_i(\theta,\phi)=\left[\sqrt{\left(\frac{\cos^2\theta\cos^2\phi}{a_j^2}+\frac{\sin^2\theta\cos^2\phi}{b_j^2}+\frac{\sin^2\phi}{c_j^2}\right)}\right]^{-1}, \tag{B9}$$

where j is the vowel index ($j=1$ for vowel 1, $j=2$ for vowel 2); a_j, b_j, and c_j are calculated from the standard deviations in data in the x_j^{**}, y_j^{**}, and z_j^{**} directions, respectively [Eqs. (B10)–(B12)]:

$$a_j=m\sigma_{x^{**}j}, \tag{B10}$$

$$b_j=m\sigma_{y^{**}j}, \tag{B11}$$

$$c_j=m\sigma_{z^{**}j}. \tag{B12}$$

Once followed for both vowels, the steps above yield two ellipsoids, each representing a vowel distribution. The representative ellipsoid for vowel 1 (referred to as ellipsoid 1) is shown along with original data for vowel 1 in Fig. 4(b). The representative ellipsoid for vowel 2 (referred to as ellipsoid 2) is shown along with original data for vowel 2 in Fig. 5(b).

The overlap calculations are based upon the volume of overlap of the two ellipsoids in (x,y,z) space. A graphical representation of ellipsoid j can be created in (x,y,z) space by generating parametric data E_j^{**} using Eq. (B9) for varying values of (θ,ϕ) in $(x_j^{**},y_j^{**},z_j^{**})$ space, and then rotating and translating this data from $(x_j^{**},y_j^{**},z_j^{**})$ space to (x,y,z) space. Assuming that we have E_j^{**}, the transformation is given by Eqs. (B13) and (B14):

$$\mathbf{E}_j^{*}=\mathbf{E}_i^{**}\begin{bmatrix}\cos\phi_j & 0 & -\sin\phi_j\\ 0 & 1 & 0\\ \sin\phi_j & 0 & \cos\phi_j\end{bmatrix}, \tag{B13}$$

$$\mathbf{E}_j=\mathbf{E}_j^{*}\begin{bmatrix}\cos\theta_j & -\sin\theta_j & 0\\ \sin\theta_j & \cos\theta_j & 0\\ 0 & 0 & 1\end{bmatrix}+E_{o,j}, \tag{B14}$$

where the mean values of F2, F1, and duration for the vowel distribution, $E_{o,j}$, are added to the coordinates for the ellipsoid representing that vowel (Fig. 6). In summary, the para-

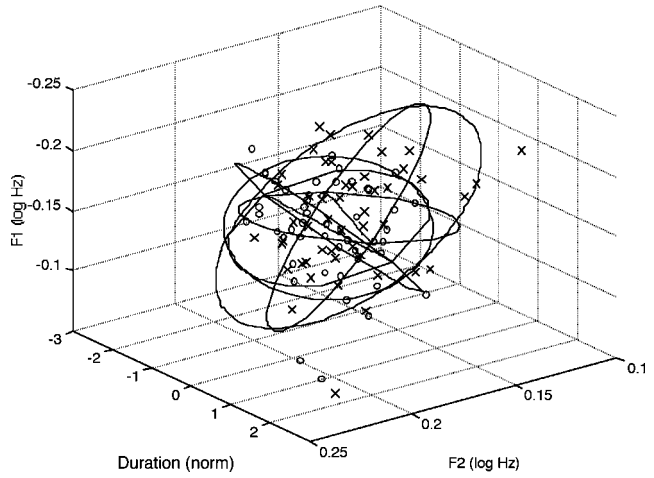


FIG. 6. Scatter data and representative ellipsoids for /i/ (x-marks) and /i:/ (o-marks) of speaker TX. For each ellipsoid, the line represents the major axis.

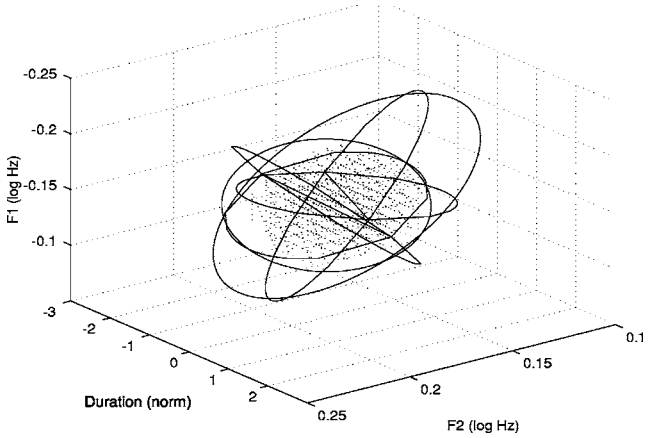


FIG. 7. Representative ellipsoids for /i/ and /i:/ of speaker TX, male, Jamaican Creole. The overlap region between the two vowels is shaded. Calculated 3-D overlap is $\Omega = 23\%$.

metrically generated ellipsoid data go through the reverse of the process used to align the original scatter data to the principal axes.

b. Calculating vowel overlap

To calculate the extent of overlap between the two ellipsoids, a somewhat "brute-force" method is employed. A space slightly larger than the space occupied by both ellipsoids is divided into a grid of evenly spaced test points. For most test cases, a $20\times20\times20$ test grid was found to give a good compromise between precision and computation time. For each test point, l, it is calculated whether or not the test point falls within ellipsoid j. The test point (x_l, y_l, z_l) is transformed into $(x_{lj}^{**}, y_{lj}^{**}, z_{lj}^{**})$ space, and the coordinate system aligned with ellipsoid j, by subtracting mean values and rotating the coordinate system [Eqs. (B15) and (B16)]:

$$[x_{lj}^{*} \quad y_{lj}^{*} \quad z_{lj}^{*}] = [(x_{lj} - x_{o,j}) \quad (y_{lj} - y_{o,j}) \quad (z_{lj} - z_{o,j})]\mathbf{R}_{\theta j}, \tag{B15}$$

$$[x_{lj}^{**} \quad y_{lj}^{**} \quad z_{ij}^{**}] = [x_{lj}^{*} \quad y_{ij}^{*} \quad z_{lj}^{*}]\mathbf{R}_{\phi j}. \tag{B16}$$

Next, the distance from the test point to the center of the ellipsoid is calculated [Eq. (B17)]:

$$d_{lj} = \sqrt{(x_{lj}^{**})^2 + (y_{ij}^{**})^2 + (z_{lj}^{**})^2}. \tag{B17}$$

Second, a radius is determined for ellipsoid j at the angle from the $\mathbf{x}^{**}$ axis to the line from the ellipsoid center to test point l. The angles are calculated from Eqs. (B18) and (B19):

$$\theta_{lj} = \tan^{-1}(y_{lj}^{**}/x_{lj}^{**}), \tag{B18}$$

$$\phi_{ij} = \tan^{-1}\left(\frac{z_{lj}^{**}}{\sqrt{(x_{lj}^{**})^2 + (y_{lj}^{**})^2}}\right), \tag{B19}$$

which can then be entered into the ellipse formula [Eq. (B19)] for $r_{lj}(\theta, \phi)$.

If the local radius is greater than or equal to the test point distance $[r_{lj}(\theta, \phi) >= d_{lj}]$, the test point is taken to be within the ellipsoid. This process is repeated for the test point relative to each of the ellipsoids $(j = 1, 2)$. The entire procedure is repeated for each test point l. The results are tallied for each of the test points: Is it in ellipsoid 1? Is it in ellipsoid 2? Is it in both ellipsoids?

The number and coordinates of all test points within the overall grid that are contained within each vowel are tracked. (The number of test points in an ellipsoid is approximately proportional to the "volume" of that ellipsoid.) The set of points that are contained both within vowel 1 and vowel 2 define the overlap region. Initially the overlap region is calculated relative to each vowel. The overlap fraction of vowel 1, Ω_1, is calculated [Eq. (B20)] as the total of test points in the overlap region between the two ellipsoids, N_{both}, divided by the total test points in ellipsoid 1, N_1,

$$\Omega_1 = N_{\text{both}}/N_1. \tag{B20}$$

Similarly, the overlap fraction of vowel 2, Ω_2, is calculated [Eq. (B21)] as the total of test points in the overlap region, N_{both}, divided by the total test points in ellipsoid 2, N_2.

$$\Omega_2 = N_{\text{both}}/N_2. \tag{B21}$$

The overlap fraction of the vowel pair, Ω, is defined [Eq. (B22)] as the larger of these two values:

$$\Omega = \max[\Omega_1, \Omega_2]. \tag{B22}$$

Since the two ellipsoids may be of different volumes, or because one ellipsoid may be contained within the other, the larger of the resulting calculations is taken as the overlap percentage. That is, it is desirable to know when one of the volumes protrudes substantially onto the other or is enveloped by the other. Note that this method does not tell us the region in three-dimensional space where the overlap is most extensive; the overlap percentage reflects only the extent of overlap. In Fig. 7, the representative ellipsoids are shown along with a shaded area representing the overlap region.

c. Calculation of vowel distribution overlap in two dimensions

The method for calculation of vowel distribution overlap in two dimensions is analogous to the three-dimensional case. Vowels are modeled as ellipses oriented at an angle to

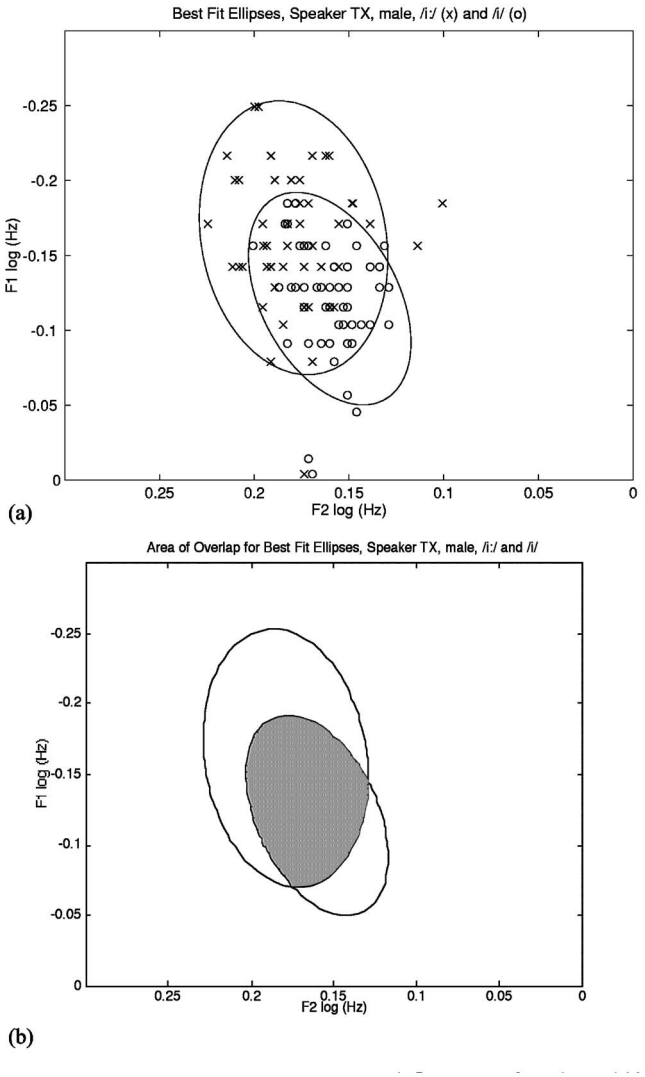


FIG. 8. (a) Best-fit ellipses fitted to scatter for vowel /i:/ (represented by "x") and /i/ ("o") in F1×F2 space. (b) Best-fit ellipses, in F1×F2 space, following calculation of overlap area (shaded). Calculated overlap fraction Ω=76%. For Jamaican Creole speaker TX, male, vowels /i:/ and /i/.

F1×F2 axes, using least squares fitting. Overlap is determined using a uniform grid of test points. Results of the two-dimensional method (for vowels /i/ and /i:/ for one male Jamaican Creole speaker) are shown in Fig. 8.

For the example we have followed throughout this Appendix, note that the calculated overlap Ω in the three-dimensional case (23%) is significantly lower than that (76%) calculated in the two-dimensional case where duration is not considered. This result suggests that duration contributes significantly to the contrast between /i/ and /i:/ for this speaker.

[1]That is, a systematic difference between the members of a phonological opposition. The convention of classifying vowel systems assumes similar spectral/temporal relations between all contrasting pairs in the system. This is an oversimplification, accomplished for theoretical convenience. In fact, empirical studies must take specific pairwise contrasts as a starting point, as will be accomplished in the study reported here, for reasons to be discussed in Sec. IV, below.

[2]In addition, the passage of time between the two studies raises the possibility that phonological change may also complicate comparability.

[3]Standard deviations are not provided. Furthermore, looking only at mean differences in F1 and F2 is only partly revealing. Women's and childrens' mean F1 for /a/-class words ("*bag,*" phonetically [æ]) is actually *lower* (i.e., this vowel is situated higher in inverted F1×F2 space) than /ɛ/ in the same system. Thus, they show greater raising than males in the same group.

[4]Dialect may have been a confounding factor in the Peterson and Barney study. Their 28 females and 15 children were primarily speakers of a Middle Atlantic dialect, while the 33 males represented "a much broader regional sampling." Thus, in addition to representing a different dialect than that of Hillenbrand *et al.*, the Peterson and Barney sample itself may have been drawn using speakers from different dialect regions (males versus females and children).

[5]Hillenbrand *et al.* do not give actual values for the curves presented in this graph, but hand-measuring yields the following lengths for each curve:

	Males	Females	Children
/æ/ (mm)	7	18	16
/ɛ/	6	6	7

[6]/h_d/ data were not able to be collected for the Jamaican speakers due to the deletion of /h/ in stressed, word-initial positions in all varieties of this language (Akers, 1981).

[7]The symbol /ɔ/ represents the fact that this word class was elicited for all speakers, all varieties. In the PNWEng sample, however, this phonetic quality did not exist. This word class was merged with /a:/.

[8]*Cot* and *caught* word classes are merged in this variety of American English. Thus, the /a:/ category in this study contains pooled data from both classes.

Adank, P., Smits, R., and van Hout, R. (**2004**). "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**, 3099–3107.

Akers, G. (**1981**). *Phonological Variation in the Jamaican Continuum* (Karoma, Ann Arbor, MI).

Crothers, J. (**1978**). "Typology and universals in vowel systems," in *Universals of Human Language*, edited by J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (Stanford U.P., Stanford), pp. 93–152.

Deterding, D. (**1990**). "Speaker normalization for automatic speech recognition," Ph.D. thesis, University of Cambridge.

Disner, S. F. (**1980**). "Evaluation of vowel normalization procedures," J. Acoust. Soc. Am. **67**, 253–261.

Ericsdotter, C., and Ericsson, A. M. (**2001**). "Gender differences in vowel duration in read Swedish: preliminary results," in *Proc. of Phonetic 2001, the XIVth Swedish Phonetics Conferences, Örenaäs*, 30 May 30–1 June 2001. Reprinted in *Working Papers*, *Department of Linguistics*, *Lund University*, Vol. **49**, pp. 34–37.

Hagiwara, R. (**1997**). "Dialect variation and formant frequency: the American English vowels revisited," J. Acoust. Soc. Am. **102**, 655–658.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hindle, D. (**1978**). "Approaches to normalization in the study of natural speech," in *Linguistic Variation, Models and Methods*, edited by D. Sankoff (Academic, New York), pp. 161–171.

Kluender, K., Diehl, R., and Wright, B. (**1988**). "Vowel-length differences before voiced and voiceless consonants: an auditory explanation," J. Phonetics **16**, 153–169.

Labov, W. (**1994**). *Principles of Linguistic Change: Internal Factors* (Blackwell, Oxford).

Labov, W. (**2001**). *Principles of Linguistic Change: Volume 2, Social Factors* (Blackwell, Oxford).

Ladefoged, P. (**1996**). Plot Formants software, UCLA Phonetics Laboratory, http://www.linguistics.UCLA.edu/faciliti/facilities/acoustics/acoustics.html

Lehiste, I. (**1970**). *Suprasegmentals* (MIT, Cambridge, MA), pp. 18–33.

Lobanov, B. M. (**1971**). "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. **49**, 606–608.

Macken, M. L. (**1980**). "Phonological universals and variation: on pattern and diversity in acquisition," unpublished Ph.D. dissertation, Stanford University.

Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, 2114–2134.

Moore, B. C. J. (**1989**). *An Introduction to the Psychology of Hearing*, 3rd ed. (Academic, London).

Nearey, T. M. (**1977**). "Applications of generalized linear modeling to vowel data," Proc. ICLSP **92**, 583–586.
Nearey, T. M. (**1989**). "Static, dynamic, and relational properties in speech perception," J. Acoust. Soc. Am. **85**, 2088–2113.
Nordström, P. -E., and Lindblom, B. (**1975**). "A normalization procedure for vowel formant data," Proc. ICPhS **8**, 212.
Perkell, J. S., and Matthies, M. L. (**1992**). "Temporal measures of anticipatory labial coarticulation for the vowel /u/: within- and cross-subject variability," J. Acoust. Soc. Am. **91**, 2911–2925.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. **24**, 175–184.
Shirai, S. (**2004**). "Lexical effects in Japanese vowel reduction," unpublished Ph.D. thesis, University of Washington.
Simpson, A., and Ericsdotter, C. (**2003**). "Sex-specific durational differences in English and Swedish," *Proc. of the XVth ICPhS*, Barcelona, 3–9 August **2003**, pp. 1113–1116.
Strange, W., and Bohn, O.-S. (**1998**). "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies," J. Acoust. Soc. Am. **104**, 488–504.
Syrdal, A. K. (**1984**). "Aspects of a model of the auditory representation of American English vowels," Speech Commun. **4**, 121–135.
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.
Wakita, H. (**1976**). "Normalization of vowels by vocal-tract length and its application to vowel identification," IEEE Trans. Acoust., Speech, Signal Process. **25**(2), 183–192.
Wassink, A. B. (**1999**). "A sociophonetic analysis of Jamaican vowels," unpublished Ph.D. dissertation, University of Michigan, Ann Arbor.
Wassink, A. B. (**2001**). "Theme and variation in Jamaican vowels," Language Variation Change, **13**(2), 135–159.

Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics

Roger W. Chan[a)]
Otolaryngology—Head and Neck Surgery; and Biomedical Engineering, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9035

Ingo R. Titze
National Center for Voice and Speech, The Denver Center for the Performing Arts, Denver, Colorado 80204

(Received 1 July 2005; revised 4 January 2006; accepted 16 January 2006)

Analytical and computer simulation studies have shown that the acoustic impedance of the vocal tract as well as the viscoelastic properties of vocal fold tissues are critical for determining the dynamics and the energy transfer mechanism of vocal fold oscillation. In the present study, a linear, small-amplitude oscillation theory was revised by taking into account the propagation of a mucosal wave and the inertive reactance (inertance) of the supraglottal vocal tract as the major energy transfer mechanisms for flow-induced self-oscillation of the vocal fold. Specifically, analytical results predicted that phonation threshold pressure (P_{th}) increases with the viscous shear properties of the vocal fold, but decreases with vocal tract inertance. This theory was empirically tested using a physical model of the larynx, where biological materials (fat, hyaluronic acid, and fibronectin) were implanted into the vocal fold cover to investigate the effect of vocal fold tissue viscoelasticity on P_{th}. A uniform-tube supraglottal vocal tract was also introduced to examine the effect of vocal tract inertance on P_{th}. Results showed that P_{th} decreased with the inertive impedance of the vocal tract and increased with the viscous shear modulus (G'') or dynamic viscosity (η') of the vocal fold cover, consistent with theoretical predictions. These findings supported the potential biomechanical benefits of hyaluronic acid as a surgical bioimplant for repairing voice disorders involving the superficial layer of the lamina propria, such as scarring, sulcus vocalis, atrophy, and Reinke's edema. [DOI: 10.1121/1.2173516]

PACS number(s): 43.70.Aj, 43.70.Bk [AL] Pages: 2351–2362

I. INTRODUCTION

Phonation threshold pressure (P_{th}), the minimum subglottal pressure or lung pressure required to initiate and sustain vocal fold oscillation, has been the focus of numerous theoretical and experimental studies in the past decade. P_{th} is an important indicator of vocal function, because it can be quantified in terms of some important biomechanical and physiological variables of phonatory control (Titze, 1988), plus it can be estimated *in vivo* by noninvasive means (Fisher and Swank, 1997; Jiang *et al.*, 1999b). Various approaches have been taken to test the theories associated with and the clinical significance of phonation threshold pressure, through the use of analytical models (e.g., Lucero, 1996, 1998, 1999), physical models (Chan *et al.*, 1997; Titze *et al.*, 1995), excised larynges (e.g., Jiang *et al.*, 1999a, 2000), and human subjects (e.g., Fisher *et al.*, 2001; Milbrath and Solomon, 2003; Roy *et al.*, 2003; Sivasankar and Fisher, 2002; Verdolini *et al.*, 2002). Functionally, P_{th} is an index of the minimum energy required to produce tissue oscillation, and provides an objective indication of such subjective sensations as vocal effort and the "ease of phonation" (Titze, 1988). P_{th} is associated with situations involving small-amplitude vocal fold oscillation, e.g., phonation onset or offset, where energy supplied to the vocal folds by the glottal airstream is just large enough to overcome energy dissipated during oscillation due to tissue damping, such that the oscillation can become self-sustained (Lucero, 1999). Based on a linearized analysis of the body-cover layered structure of the vocal fold (Hirano, 1975), Titze (1988) derived the fundamental analytical relationships that characterize the linear, small-amplitude self-oscillation of the vocal fold, including those that quantifyP_{th}. He showed that the propagation of a shear wave on the surface of the vocal fold, the mucosal wave, creates an asymmetric driving pressure that is more positive during vocal fold opening than during vocal fold closing (i.e., an intraglottal pressure in phase with the tissue velocity), establishing a key mechanism for the transfer of energy from the glottal airflow to the oscillating tissue. It was also shown that the acoustic reactance (the imaginary part of input impedance) of the supraglottal vocal tract is generally positive at a fundamental frequency (f_0) below the first formant frequency (F_1), leading to an inertive reactance or *inertance* that effectively reduces the total impedance of the vocal tract and contributes to sustain self-oscillation of the vocal fold (Rothenberg, 1981; Titze, 1988, 2004). This inertance describes the inertia or "sluggishness" of air par-

[a)]Author to whom correspondence should be addressed: Roger W. Chan, Otolaryngology—Head & Neck Surgery, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, Texas 75390-9035. Telephone: (214) 648-0386; fax: (214) 648-9122; electronic mail: roger.chan@utsouthwestern.edu

ticles accelerating or decelerating in the immediate supraglottal region of the vocal tract, commonly called the epilarynx tube. Such inertial changes lead to supraglottal pressure changes in phase with the direction of movement of the oscillating vocal folds, which contribute to an asymmetric intraglottal pressure variation during the glottal cycle for self-oscillation (Titze, 1988). However, quantitative relationships have only been developed for two separate scenarios describing the contributions of mucosal wave propagation and vocal tract inertance independent of each other. As both of these mechanisms likely contribute to self-oscillation in physiological situations, in the present study we attempt to develop some quantitative relationships that could take into account the effect of both energy transfer mechanisms simultaneously.

The hypothesis that inertive reactance (inertance) of the vocal tract could facilitate self-sustained vocal fold oscillation has been supported by findings from previous computer simulation studies (e.g., Titze and Story, 1997; Titze, 2004) as well as an excised hemilarynx study (Alipour *et al.*, 2001), but it has not been tested systematically. Therefore, the effect of vocal tract inertance on P_{th} was investigated in the present study through the development of revised analytical relationships, and empirical observations of a physical model of the larynx similar to one reported previously (Chan *et al.*, 1997; Titze *et al.*, 1995). The physical model was designed to simulate some key geometrical and biomechanical characteristics of the human vocal fold cover (the superficial layer of the lamina propria or mucosa), which is the major vibratory portion of the vocal fold in small-amplitude conditions. The design of the physical model minimized or eliminated some variabilities inherent in experiments involving excised larynges or human subjects, allowing for the isolation and precise control of crucial variables for testing analytical relationships empirically. In this study, a uniform-tube supraglottal vocal tract with an area function corresponding to the neutral schwa vowel was coupled to the physical model and its effect on P_{th} was measured.

In addition to acoustic impedance of the vocal tract, the biomechanical properties of the vocal fold could also have a profound impact on vocal fold oscillation dynamics and P_{th}. It has been shown that P_{th} is directly related to the viscosity, or more precisely the viscous shear properties, of the vocal fold cover (Chan, 1998; Titze, 1988). In the previous studies using the physical model (Chan *et al.*, 1997; Titze *et al.*, 1995), a synthetic polymer solution (sodium carboxymethyl cellulose) was incorporated into the artificial vocal fold cover and its viscosity was manipulated in order to test the theory. Results supported a roughly linear relationship between P_{th} and vocal fold tissue viscosity, consistent with the theoretical predictions. The polymer solution used for those studies was not a biomaterial for the vocal fold, however, and only its viscosity was varied, without taking into account the tissue viscoelasticity of the vocal fold cover. The viscoelastic shear properties of the vocal fold cover and a variety of implantable biomaterials have been reported (Chan and Titze, 1999a, 1999b), and computer simulation studies have shown that the viscoelastic shear properties of the cover are critical for dictating vocal fold oscillation energetics and dynamics, including both elastic and viscous properties (Alipour *et al.*, 2000; Berry, 2001). Hence, another goal of this study was to investigate the relationship between P_{th} and vocal fold tissue viscoelasticity, through empirical observations of the physical model with viscoelastic biomaterials implanted into the vocal fold cover.

Some recent studies have identified the glycosaminoglycan hyaluronic acid (HA) as a potentially optimal bioimplant for the surgical management of voice disorders involving the lamina propria, particularly the superficial layer or the vocal fold cover, from both a biological perspective and a biomechanical perspective. In the lamina propria extracellular matrix (ECM), increased HA levels have been found to be associated with reduced fibrosis and scar tissue formation during wound healing, with HA interacting with fibroblasts and other ECM proteins to mediate the cellular synthesis and degradation of various ECM proteins such as collagen, proteoglycans, and fibronectin (Ward *et al.*, 2002; Hirano *et al.*, 2003a, 2003b; Thibeault *et al.*, 2004). Biomechanically, the viscous shear properties (viscous shear modulus G'' and dynamic viscosity η') of HA at specific concentrations and molecular weights were quite close to those of the normal human vocal fold cover, and the selective enzymatic removal of HA from the cover significantly altered its viscoelasticity (Chan and Titze, 1999a; Chan *et al.*, 2001). In order to quantify the effect of vocal fold tissue viscoelasticity on P_{th}, HA was chosen as one of the biomaterials implanted into the physical model in the present study, as well as fibronectin and fat (human abdominal subcutaneous adipose tissue), which are among some of the most promising implantable biomaterials for the surgical repair of pathologies in the vocal fold lamina propria (Chan and Titze, 1998, 1999a; Dahlqvist *et al.*, 2004; Sataloff *et al.*, 1997). Results of such experiments would allow us to further examine the biomechanical benefits of these biomaterials as potential surgical implants for repairing voice disorders that damage the viscoelasticity of the lamina propria, such as scarring, sulcus vocalis, atrophy, and Reinke's edema.

II. THEORETICAL ANALYSIS

Titze's (1988) approach for the analysis of a self-sustained small-amplitude oscillation of the vocal fold is followed, based on the body-cover theory of the vocal fold layered structure (Hirano, 1975). Figure 1 shows the mid-coronal section of such a model with an idealized glottal geometry, as a linear approximation. The vocal fold body (muscle) and ligament are assumed to be stationary, with small-amplitude oscillation observed only on the vocal fold cover (superficial layer and intermediate layer of the lamina propria). The vocal folds are also assumed to be slightly abducted prior to the onset of oscillation, eliminating any nonlinearity associated with vocal fold collisions. Such assumptions are not expected to be valid in general, but they may be observed in the production of a very soft voice, i.e., during voicing onset and offset conditions, and in the production of a low-intensity falsetto voice with tissue oscillation primarily confined to the most superficial portion of the vocal fold lamina propria. Tissue motion with the propaga-

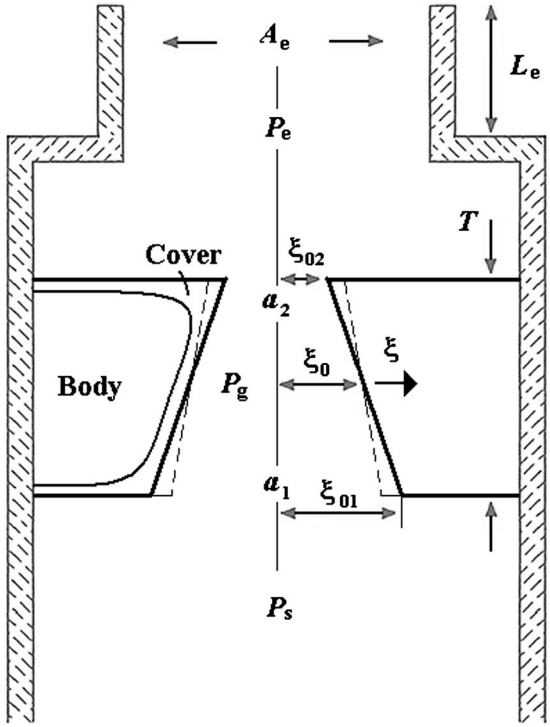


FIG. 1. A body-cover model of the vocal fold in the midcoronal section for linearized theoretical analysis of small-amplitude oscillation, with mucosal wave propagation and coupling to a supraglottal vocal tract (epilarynx tube). The vocal folds are shown in prephonatory positions (with prephonatory glottal half-width ξ_0 and tissue displacement ξ). The dotted lines indicate mucosal wave motion like a rotating plate in a bar-plate model (Titze and Story, 2002).

tion of a surface wave on the cover could be described by a mucosal wave model, equivalent to a rotating bar mass (dotted line) of a low-dimensional bar-plate model (Titze and Story, 2002). Analytical expressions derived by Titze (1988) for phonation threshold pressure P_{th} have been revised, taking into account both the effect of a mucosal wave and inertive acoustic impedance of a vocal tract simultaneously. The results are summarized next.

A. Small-amplitude oscillation in a mucosal wave model

Consider the body-cover model in Fig. 1, capable of a mucosal wave motion on the surface of the vocal fold, and acoustic loading of the supraglottal vocal tract, as given by the epilarynx tube, i.e., the immediate supraglottal region of the vocal tract. For simplification, acoustic loading of the subglottal tract (trachea) is not considered in the present model. The equation of motion for the vocal fold cover is given by that of a mass-spring-damper oscillator in a lumped-element representation, which has been shown to be equivalent to that of a low-dimensional bar-plate model (Titze and Story, 2002):

$$LTP_g = M\ddot{\xi} + B\dot{\xi} + K\xi, \tag{1}$$

where L is the vocal fold length, T is the vertical (superior-inferior) thickness of the vocal fold, P_g is the intraglottal pressure, M, B, K are the mass, damping, and stiffness of the oscillating tissue, respectively, and ξ, $\dot{\xi}$, $\ddot{\xi}$ are tissue displacement, velocity, and acceleration, respectively. In order to derive the expressions for phonation threshold pressure, the intraglottal pressure P_g has to be expressed in terms of some function of the tissue displacement and velocity in the equation of motion. P_g is the driving pressure for vocal fold oscillation, and was found to be a function of the epilarynx tube pressure P_e, the subglottal pressure P_s, as well as the ratio of the glottal areas at the entry and the exit of the glottis, a_1 and a_2 (at the inferior and superior edges of the vocal fold), respectively (Titze, 1988):

$$P_g = P_e + (P_s - P_e)\left(1 - \frac{a_2}{a_1}\right). \tag{2}$$

The epilarynx tube pressure P_e is given by the acoustic pressure of the vocal tract:

$$P_e = I\dot{u}, \tag{3}$$

where I is the inertive reactance or inertance of the vocal tract, u is glottal airflow at exit of the glottis, and $\dot{u}$ is the rate of change of flow (the time derivative of u). Resistive loss of the vocal tract is not considered here, because the acoustic impedance of the vocal tract is primarily inertive when the fundamental frequency is low compared with the formant frequencies ($f_0 < F_1$) for a uniform-tube vocal tract (Rothenberg, 1981; Titze, 1988, 2004). Computer simulation has also shown that the input impedance of a vocal tract with a narrow epilarynx tube remains inertive at even higher fundamental frequencies (Titze and Story, 1997). For quasisteady flow, by applying the continuity law of fluid flow, $u = a_2 v_2$, where v_2 is the air particle velocity at the glottal exit, the flow derivative $\dot{u}$ (which indicates acceleration of the air column in the epilarynx tube) can be expressed as

$$\dot{u} = \dot{a}_2 v_2 + a_2 \dot{v}_2, \tag{4}$$

where $\dot{v}_2$ is significant only for very small glottal areas. Since the present model assumes a small-amplitude oscillation of the tissue around an abducted glottis, the glottal areas are finite, and $\dot{v}_2$ becomes negligible because the viscous pressure loss at the glottis becomes insignificant. The glottal area at exit, a_2, and the glottal area at entry, a_1, are functions of the tissue displacement ξ from the prephonatory position as well as the tissue velocity $\dot{\xi}$ (Titze, 1988):

$$a_2 = 2L(\xi_{02} + \xi - \tau\dot{\xi}), \tag{5}$$

$$a_1 = 2L(\xi_{01} + \xi + \tau\dot{\xi}), \tag{6}$$

where L is length of the vocal fold, ξ_{01} and ξ_{02} are prephonatory glottal half-widths at glottal entry and glottal exit, respectively, and τ is a time delay given by $T/2c$, with T being the vertical thickness of the vocal fold and c being the mucosal wave propagation velocity. Hence, the flow derivative $\dot{u}$ becomes

$$\dot{u} = 2Lv_2(\dot{\xi} - \tau\ddot{\xi}), \tag{7}$$

and the epilarynx tube pressure P_e is

$$P_e = 2ILv_2(\dot{\xi} - \tau\ddot{\xi}). \tag{8}$$

With the glottal areas and the epilarynx tube pressure in Eqs. (5), (6), and (8), it is now possible to express the intraglottal pressure P_g in terms of vocal fold tissue displacement, velocity, and acceleration [re: Eq. (2)]:

$$P_g = 2IL\nu_2(\dot{\xi} - \tau\ddot{\xi}) + [P_s - 2IL\nu_2(\dot{\xi} - \tau\ddot{\xi})]\left(1 - \frac{\xi_{02} + \xi - \tau\dot{\xi}}{\xi_{01} + \xi + \tau\dot{\xi}}\right). \quad (9)$$

It is now clear that P_g has several components in phase with the tissue velocity $\dot{\xi}$, with both the epilarynx tube pressure (due to vocal tract inertance) and the glottal area change (due to mucosal wave propagation) contributing to an asymmetric driving pressure for self-sustained oscillation. However, the last term of this expression is highly nonlinear and has to be simplified to allow for further analysis. To do so, let there be two components of the tissue displacement ξ by considering a small-amplitude oscillation of the tissue around a prephonatory equilibrium (or abducted) position:

$$\xi = \bar{\xi} + \tilde{\xi}, \quad (10)$$

where $\bar{\xi}$ is the static component and $\tilde{\xi}$ is the dynamic or oscillatory component. The last term in Eq. (9) can be simplified by expanding the denominator binomially around $(\xi_{01}+\bar{\xi})$, and by assuming that $(\xi_{01}+\bar{\xi})^2 \gg (\tilde{\xi}+\tau\dot{\tilde{\xi}})^2$ as a linear, small-amplitude approximation:

$$\frac{(\xi_{01} - \xi_{02} + 2\tau\dot{\tilde{\xi}})(\xi_{01} + \bar{\xi} - \tilde{\xi} - \tau\dot{\tilde{\xi}})}{(\xi_{01} + \bar{\xi})^2}. \quad (11)$$

With this term and Eq. (10), the intraglottal pressure P_g can be expressed as

$$P_g = 2IL\nu_2(\dot{\tilde{\xi}} - \tau\ddot{\tilde{\xi}}) + [P_s - 2IL\nu_2(\dot{\tilde{\xi}} - \tau\ddot{\tilde{\xi}})] \times \left(\frac{(\xi_{01} - \xi_{02} + 2\tau\dot{\tilde{\xi}})(\xi_{01} + \bar{\xi} - \tilde{\xi} - \tau\dot{\tilde{\xi}})}{(\xi_{01} + \bar{\xi})^2}\right). \quad (12)$$

Substituting this expression for intraglottal pressure into the equation of motion [Eq. (1)] that specifies P_g as the driving force for self-oscillation, then grouping together the coefficients of the dynamic displacement $\tilde{\xi}$ and its derivatives, and eliminating all product terms of higher order as a linear approximation, the equation of motion becomes

$$(M\ddot{\tilde{\xi}} + B\dot{\tilde{\xi}} + K\tilde{\xi} + K\bar{\xi})\frac{\xi_{01} + \bar{\xi}}{2L^2T\nu_2} = -I\tau(\xi_{02} + \bar{\xi})\ddot{\tilde{\xi}} + [I(\xi_{02} + \bar{\xi}) + \beta\tau(\xi_{01} + \xi_{02} + 2\bar{\xi})]\dot{\tilde{\xi}} - \beta(\xi_{01} - \xi_{02})\tilde{\xi} + \beta(\xi_{01} - \xi_{02})(\xi_{01} + \bar{\xi}), \quad (13)$$

where $\beta = P_s/2L\nu_2(\xi_{01}+\bar{\xi})$. Comparing both sides of the equation, the effective mass M^*, effective damping B^*, and effective stiffness K^* of the tissue under small-amplitude oscillation can be readily obtained:

$$M^* = M + \frac{2I\tau L^2T\nu_2(\xi_{02} + \bar{\xi})}{\xi_{01} + \bar{\xi}}, \quad (14)$$

$$B^* = B - \frac{LT}{\xi_{01} + \bar{\xi}}\left(2IL\nu_2(\xi_{02} + \bar{\xi}) - \frac{\tau P_s(\xi_{01} + \xi_{02} + 2\bar{\xi})}{\xi_{01} + \bar{\xi}}\right), \quad (15)$$

$$K^* = K + \frac{LTP_s(\xi_{01} - \xi_{02})}{(\xi_{01} + \bar{\xi})^2}, \quad (16)$$

with the static or equilibrium tissue displacement $\bar{\xi}$ given by

$$\bar{\xi} = -\frac{\xi_{01}}{2} + \left[\left(\frac{\xi_{01}}{2}\right)^2 + \frac{LTP_s(\xi_{01} - \xi_{02})}{K}\right]^{1/2}. \quad (17)$$

B. Phonation threshold pressure

Equations (13)–(17) summarize the equation of motion of a small-amplitude vocal fold oscillation with the propagation of a surface mucosal wave on the vocal fold cover and the influence of inertive acoustic loading of the supraglottal vocal tract. By definition, when the effective damping $B^* = 0$, there is no net energy loss and vocal fold oscillation can be sustained indefinitely, i.e., a subcritical Hopf bifurcation has been reached at oscillation onset, generating an unstable limit cycle (Lucero, 1999). Therefore, from Eq. (15),

$$P_{\text{th}} = \frac{B(\xi_{01} + \bar{\xi})^2 - 2IL^2T\nu_2(\xi_{01} + \bar{\xi})(\xi_{02} + \bar{\xi})}{\tau LT(\xi_{01} + \xi_{02} + 2\bar{\xi})}, \quad (18)$$

where P_{th} is the phonation threshold pressure. For a rectangular prephonatory glottal geometry, the prephonatory glottal half-widths are the same, $\xi_{01} = \xi_{02} = \xi_0$, and Eq. (18) can be simplified as

$$P_{\text{th}} = \frac{\xi_0 + \bar{\xi}}{\tau}\left(\frac{B}{2LT} - IL\nu_2\right). \quad (19)$$

According to Eq. (17), the static or equilibrium tissue displacement $\bar{\xi}$ is zero for the rectangular prephonatory glottis. Recall that the time delay τ is related to the mucosal wave velocity $c(\tau = T/2c)$; hence

$$P_{\text{th}} = c\xi_0\left(\frac{B}{LT^2} - \frac{2IL\nu_2}{T}\right), \quad (20)$$

or

$$P_{\text{th}} = \frac{Bc\xi_0}{LT^2} - \frac{2IL\nu_2 c\xi_0}{T}. \quad (21)$$

These equations document the combined effects of vocal fold geometry, tissue properties, and vocal tract inertive acoustic loading on phonation threshold pressure for a rectangular glottal shape. The first term in Eq. (21) describes the contribution of tissue damping, which tends to raise P_{th}, given its direct relationship with the viscous damping coefficient B. The second term describes the contribution of the inertive impedance of the supraglottal vocal tract, which

tends to lower P_{th} with its negative sign. This finding supports the notion that vocal tract inertance contributes to facilitate self-sustained vocal fold oscillation.

Based on Eq. (21), it can be seen that small-amplitude self-oscillation of the vocal fold is facilitated by the following conditions (that would lower the phonation threshold pressure): (a) decreasing the tissue damping constant B, (b) decreasing the mucosal wave velocity c, (c) decreasing the prephonatory glottal half-width ξ_0, (d) increasing the vocal fold length L, (e) increasing the vocal fold vertical thickness T, and (f) increasing the vocal tract inertance I.

In order to relate this analysis to measurable, standard terminology for the biomechanical properties of the vocal fold, the viscous damping constant B and the mucosal wave velocity c can be expressed in terms of well-defined geometrical and viscoelastic parameters of the vocal fold. First of all, it is necessary to describe the damping constant B with the concept of a damping ratio. During cyclic loading of a viscoelastic material or biological soft tissue, the unloading process does not follow the same path as the loading process in their stress-strain relationship, because of frictional or viscous dissipation of internal energy. This time-dependent behavior can be characterized as hysteresis, or stress relaxation, with a decrease of stress in the unloading path compared to the loading path (Fung, 1993). The amount of viscous damping can be quantified by the damping ratio ζ, which is defined as the ratio of viscous damping to critical damping (Thomson, 1993):

$$\zeta = \frac{B}{B_c}, \tag{22}$$

where B_c is the critical damping coefficient. Critical damping can be defined as the amount of damping for the limiting condition between oscillation and no oscillation, where the energy to sustain oscillation is exactly lost to damping (when $\zeta = 1.0$). For a lumped-element mechanical oscillator system, i.e., the mass-spring-damper model or low-dimensional bar-plate model described by Eq. (1), it is given by

$$B_c = 2\sqrt{KM}, \tag{23}$$

where K is the effective stiffness (spring constant) and M is the effective mass of the system. Thus, the damping coefficient B is related to the damping ratio ζ as follows:

$$B = 2\zeta\sqrt{KM}. \tag{24}$$

The effective stiffness K of the system can be related to its elastic shear modulus G', where $K = 2G'LT/D$ (Titze and Story, 2002), with D being the vocal fold depth in the medial-lateral direction. Hence,

$$B = \zeta\sqrt{\frac{8G'MLT}{D}}. \tag{25}$$

Next, the mucosal wave velocity c can be expressed in terms of the shear modulus and the inertial modulus (density) of the cover (Titze and Story, 2002):

$$c = \sqrt{\frac{G'}{\rho_v}}, \tag{26}$$

where ρ_v is the density of vocal fold tissue. Substituting the expressions for B and c [Eqs. (25) and (26)] into Eq. (21), with the viscous shear modulus $G'' = \zeta G'$ (Chan and Titze, 1999b) and the density $\rho_v = M/LTD$, P_{th} can be expressed in terms of some standard linear viscoelastic functions of the vocal fold cover:

$$P_{th} = \frac{2\xi_0}{T}\left(\sqrt{2}G'' - IL\nu_2\sqrt{G'/\rho_v}\right). \tag{27}$$

As an approximation, the air particle velocity at glottal exit, ν_2, is related to the subglottal pressure P_s for quasi-steady flow, assuming no resistive loss in the vocal tract (Titze, 1988):

$$\nu_2 = \sqrt{\frac{2P_s}{\rho_v}}, \tag{28}$$

where P_s is equal to the phonation threshold pressure P_{th} by definition. A quadratic equation of P_{th} can be obtained by substituting Eq. (28) into Eq. (27), and by expanding and grouping the coefficient terms together,

$$\frac{T^2\rho_v}{4I^2L^2\xi_0^2G'}P_{th}^2 - \frac{\sqrt{2}TG''\rho_v^2 + 2I^2L^2\xi_0G'}{I^2L^2\xi_0G'\rho_v}P_{th} + \frac{2G''^2\rho_v}{I^2L^2G'} = 0. \tag{29}$$

Applying the general solution $x = (-B \pm \sqrt{B^2 - 4AC})/2A$ for quadratic equations of the form $Ax^2 + Bx + C = 0$, taking the negative sign solution according to Eq. (27), and by neglecting higher-order terms as a small-amplitude approximation, the general expression of P_{th} taking into account both the effect of the mucosal wave and inertive acoustic impedance of the supraglottal vocal tract can be obtained:

$$P_{th} = \frac{2\xi_0}{T}\left(\sqrt{2}G'' - \frac{2\sqrt[4]{2}ILG'}{\rho_v}\sqrt{\frac{\xi_0\zeta}{T}}\right). \tag{30}$$

This expression allows one to quantify the dependence of phonation threshold pressure on vocal fold geometry, linear viscoelastic shear properties of the vocal fold cover, as well as vocal tract inertance. As indicated by the first term of the equation, the dominant effect of the prephonatory glottal half-width ξ_0 on P_{th} is a direct positive relationship, whereas the dominant effect of vocal fold thickness T on P_{th} is an inverse relationship, consistent with previous analysis (Titze, 1988). The effects of ξ_0 and T in the second term of the equation on P_{th} will depend on the other geometrical and viscoelastic parameters, and are probably not as significant because the damping ratio ζ of the vocal fold cover is rather low at phonatory frequencies (around 0.1) (Chan and Titze, 2000). This expression also shows that P_{th} increases with the viscous shear properties of the vocal fold cover, i.e., the viscous shear modulus G'', or the dynamic viscosity η', since $\eta' = G''/\omega$, where ω is the angular frequency of oscillation. To test this hypothesis, biomaterials with varying viscoelastic shear properties were implanted into the vocal fold cover of a physical model of the larynx and their effect on P_{th} was

observed experimentally. Equation (30) also shows that the supraglottal vocal tract inertance I would contribute to lower P_{th}, facilitating vocal fold oscillation by effectively reducing the effect of tissue damping due to viscous energy loss of the vocal fold cover, and that this facilitative effect becomes more significant for a higher elastic shear modulus G' of the vocal fold cover (i.e., for stiffer tissue) and for longer vocal folds. This inertive impedance effect of the vocal tract takes place when the fundamental frequency is low compared to the first formant frequency ($f_0 < F_1$) (Titze, 1988, 2004). To test this hypothesis, a supraglottal vocal tract with significant inertive acoustic loading was coupled to the physical model and its effect was observed empirically.

III. METHOD

A. The physical model of the vocal fold cover

The previous design of the physical model involved simulating the vocal fold cover with a viscous polymer fluid (sodium carboxymethyl cellulose) encapsulated in a silicone membrane, with the vocal fold epithelium modeled by the membrane, and the superficial layer of the lamina propria modeled by the fluid (Chan *et al.*, 1997; Titze *et al.*, 1995). The artificial vocal fold cover was coupled to a trapezoidal stainless steel rod to simulate the vocal fold body in a passive state, and mounted in a rectangular Plexiglas airflow channel to simulate the glottal airway. Note that only the vocal fold cover was targeted by the physical model, but not the vocal fold body, because it has been observed to be nearly stationary during small-amplitude oscillation situations (Saito *et al.*, 1985). A compressed air supply served as the driving pressure or energy source for sustaining flow-induced oscillation of the vocal fold cover. An artificial lung and subglottal system was attached to the compressed air supply in line to reduce subglottal resonances, to minimize turbulence, and to achieve laminar flow.

In the present model (Fig. 2), the major improvement in design was that instead of injecting a viscous fluid to be encapsulated by the silicone membrane, the superficial layer of the lamina propria was simulated by a viscoelastic biomaterial that was implanted in between the silicone epithelial membrane and the stainless steel vocal fold body. The epithelial membrane was fabricated by molding a silicone dispersion fluid (Factor II A-1009, Lakeside, AZ) to a thickness of around 70 μm, and the superficial layer with a thickness of around 1.0 mm was created by evenly distributing the biomaterial on the medial surface of the body before attaching the epithelial membrane with a silicone adhesive (Dow Corning Medical Type A, Midland, MI). There were two advantages for this design: (1) The implantation instead of injection was intended to mimic phonosurgical procedures involving the direct placement of implantable biomaterials into the lamina propria, such as minithyrotomy (Gray *et al.*, 1999). (2) The use of viscoelastic biomaterials instead of a viscous polymer fluid provided a much more realistic simulation of the vocal fold cover and its vibratory behavior, as the human vocal fold cover is not only viscous, but viscoelastic (Chan and Titze, 1999b). Three viscoelastic biomaterials commonly used by surgeons for treating lamina propria deficiencies were used in the present study, namely fat (adipose tissue), hyaluronic acid, and fibronectin.

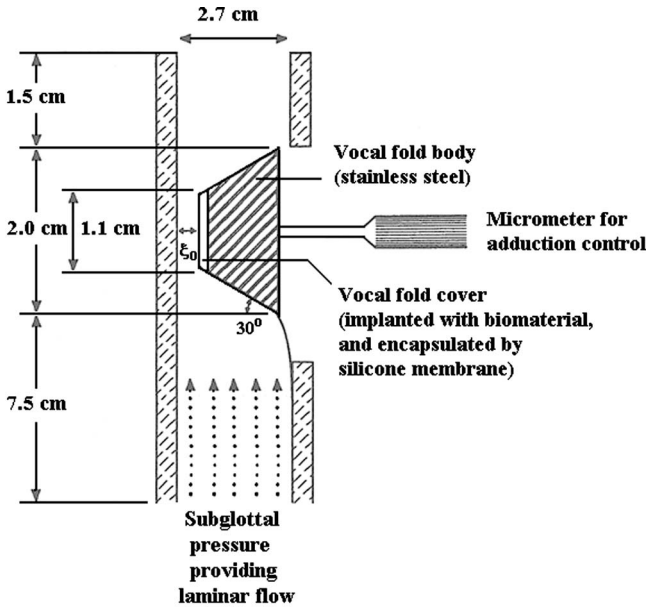


FIG. 2. Cross-sectional schematic of the physical model of the larynx with the simulated vocal fold cover, body, and the rectangular glottal airway, showing a rectangular prephonatory glottal geometry (not to scale). The transverse dimension of the airway is 2.22 cm. A biomaterial simulating a 1.0 mm thick superficial layer of the lamina propria is directly implanted into the vocal fold cover bounded by the epithelium and the vocal fold body.

The procedure for each experimental trial was similar to that of the previous model, where self-oscillation of the vocal fold cover was initiated by slowly increasing the subglottal pressure from zero with a pressure regulator with a range of 0–2 psi (Fairchild Model 10, Winston-Salem, NC), until a stable "mucosal wave" motion was observed on the artificial vocal fold cover. The subglottal pressure at that point was measured by a water manometer with a range of 60 cm H_2O and a resolution of 0.2 cm H_2O (Dwyer 1230-8, Michigan City, IN) and recorded as the onset threshold pressure. Next, subglottal pressure was slowly decreased until the mucosal wave motion ceased, at which time the subglottal pressure on the manometer was recorded as the offset threshold pressure. Average airflow through the flow channel (mean glottal flow) was measured by a rotameter (Gilmont Model J197, New York, NY). The oscillation of the vocal fold cover was observed with a Sony Handycam video camera mounted directly above the physical model, with a digital stroboscope (Pioneer Model DS-303, Cedarhurst, NY) triggered to be synchronized with the oscillation frequency such that still images and slow-motion video imaging of the vocal fold oscillation were obtained. The fundamental frequency was estimated from the stroboscopic system frequency, assuming periodic vibration. Frame-by-frame image analysis at a standard rate of 30 frames per second was conducted on the slow-motion video recording, allowing for the amplitude of oscillation to be measured with calibration of linear dimensions on a video screen.

B. Experiments with phonosurgical biomaterials

To manipulate the viscoelastic properties of the vocal fold cover while simulating clinical phonosurgical proce-

dures, biomaterials with varying viscous shear properties were implanted into the vocal fold cover of the physical model, including fat, hyaluronic acid (HA), and fibronectin, some of the most promising candidate materials for treating lamina propria disorders. Human subcutaneous adipose tissue (fat) was excised from the periumbilical anterior abdomen or the posterolateral abdomen of two cadavers within 24 h postmortem. The fat tissues harvested were divided into two portions, with one portion rinsed with saline solution and not subjected to any further processing. The other portion was minced into small pieces (2–3 mm in size) with fibrous tissue and fascia being removed, and rinsed with saline solution. Fat was processed by either of these two methods before being implanted into the physical model, because some phonosurgeons advocate minimizing fat tissue processing in order to preserve viable fat cells, while others prefer some kind of mechanical processing for easier injection (Sataloff *et al.*, 1997; Shaw *et al.*, 1997; Hsiung *et al.*, 2000).

Hyaluronic acid samples (Sigma-Aldrich Co., St. Louis, MO) from human umbilical cords with molecular weights of $\mathrm{Mr} \approx 3\times$ to 6×10^6 were prepared at concentrations of 0.01%, 0.1%, and 1.0%. These concentrations of HA were chosen to represent various levels of the molecular entanglement of spherical hyaluronate coil structures in solution, with 0.01% corresponding to little or no entanglement, 0.1% corresponding to some entanglement, and 1.0% corresponding to a fully (96%–100%) entangled dynamic molecular network (Laurent, 1987; Morris *et al.*, 1980). Fibronectin samples (Sigma-Aldrich Co., St. Louis, MO) from human foreskin fibroblasts with a molecular weight of $\mathrm{Mr} \approx 2\times10^5$ were mixed with 0.01% HA and 0.1% HA at a concentration of 0.033%. The physiological level of fibronectin in the vocal fold cover is not known to the authors, but a concentration of 0.033% was chosen because it was the approximate physiological level found in human synovial fluid, an extracellular matrix in the human body where the viscoelasticity of HA is critical (Abdullin *et al.*, 1988).

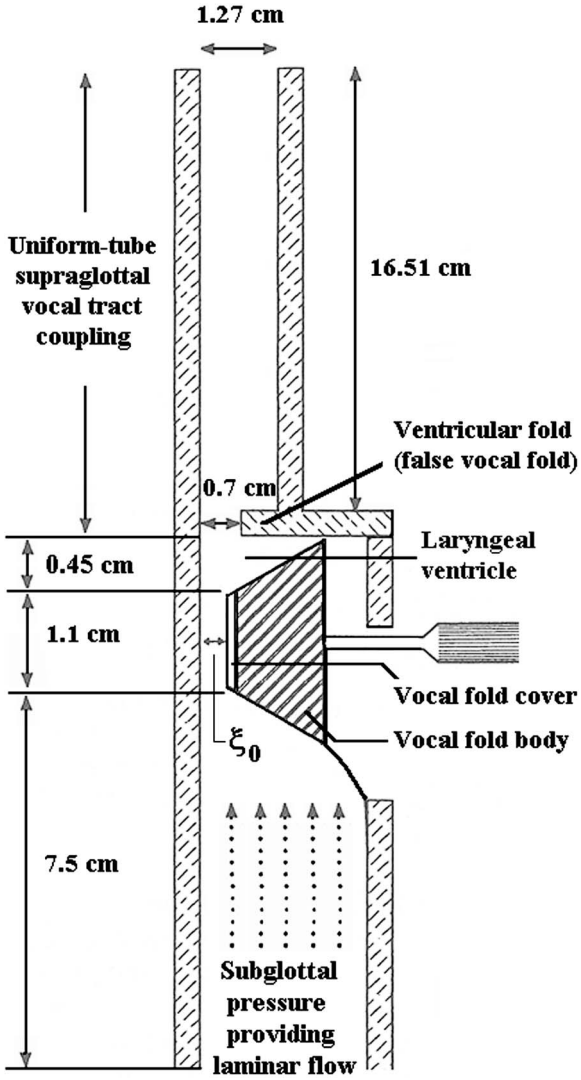


FIG. 3. Schematic of the physical model of the larynx with coupling to a rectangular uniform-tube supraglottal vocal tract (not to scale).

C. Experiments with uniform-tube vocal tract

A rectangular Plexiglas uniform tube with the same interior dimensions as the simulated glottal airway of the physical model was fabricated and coupled to the original model without air leakage, creating a "supraglottal vocal tract" that is continuous with the flow channel of the model (Fig. 3). The length of the tube (l) was 16.51 cm, and its cross-sectional area (A) was 2.82 cm^2. With these dimensions, its formant frequencies F_n can be calculated by the resonances of a quarter-wavelength closed-open tube:

$$F_n = \frac{(2n-1)s}{4l}, \tag{31}$$

where s is the speed of sound in a warm, humid vocal tract (35 000 cm/s). The first three formant frequencies of the uniform tube were found to be 530, 1590, and 2650 Hz, respectively, close to those for a typical human male vocal tract with a neutral schwa vowel /ə/ configuration. The horizontal wall coupling the base of the tube with the top of the original model was extended into the flow channel to create a rectangular-shaped artificial "ventricular fold" or false vocal fold. The ventricular fold was positioned 0.45 cm above the upper margin of the vocal fold cover and with an "adduction" level (false fold glottal half-width) of 0.7 cm. These values were close to those observed for an /a/ vowel for four normal male speakers based on *in vivo* x-ray laminagraphic data (Agarwal *et al.*, 2003), with an average false vocal fold height of 0.53 cm above the true folds, and an average gap of 0.613 cm between the false folds (i.e., a half-width of 0.31 cm). These geometrical dimensions were chosen so that the false fold could be scaled in proportion with the dimensions of the true vocal fold and the flow channel of the physical model.

The inertive reactance or inertance I of the uniform-tube vocal tract is given by

$$I = \frac{\rho l}{A} + \frac{\rho l_f}{A_f} \tag{32}$$

where ρ is density of air in the vocal tract (1.14 $\mathrm{kg/m}^3$), l is the vocal tract length (16.51 cm), A is the vocal tract

cross-sectional area (2.82 cm^2), l_f is the superior-inferior length of the false vocal fold (0.32 cm), and A_f is its corresponding cross-sectional area in the airway (1.55 cm^2). With the given dimensions, the inertance for this vocal tract was 690.90 kg/m^4, primarily due to the inertance of the uniform tube (667.43 kg/m^4 as compared to 23.47 kg/m^4, the inertance of the false vocal fold). For the "no vocal tract" condition in the original physical model (Fig. 2), there was actually a short section (1.5 cm) of the glottal airway that was immediately supraglottal. The cross-sectional area of this short section was 6.0 cm^2, resulting in a relatively insignificant inertance of 28.5 kg/m^4, an order of magnitude below that of the uniform tube.

IV. RESULTS

Self-sustained oscillation of the vocal fold cover in the physical model was readily observed with the biomaterials implanted, except for the trials with fat and hyaluronic acid of the highest concentration (1.0%). For all of the trials with sustained oscillation, a hysteresis phenomenon was always observed, with the onset threshold pressure consistently higher than the offset threshold pressure, as reported previously (Lucero, 1999; Titze *et al.*, 1995). Videostroboscopic analysis of the oscillation revealed that the fundamental frequency f_0 ranged from 100–150 Hz, and a frame-by-frame image analysis of slow-motion video of the oscillation showed that the maximum oscillation amplitude was around 1.0 mm at a subglottal pressure of 0.5 kPa. Interestingly, no stable, sustained self-oscillation of the vocal fold cover could be established when adipose tissue specimens (human abdominal subcutaneous fat) were implanted into the physical model, regardless of the size of the fat tissue specimens used, the method of fat tissue processing, and the anatomical location from which the fat tissues were harvested. Intermittent, and irregular or chaotic oscillation was observed at very high subglottal pressures (>3.0 kPa), but sustained, continuous, and stable phonation was not achieved. Likewise, for hyaluronic acid of the highest concentration (1.0%), no stable and sustained self-oscillation of the vocal fold cover was observed when it was implanted into the physical model, except under very high subglottal pressures beyond the physiological range (>3.0 kPa).

A. Effect of vocal fold tissue viscoelasticity

For the two lower concentrations of HA, stable self-oscillation of the vocal fold cover was consistently observed, and phonation threshold pressure measurements are summarized in Fig. 4. A linear relationship between P_{th} and prephonatory glottal half-width (ξ_0) is apparent, in agreement with previous results (Chan *et al.*, 1997; Titze *et al.*, 1995) as well as the present analysis [Eq. (30)]. Higher P_{th} values were consistently measured for 0.1% HA compared to 0.01% HA. This concentration effect became more profound in the presence of fibronectin, compared to HA without fibronectin. As shown in Fig. 5, an approximately 0.05–0.12 kPa difference in P_{th} was observed between the two concentrations of HA with fibronectin, versus only a 0.02–0.07 kPa difference between the two concentrations of HA without fibronectin (Fig. 4). Figure 6 shows the effect of the presence of fibronectin itself on P_{th} of the physical model. Higher P_{th} values were consistently observed for HA with fibronectin than for HA without fibronectin. As shown in Fig. 7, this fibronectin effect was more prominent for the higher concentration of HA (0.1%), where the presence of fibronectin was associated with an increase in P_{th} by around 0.07–0.12 kPa, as compared to the lower concentration of HA (0.01%), where the fibronectin-related increase was only about 0.02–0.08 kPa (Fig. 6).

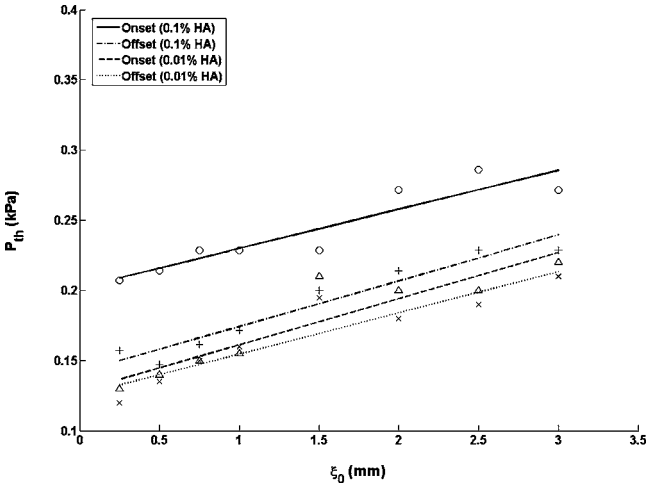


FIG. 4. Phonation threshold pressure (onset and offset P_{th}) of the physical model as a function of prephonatory glottal half-width (ξ_0), with the biomaterial hyaluronic acid (0.01% and 0.1% HA) implanted into the vocal fold cover.

B. Effect of vocal tract inertance

With all other key geometrical and biomechanical variables of the physical model being identical (i.e., prephonatory glottal half-width ranging from 0.25 to 3.0 mm, the glottal shape being rectangular, and tissue viscoelasticity of the vocal fold cover being constant), Fig. 8 shows that P_{th} was consistently lower for the physical model with supraglottal vocal tract coupling than for situations without vocal tract

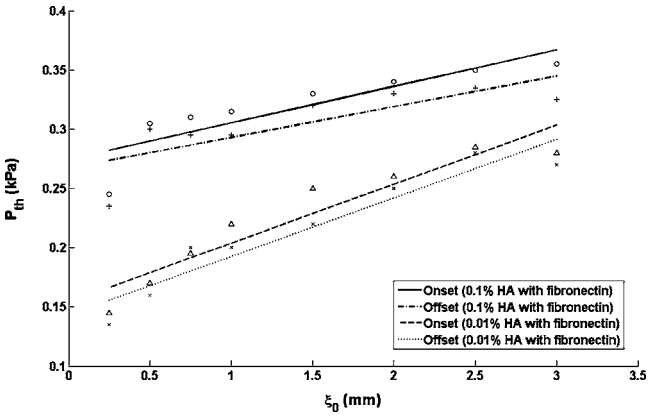


FIG. 5. Phonation threshold pressure (onset and offset P_{th}) of the physical model as a function of prephonatory glottal half-width (ξ_0), with the biomaterials hyaluronic acid (0.01% and 0.1% HA) and fibronectin (0.033%) implanted into the vocal fold cover.

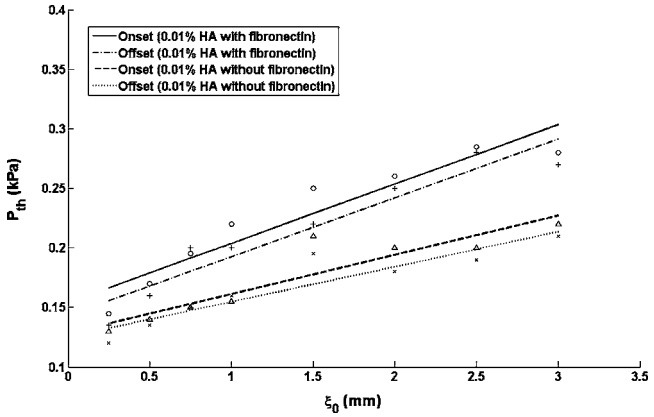


FIG. 6. Phonation threshold pressure (onset and offset P_{th}) of the physical model as a function of prephonatory glottal half-width (ξ_0), with the biomaterials hyaluronic acid (0.01% HA) and fibronectin (0.033%) implanted into the vocal fold cover.

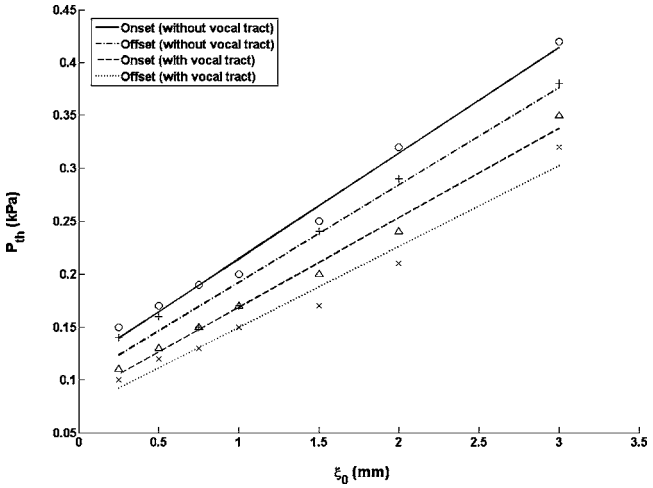


FIG. 8. Phonation threshold pressure (onset and offset P_{th}) of the physical model as a function of prephonatory glottal half-width (ξ_0), with and without a uniform-tube vocal tract.

coupling. This was true for both onset threshold pressures, where the P_{th} values were 0.03–0.08 kPa lower in the presence of the vocal tract, and offset threshold pressures, where the drop in P_{th} was 0.03–0.07 kPa.

V. DISCUSSION

There was no stable sustained oscillation of the vocal fold cover observed with fat tissue implanted in the physical model, even under very high subglottal pressures beyond the physiological range (>3.0 kPa). This finding was somewhat surprising, and was inconsistent with previous experimental and clinical studies that showed generally satisfactory results with the use of fat for repairing the vocal fold lamina propria (e.g., Sataloff *et al.*, 1997; Shaw *et al.*, 1997; Hsiung *et al.*, 2000). This apparent discrepancy could be related to a large variation in the method by which fat tissue is harvested, processed, and prepared for injection or implantation among different phonosurgeons. In the present study, similar effects on the physical model were observed for the two methods used to prepare the fat implant, including minimal processing with only rinsing by saline solution, and cutting of the adipose tissues to smaller pieces (2–3 mm). The inability of the physical model implanted with fat to sustain stable oscillation could also be due to the fact that the entire vocal fold cover was completely filled with fat in the present study, as opposed to the lamina propria being only partially filled with fat in surgical scenarios (Shaw *et al.*, 1997; Hsiung *et al.*, 2000). Our findings suggested that adipose tissue on its own, without a matrix of extracellular fibrous proteins and interstitial proteins and without any kind of active host response, may not be easily set into oscillation. On the other hand, fat assimilated with ECM proteins in clinical situations may be more favorable for vibration due to rheological interactions and biological activities. In particular, with viable fat cells, adipose tissue *in vivo* may trigger a tissue remodeling response that would yield a viscoelasticity more favorable for phonation.

When HA was implanted into the vocal fold cover of the physical model, it was found that a higher P_{th} was associated with a higher HA concentration, particularly for HA mixed with fibronectin. Similarly, a higher P_{th} was observed with fibronectin for the same HA concentration. These effects could be interpreted in terms of the differences in viscous

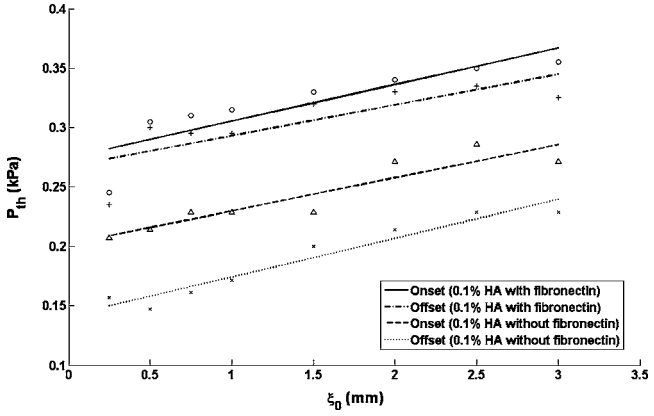


FIG. 7. Phonation threshold pressure (onset and offset P_{th}) of the physical model as a function of prephonatory glottal half-width (ξ_0), with the biomaterials hyaluronic acid (0.1% HA) and fibronectin (0.033%) implanted into the vocal fold cover.

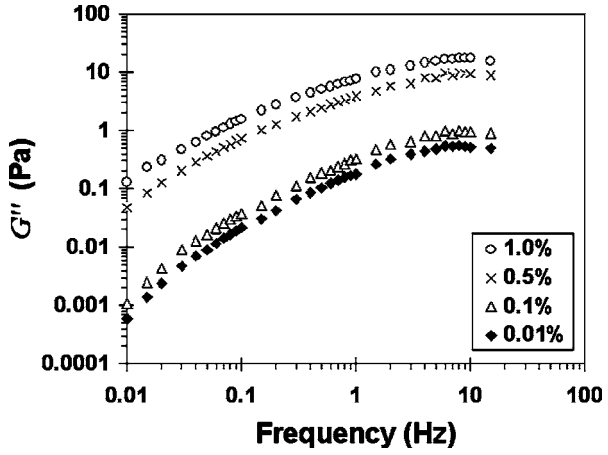


FIG. 9. Viscous shear modulus (G'') of sodium hyaluronate (HA) in phosphate buffered solution (pH 7.0) at four concentrations, from human umbilical cords (from Chan and Titze, 1999a).

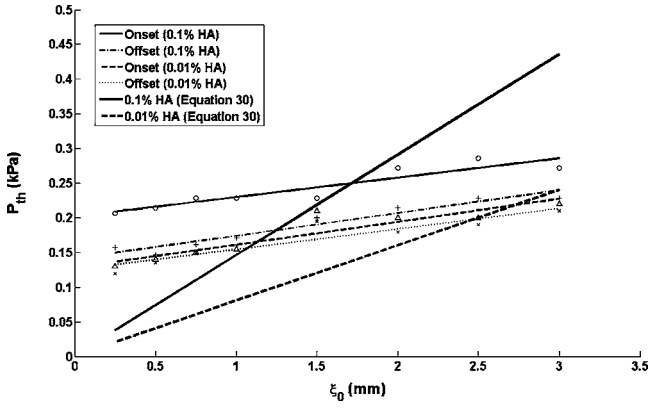


FIG. 10. Phonation threshold pressure (P_{th}) as a function of prephonatory glottal half-width (ξ_0) as predicted from Eq. (30), based on the viscoelastic data of 0.01% and 0.1% hyaluronic acid (HA) reported in Chan and Titze (1999a). Also shown are empirical values of onset and offset P_{th} of the physical model with 0.01% and 0.1% HA implanted into the vocal fold cover.

shear properties between the different biomaterials. Figure 9 shows the viscous shear modulus (G'') of HA at four concentrations, as quantified by a parallel-plate controlled stress rheometer in Chan and Titze (1999a). It can be seen that the viscous shear modulus (G'') of 1.0% and 0.5% HA was approximately 10–20 times higher than that of 0.1% HA throughout the frequency range tested (0.01–15 Hz), and that the G'' of 0.1% HA was about 1.5 to 2.0 times higher than that of 0.01% HA. These viscoelastic data correlated quite well with the findings of the present study, where P_{th} for 1.0% HA was unrealistically high (with subglottal pressures beyond the physiological range), and the phonation threshold pressure for 0.1% HA was approximately 1.5–2.0 times higher than that for 0.01% HA (Figs. 4 and 5). These correlations between the effect of the biomaterials on P_{th} and the viscoelastic data of the biomaterials supported the dependence of P_{th} on G'', as presented in the theoretical analysis in Sec. II [Eq. (30)].

In order to assess the numerical accuracy of the theoretical analysis in Sec. II, the predictions of phonation threshold pressure values based on Eq. (30) were compared to the empirical data of P_{th} values measured with the physical model with 0.1% HA and 0.01% HA implanted into the vocal fold cover (Fig. 4). The following parameter values of the physical model were used for the numerical predictions: $T=1.1$ cm, $L=2.2$ cm, $I=28.5$ kg/m^4 (for the "no vocal tract" condition in Fig. 2) [re: Eq. (32)], $\rho_v=1020$ kg/m^3 (Perlman, 1985), $G'=3.198$ Pa, $G''=0.587$ Pa (for 0.1% HA), and $G'=1.905$ Pa, $G''=0.324$ Pa (for 0.01% HA) [based on empirical data reported by Chan and Titze (1999a) extrapolated to 125 Hz]. Figure 10 shows the results of this comparison. It can be seen that Eq. (30) was capable of order-of-magnitude estimations of the dependence of P_{th} on the viscoelastic shear properties of the vocal fold cover, with higher P_{th} values predicted for 0.1% HA compared to 0.01% HA, given their different viscoelasticity (0.1% HA with a higher viscous shear modulus G''). This prediction was consistent with empirical findings of the physical model. Next, P_{th} was predicted to be directly related to the prephonatory glottal half-width ξ_0, also consistent with empirical observations. Nevertheless, the slope of the dependence of P_{th} on ξ_0 based on Eq. (30) did not agree exactly with that of the empirical data, leaving P_{th} underpredicted at small values of ξ_0 and overpredicted at large ξ_0. This discrepancy could be related to the effect of viscous pressure losses associated with small glottal widths (Lucero, 1996), which was not accounted for in the present analysis, or the incomplete understanding of the exact aeroacoustic interactions of vocal tract acoustic pressures with glottal Bernoulli and jet pressures (Titze, 2004), which could involve more than linear interactions, as assumed in the present model. Further studies along these directions are required.

For the effect of vocal tract inertance, lower phonation threshold pressures were observed with the presence of the uniform-tube vocal tract, when compared to the condition without the vocal tract. This finding could be related to the regulation of glottal airflow with the maximum power transfer principle, and the acoustic interactions between the glottal source and the vocal tract (Titze, 2002, 2004). Titze's (2002) simulations with a low-dimensional bar-plate vocal fold model showed that maximum acoustic power is transferred from the larynx to the supraglottal vocal tract when the input impedance of the vocal tract (load impedance) approaches that of the glottis (source impedance). In particular, it was found that the immediate supraglottal region of the vocal tract, or the epilarynx tube, tends to raise the vocal tract impedance to match with the glottal impedance (Titze and Story, 1997; Titze, 2002). This impedance matching has been shown to result in a larger amplitude of vocal fold oscillation, a higher degree of skewing of the glottal flow waveform (enhanced airflow asymmetry), and large asymmetric fluctuations in the intraglottal pressure due to superposition of the vocal tract acoustic pressure onto the Bernoulli pressure and jet pressure at the glottis (Titze, 2004). These effects of a higher vocal tract impedance contributed to a more efficient energy transfer from the glottal airflow to the tissue of the vocal folds for sustaining self-oscillation, particularly with the asymmetric intraglottal pressure. This situation of impedance matching was reasonably approximated by coupling the uniform-tube vocal tract to the physical model in the present study, as illustrated by comparing the load impedance and the source impedance. As discussed in Sec. II, the load impedance of the vocal tract for low fundamental frequencies (when $f_0 < F_1$) was primarily an inertance, and its magnitude Z_{load} can be approximated by the impedance of the epilarynx tube (Titze and Story, 1997), as given by

$$Z_{load} = \frac{2\pi f_0 \rho L_e}{A_e} \tag{33}$$

where L_e and A_e are length (0.65 cm) and area (1.55 cm^2) of the epilarynx tube of the vocal tract, respectively (Figs. 1 and 3). With an average f_0 of 125 Hz, Z_{load} was around 37 450 kg/s m^4 for the uniform-tube vocal tract. The glottal source impedance can be approximated by the ratio of mean subglottal pressure to mean glottal flow U:

$$Z_{\text{source}} = \frac{P_s}{U}. \quad (34)$$

For the physical model with the uniform-tube vocal tract attached, P_s was in the range of 0.1–0.3 kPa, whereas U was around 400–500 mL/s. Therefore, Z_{source} can be estimated to be in the range of 20 000 to 75 000 kg/s m^4. The magnitude of the epilarynx tube impedance Z_{load} was reasonably close to these magnitudes of the glottal source impedance, thereby elevating the impedance of the supraglottal vocal tract to match with the glottal impedance. According to the maximum power transfer principle, such impedance matching contributed to a more efficient energy transfer from the glottal source to the vocal tract, resulting in a higher acoustic output power or intensity (Titze, 2002). The elevated vocal tract inertance also contributed to facilitate the energy transfer to sustain self-oscillation of the vocal folds, by enhancing the asymmetry of the intraglottal driving pressure (Titze, 2004), resulting in lower phonation threshold pressures.

VI. CONCLUSION

A linear, small-amplitude oscillation theory had previously shown that vocal fold oscillation is facilitated by the presence of a vocal tract, with phonation threshold pressure lowered by an inertive acoustic loading of the vocal tract (Titze, 1988). This linear theory has been revised in the present study to simultaneously account for both the effect of a mucosal wave and the inertive impedance (positive reactance) of the vocal tract as the primary energy transfer mechanisms for sustaining flow-induced self-oscillation of the vocal fold. Analytical expressions of phonation threshold pressure (P_{th}) were derived, relating P_{th} to the linear viscoelastic shear properties of the vibratory portion of the vocal fold, and to the inertive reactance of the supraglottal vocal tract. Results showed that P_{th} is positively related to the viscous shear modulus (G'') or dynamic viscosity (η') of the vocal fold cover, and negatively related to the inertance of the vocal tract, consistent with former theoretical and modeling results (Titze, 1988, 2004).

To verify these theoretical findings, experiments were performed with a self-oscillating physical model of the larynx, with coupling to a supraglottal uniform-tube vocal tract, such that the effect of vocal tract inertance on phonation threshold pressure could be examined. Results showed that P_{th} was consistently lowered by significant inertance of the vocal tract. In particular, the epilarynx tube of the vocal tract contributed to raise the vocal tract inertive impedance to match reasonably with the glottal impedance (Titze, 2002), and this inertive effect likely facilitated the transfer of energy from the glottal airflow to the tissue of the vocal folds for self-sustained oscillation, by contributing to an asymmetric intraglottal pressure (Titze, 2004).

Additional experiments were performed with the physical model, with viscoelastic biomaterials implanted into the artificial vocal fold lamina propria (cover) of the physical model, including fat (adipose tissue), hyaluronic acid, and hyaluronic acid with fibronectin, such that the empirical effect of vocal fold tissue viscoelasticity on phonation threshold pressure could be assessed. Biomaterials were implanted directly into the lamina propria, instead of being injected, so as to simulate phonosurgical procedures that have been shown to be effective for repairing the vibratory layer (cover) of the vocal fold (Gray *et al.*, 1999). Self-sustained oscillation of the vocal fold cover in the physical model was induced with a compressed air supply, and the onset and offset phonation threshold pressures were measured with a manometer. Very high values of P_{th} were obtained with adipose tissue, with unstable oscillation of the vocal fold cover, whereas the lowest phonation threshold pressures were obtained with 0.01% hyaluronic acid (HA). When interpreted together with published viscoelastic data of HA (Chan and Titze, 1999a; Chan *et al.*, 2001), the findings of the present study supported the potential biomechanical benefits of HA for the surgical repair of deficiencies in the vocal fold lamina propria. In particular, HA at an appropriate concentration (0.01%–0.1%) could be an optimal phonosurgical implant material, because it likely lowers the phonation threshold pressure and facilitates tissue oscillation in the reconstructed vocal folds. However, due to the very short half-life of unmodified native HA molecules implanted into the vocal folds, there has to be some kind of cross-linking or molecular modification that would prevent the HA molecules from being quickly degraded by the host cells, and yet the modification should maintain their optimal viscoelastic properties (Chan *et al.*, 2001; Ward *et al.*, 2002). Such therapeutic use of HA is consistent with the results of some recent studies that showed the potential biomechanical and biological benefits of HA for the treatment of glottal insufficiency or vocal fold scarring (Dahlqvist *et al.*, 2004; Hertegård *et al.*, 2002; Thibeault *et al.*, 2004), as HA has been shown to be a key factor mediating the wound healing process in the vocal fold ECM and may contribute to prevent vocal fold scarring (Ward *et al.*, 2002; Hirano *et al.*, 2003a, 2003b).

ACKNOWLEDGMENTS

This work was supported by NIH Grants No. R01 DC006101 and No. R01 DC004347 from the National Institute on Deafness and Other Communication Disorders. The authors thank Michael R. Titze for his contributions to empirical experiments with the physical model.

Abdullin, A. R., Litvinov, R. I., Zinkevich, O. D., and Salikov, I. G. (**1988**). "Quantitative and qualitative changes in the fibronectin level in rheumatoid arthritis," Terapevticheskii Arkhiv **60**, 89–94 (in Russian).

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., Montequin, D., and Tayama, N. (**2001**). "Aerodynamic profiles of a hemilarynx with a vocal tract," Ann. Otol. Rhinol. Laryngol. **110**, 550–555.

Berry, D. A. (**2001**). "Mechanisms of modal and nonmodal phonation," J. Phonetics **29**, 431–450.

Chan, R. W. (**1998**). "Shear properties of vocal fold mucosal tissues and their effect on vocal fold oscillation," Ph.D. dissertation, The University of Iowa, Iowa City, Iowa.

Chan, R. W., and Titze, I. R. (**1998**). "Viscosities of implantable biomaterials in vocal fold augmentation surgery," Laryngoscope **108**, 725–731.

Chan, R. W., and Titze, I. R. (**1999a**). "Hyaluronic acid (with fibronectin) as a bioimplant for the vocal fold mucosa," Laryngoscope **109**, 1142–1149.

Chan, R. W., and Titze, I. R. (**1999b**). "Viscoelastic shear properties of

human vocal fold mucosa: Measurement methodology and empirical results," J. Acoust. Soc. Am. **106**, 2008–2021.
Chan, R. W., and Titze, I. R. (**2000**). "Viscoelastic shear properties of human vocal fold mucosa: Theoretical characterization based on constitutive modeling," J. Acoust. Soc. Am. **107**, 565–580.
Chan, R. W., Gray, S. D., and Titze, I. R. (**2001**). "The importance of hyaluronic acid in vocal fold biomechanics," Otolaryngol.-Head Neck Surg. **124**, 607–614.
Chan, R. W., Titze, I. R., and Titze, M. R. (**1997**). "Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **101**, 3722–3727.
Dahlqvist, A., Gärskog, O., Laurent, C., Hertegård, S., Ambrosio, L., and Borzacchiello, A. (**2004**). "Viscoelasticity of rabbit vocal folds after injection augmentation," Laryngoscope **114**, 138–142.
Fisher, K. V., and Swank, P. R. (**1997**). "Estimating phonation threshold pressure," J. Speech Lang. Hear. Res. **40**, 1122–1129.
Fisher, K. V., Ligon, J., Sobecks, J. L., and Roxe, D. M. (**2001**). "Phonatory effects of body fluid removal," J. Speech Lang. Hear. Res. **44**, 354–367.
Fung, Y. C. (**1993**). *Biomechanics. Mechanical Properties of Living Tissues*, 2nd ed. (Springer-Verlag, New York).
Gray, S. D., Bielamowicz, S. A., Titze, I. R., Dove, H., and Ludlow, C. (**1999**). "Experimental approaches to vocal fold alteration: Introduction to the minithyrotomy," Ann. Otol. Rhinol. Laryngol. **108**, 1–9.
Hertegård, S., Hallén, L., Laurent, C., Lindström, E., Olofsson, K., Testad, P., and Dahlqvist, A. (**2002**). "Cross-linked hyaluronan used as augmentation substance for treatment of glottal insufficiency: Safety aspects and vocal fold function," Laryngoscope **112**, 2211–2219.
Hirano, M. (**1975**). "Phonosurgery: Basic and clinical investigations," Otologia (Fukuoka) **21**, 239–440.
Hirano, S., Bless, D. M., Heisey, D., and Ford, C. N. (**2003a**). "Effect of growth factors on hyaluronan production by canine vocal fold fibroblasts," Ann. Otol. Rhinol. Laryngol. **112**, 617–624.
Hirano, S., Bless, D. M., Massey, R. J., Hartig, G. K., and Ford, C. N. (**2003b**). "Morphological and functional changes of human vocal fold fibroblasts with hepatocyte growth factor," Ann. Otol. Rhinol. Laryngol. **112**, 1026–1033.
Hsiung, M. W., Woo, P., Minasian, A., and Schaefer, Mojica, J. (**2000**). "Fat augmentation for glottic insufficiency," Laryngoscope **110**, 1026–1033.
Jiang, J., Ng, J., and Hanson, D. (**1999a**). "The effects of rehydration on phonation in excised canine larynges," J. Voice **13**, 51–59.
Jiang, J., O'Mara, T., Conley, D., and Hanson, D. (**1999b**). "Phonation threshold pressure measurements during phonation by airflow interruption," Laryngoscope **109**, 425–432.
Jiang, J., Verdolini, K., Aquino, B., Ng, J., and Hanson, D. (**2000**). "Effects of dehydration on phonation in excised canine larynges," Ann. Otol. Rhinol. Laryngol. **109**, 568–575.
Laurent, T. C. (**1987**). "Biochemistry of hyaluronan," Acta Oto-Laryngol. **442**, 7–24.
Lucero, J. C. (**1996**). "Relation between the phonation threshold pressure and the prephonatory glottal width in a rectangular glottis," J. Acoust. Soc. Am. **100**, 2551–2554.
Lucero, J. C. (**1998**). "Optimal glottal configuration for ease of phonation" J. Voice **12**, 151–158.
Lucero, J. C. (**1999**). "A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset-offset," J. Acoust. Soc. Am. **105**, 423–431.
Milbrath, R. L., and Solomon, N. P. (**2003**). "Do vocal warm-up exercises alleviate vocal fatigue?," J. Speech Lang. Hear. Res. **46**, 422–436.
Morris, E. R., Rees, D. A., and Welsh, E. J. (**1980**). "Conformation and dynamic interactions in hyaluronate solutions," J. Mol. Biol. **138**, 383–400.
Perlman, A. L. (**1985**). "A technique for measuring the elastic properties of vocal fold tissue," Ph.D. dissertation, The University of Iowa, Iowa City, Iowa.
Rothenberg, M. (**1981**). "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology*, edited by K. Stevens and M. Hirano (University of Tokyo Press, Tokyo, Japan), pp. 304–323.
Roy, N., Tanner, K., Gray, S. D., Blomgren, M., and Fisher, K. V. (**2003**). "An evaluation of the effects of three laryngeal lubricants on phonation threshold pressure (PTP)," J. Voice **17**, 331–342.
Saito, S., Fukuda, H., Kitahara, S., Isogai, Y., Tsuzuki, T., Muta, H., Takayama, E., Fujioka, T., Kokawa, M., and Makino, K. (**1985**). "Pellet tracking in the vocal fold while phonating—Experimental study using canine larynges with muscle activity," in *Vocal Fold Physiology*, edited by I. R. Titze and R. C. Scherer (eds.) (Denver Center for the Performing Arts, Denver, CO), pp. 169–182.
Sataloff, R. T., Spiegel, J. R., Hawkshaw, M. J., Rosen, D. C., and Heuer, R. J. (**1997**). "Autologous fat implantation for vocal fold scar: A preliminary report," J. Voice **11**, 238–246.
Shaw, G. Y., Szewczyk, M. A., Searle, J., and Woodroof, J. (**1997**). "Autologous fat injection into the vocal folds: Technical considerations and long-term follow-up," Laryngoscope **107**, 177–186.
Sivasankar, M., and Fisher, K. V. (**2002**). "Oral breathing increases P_{th} and vocal effort by superficial drying of vocal fold mucosa," J. Voice **16**, 172–181.
Thibeault, S. L., Rousseau, B., Welham, N. V., Hirano, S., and Bless, D. M. (**2004**). "Hyaluronan levels in acute vocal fold scar," Laryngoscope **114**, 760–764.
Thomson, W. T. (**1993**). *Theory of Vibration with Applications*, 4th ed. (Prentice–Hall, Englewood Cliffs, NJ).
Titze, I. R. (**1988**). "The physics of small-amplitude oscillation of the vocal folds" J. Acoust. Soc. Am. **83**, 1536–1552.
Titze, I. R. (**2002**). "Regulating glottal airflow in phonation: Application of the maximum power transfer theorem to a low dimensional phonation model," J. Acoust. Soc. Am. **111**, 367–376.
Titze, I. R. (**2004**). "A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice, " J. Voice **18**, 292–298.
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 Story, B. H. (**2002**). "Rules for controlling low-dimensional vocal fold models with muscle activation," J. Acoust. Soc. Am. **112**, 1064–1076.
Titze, I. R., Schmidt, S. S., and Titze, M. R. (**1995**). "Phonation threshold pressure in a physical model of the vocal fold mucosa," J. Acoust. Soc. Am. **97**, 3080–3084.
Verdolini, K., Min, Y., Titze, I. R., Lemke, J., Brown, K., van Mersbergen, M., Jiang, J., and Fisher, K. (**2002**). "Biological mechanisms underlying voice changes due to dehydration," J. Speech Lang. Hear. Res. **45**, 268–281.
Ward, P. D., Thibeault, S. L., and Gray, S. D. (**2002**). "Hyaluronic acid: Its role in voice," J. Voice **16**, 303–309.

Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude

Jay J. Bauer
Department of Communication Sciences and Disorders, University of Wisconsin—Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201-0413

Jay Mittal and Charles R. Larson[a)]
Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208

Timothy C. Hain
Departments of Neurology, Otolaryngology, and Physical Therapy/Human Movement Sciences, Northwestern University, 645 N. Michigan, Suite 1100, Chicago, Illinois 60611

(Received 3 May 2005; revised 10 January 2006; accepted 13 January 2006)

The present study tested whether subjects respond to unanticipated short perturbations in voice loudness feedback with compensatory responses in voice amplitude. The role of stimulus magnitude (±1,3 vs 6 dB SPL), stimulus direction (up vs down), and the ongoing voice amplitude level (normal vs soft) were compared across compensations. Subjects responded to perturbations in voice loudness feedback with a compensatory change in voice amplitude 76% of the time. Mean latency of amplitude compensation was 157 ms. Mean response magnitudes were smallest for 1-dB stimulus perturbations (0.75 dB) and greatest for 6-dB conditions (0.98 dB). However, expressed as gain, responses for 1-dB perturbations were largest and almost approached 1.0. Response magnitudes were larger for the soft voice amplitude condition compared to the normal voice amplitude condition. A mathematical model of the audio-vocal system captured the main features of the compensations. Previous research has demonstrated that subjects can respond to an unanticipated perturbation in voice pitch feedback with an automatic compensatory response in voice fundamental frequency. Data from the present study suggest that voice loudness feedback can be used in a similar manner to monitor and stabilize voice amplitude around a desired loudness level. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2173513]

PACS number(s): 43.70.Aj, 43.75.Bc [AL] Pages: 2363–2371

I. INTRODUCTION

The control of voice amplitude is an essential component of speech and singing and serves several functions. People modulate voice amplitude to attract or diminish attention to themselves, such as in role taking during dialogue, or to overcome environmental noise or distance. Amplitude, along with fundamental frequency (F_0) and duration, is used as an indicator of prosody. In this sense, amplitude is used to segment a message and place emphasis on certain words or syllables. Amplitude is also used to disambiguate confusing words or phrases during speech (Titze, 1994).

Understanding amplitude control is important for preventing and treating disorders characterized by abnormal voice. For example, people suffering from Parkinson's disease have difficulty communicating because the amplitude of their voice tends to be too low and monotonous (Ramig, 1994). Several types of neurological disorders (e.g., Parkinson's disease, spastic pseudobulbar palsy, dystonia) or mass lesions in the brain also present with instability in voice amplitude as well as voice F_0 (Ford and Connor, 2000; Ramig, 1994). Speech of schizophrenics is also seen as lacking emotional prosody and can be flat in affect (Alpert *et al.*, 2000; Leentjens *et al.*, 1998; Murphy and Cutting, 1990).

The peripheral mechanisms of amplitude control are well understood. Voice amplitude depends on complex interactions between the respiratory, laryngeal, and articulatory systems. Generally speaking, increases in lung pressure or F_0 lead to increases in voice amplitude. Adjustment, or tuning, of the supraglottal vocal tract, as done by well-trained singers, can also significantly affect voice loudness (Titze, 1994). However, given our detailed understanding of the peripheral mechanisms of amplitude control, there is a paucity of information related to neural control mechanisms of voice amplitude modulation.

Lombard, one of the first to investigate the topic of loudness control, showed the critical importance of voice loudness feedback on amplitude regulation in 1911 (as cited in Lane and Tranel, 1971). Lombard demonstrated that increased noise loudness feedback automatically raises voice amplitude to a level that can overcome environmental noise and thus enable a speaker to make him/herself heard. Adams has since proposed using the Lombard effect to elevate the voice loudness of patients with Parkinson's disease (Adams

[a)]Electronic mail: clarson@northwestern.edu

and Lang, 1992). A related phenomenon is that of side-tone amplification, where a speaker produces an increase in voice amplitude when they perceive (usually by the use of headphones) that their voice loudness feedback is too quiet for a given communication goal (Lane and Tranel, 1971). Conversely, when voice loudness is perceived as excessive, a speaker will reduce voice amplitude. Thus, speakers modulate their voice amplitude to compensate for changes in the loudness of voice auditory feedback. (We use the term amplitude in reference to voice output and loudness in reference to the voice feedback signal.)

However, one problem in interpreting these data on voice amplitude control relates to the methodology. In the studies of the Lombard phenomenon, a constant noise was added to the feedback signal, and in the side-tone studies, the voice feedback loudness was adjusted and remained at a constant level throughout the testing session (Lane and Tranel, 1971; Siegel and Pick, Jr., 1974; Siegel and Kennard, 1984). Given the ongoing presence of the feedback signals (noise or modulated voice), it is unknown whether the subjects were making voluntary adjustments to the signal or whether the responses were generated as an automatic response to the feedback. This methodology also precluded any temporal response measures (i.e., latency) of the vocal responses.

In a series of recent studies, it has been found that speakers compensate for unanticipated, brief duration perturbations in voice *pitch* feedback by modulating their voice F_0 (Burnett *et al.*, 1998; Burnett and Larson, 2002; Hain *et al.*, 2000, 2001; Jones and Munhall, 2000, 2002; Kawahara and Williams, 1996; Larson *et al.*, 2001, 2000; Natke *et al.*, 2003; Natke and Kalveram, 2001). These studies demonstrated that such compensatory responses have latencies of approximately 130 ms and do not seem to be volitionally controlled. The terms pitch-shift reflex, or response (PSR), has been used to refer to this process. The PSR has been studied during neutral vowel sounds, during pitch glissandos, and as subjects produce nonsense syllables, prolonged vowels during speech and during normal speech. Results from these studies suggest that the PSR helps to stabilize voice F_0 around an actual or intended target F_0.

More recently, Heinks-Maldonado and Houde (2005) described results of presenting brief perturbations in voice loudness to vocalizing subjects. In this study, the perturbations, ±10 dB, 400-ms duration, resulted in compensatory changes in voice amplitude with latencies of approximately 170 ms. The timing and compensatory nature of these responses were very similar to the results of the pitch-shift studies described above, and suggest similar vocal control mechanisms between F_0 and amplitude. However, since only perturbations of 10 dB were presented in the Heinks-Maldonado and Houde (2005) study, it is not clear whether the results generalize to less intense stimuli, such as those more typically encountered during natural speech. Also, the subjects in this task were asked to sustain voice amplitude at a single amplitude level. Thus, it is not known if the same results would be obtained at a lower amplitude level.

Given the wide range of vocal activities in which the PSR has been observed and the similarity of the compensatory nature of the PSR with that of the side-tone amplification studies mentioned above, we hypothesized that subjects would also respond to brief, unanticipated perturbations in voice loudness feedback with compensatory response in voice amplitude. Considering the results of the Heinks-Maldonado and Houde (2005) study mentioned above in which 10-dB stimuli were presented to the subjects, we chose magnitudes of 1, 3, and 6 dB to determine if the results would scale to less intense perturbations which are more typical of those observed in natural speech.

As mentioned previously, it was suggested that the PSR functions to help stabilize voice F_0. Recent data indicate that the PSR may also function in a task-dependent manner where the latency of the vocal response is modulated according to the intended speech goals (Bauer, 2004; Xu *et al.*, 2004). If the neural mechanisms responsive to perturbations in loudness feedback help to stabilize voice amplitude, they may also do so in a task-dependent manner. Specifically, subjects may be more sensitive to loudness perturbations during difficult vocal tasks compared to easier vocal tasks. One such difficult task may be vocalization near phonation threshold pressure. In such a situation, if the voice amplitude drops below phonation threshold pressure, vocalization will cease altogether. Controlling vocalization at soft vocal amplitude may require closer monitoring of auditory feedback to insure that vocal amplitude remains above phonation threshold pressure. If subjects monitor auditory feedback more closely at soft vocal amplitude, their responses to perturbations in voice loudness feedback may be greater in magnitude than those seen at normal vocal amplitude. Therefore, an additional hypothesis was tested in this study to assess whether response magnitudes to loudness-shifted voice feedback would be larger as subjects maintain a relatively soft voice compared to normal voice amplitude.

II. METHODS

A. Subjects

Twenty normal young adults (10 male, 10 female; ages 18–22) volunteered as subjects. No subjects reported a history of neurological, speech, or hearing disorders, and all subjects passed a hearing screening at 40-dB HL bilaterally at 500, 1000, and 2000 Hz. Subjects signed informed consent approved by the Northwestern University IRB and were paid for their participation.

B. Apparatus

Subjects were comfortably seated in a sound-attenuated chamber (IAC model 1201) and wore AKG headphones with attached boom-set condenser microphone (AKG HSC 200) at a 1-in. microphone-to-mouth distance during the experiment. The vocal signal was preamplified with a Mackie mixer, modulated for loudness feedback perturbations (loudness-shifted) with a Roland VS 880 EX effects processor coupled to an Ebtech line level shifter (model LLS-2), amplified with a Crown D75 amplifier and routed back to the AKG headphones after attenuation with HP decibel attenuators (model 350D). Masking pink noise (60 dB SPL) was presented to the subjects throughout the experiments using a Goldline audio noise source (model PN2; spectral frequen-

cies 1 to 5000 Hz). Subjects viewed a Dorrough loudness monitor (model 40 A) 0.5 m in front of them throughout the experiment in order to maintain target vocal amplitude and reduce vocal amplitude drift. Marks on the meter indicated calibrated voice amplitude levels of 75 and 60 dB SPL. The loudness monitor provided visual feedback related to the amplitude of the sustained vowel. Subjects were instructed to maintain constant voice amplitude near targeted levels throughout the entire utterance with minimal fluctuations (±3 dB). The Roland effects processor was controlled by MIDI software (MAX/MSP v4.5 by Cycling '74) from a laboratory computer. The voice, auditory feedback, and control signals (MIDI synchronization pulses and TTL pulses) were digitized on line with a PowerLab A/D converter by AD Instruments (10 kHz, 12-bit sampling, 5-kHz digital low-pass filter).

It was recognized that observation of the loudness monitor might be a confounding factor, because subjects could potentially respond to the visual changes in the monitor. However, the loudness perturbation was not displayed on the monitor, only the subject's voice output. Moreover, the monitor displays blinking lights that rapidly move left and right, and it was considered highly unlikely that subjects could adjust the amplitude of their voice rapidly enough and in synchrony with the lights so as to respond to individual fluctuations. It was considered more advantageous for the purpose of this study to help the subjects to maintain a relatively constant amplitude level over the course of their vocalizations rather than allow for vocal amplitude drift. As Heinks-Maldonado and Houde (2005) observed, natural speech has a tendency to decrease in amplitude as breath support is diminished throughout an utterance, and we wished to reduce this tendency as much as possible by providing a visual feedback guide of "general" vocal amplitude.

C. Procedures

Subjects were instructed to repeatedly sustain the vowel /u/ for approximately 5-s durations at either a normal (≈75 dB SPL) or soft amplitude level (≈60 dB SPL). Production of ten consecutive vocalizations constituted an experimental block. For each vocalization within a block, the voice loudness feedback was increased or decreased four times in succession, resulting in 20 increasing and 20 decreasing perturbations within each block. The duration of each perturbation was 200 ms and the magnitude was held constant at 1, 3, or 6 dB SPL within each block. In order to reduce potential predictability effects, the initial loudness perturbation occurred between 300 and 600 ms after vocal onset, while successive stimuli were presented with an interstimulus interval ranging from 900 to 1200 ms. Likewise, within each block, subjects were instructed to consistently maintain their voice amplitude at either the normal or soft phonation amplitude level. Overall, there were 12 experimental conditions collected across six blocks of vocalizations (2 voice amplitude levels ×2 stimulus directions ×3 stimulus magnitudes). The order of completion of the six experimental blocks was randomized across subjects to prevent potential order-related effects.

Digitized signals were analyzed off-line on a laboratory computer by converting the voice signal to a root-mean-square (rms) voltage signal using IGOR PRO software (v. 4.0 by Wavemetrics). Voice rms was calculated using the formula

$$\mathrm{rms}(x) = \sqrt{\frac{1}{N}\sum_{n-25}^{n+25} x^2}, \tag{1}$$

where x= the value of each data point, and N= total number of data points. Voice rms voltage measures were then converted to dB SPL using the following formula:

$$\mathrm{voice(dB)} = 20 \times \{\log[\mathrm{rms}(x)/c]\} + 75, \tag{2}$$

where x= the voltage level corresponding to each data point and c=0.323, which is the rms voltage corresponding to a vocal level of 75 dB SPL that was obtained through calibration procedures (see below). The vocal rms waveform of all 40 trials per block for each subject was then time aligned to the onset of the loudness-shift trigger stimulus, sorted based on stimulus direction, and averaged to produce one event-related averaged response per experimental condition per subject. From each average signal a prestimulus period of 200 ms was used to calculate a mean prestimulus baseline voice rms level. A valid average vocal response was operationally defined as a change in voice rms amplitude of at least two standard deviations (2 SD) from this mean baseline within a poststimulus response window of 900 ms for a duration of at least 50 ms and with a latency greater than 50 ms. Response latency was measured at the point where the rms average wave first crossed the 2-SD threshold after 50 ms. Response magnitude was measured as the greatest point of rms divergence following the response latency. A nonresponse was identified as not meeting the response criteria outlined above. This procedure for determining a valid vocal amplitude response is similar to that used previously to calculate a voice F_0 response (Bauer and Larson, 2003; Burnett *et al.*, 1998; Sivasankar *et al.*, 2005). To verify that subjects produced a louder voice for the normal compared to the soft condition, the mean of the prestimulus voice amplitudes under the two voice conditions were submitted to a one-way ANOVA.

Measures of magnitude, gain, and latency were each submitted to statistical analysis using separate three-way factorial ANOVAs with Bonferroni *posthoc* tests. Square root transformations of response magnitude, gain, and latency measures were done to achieve normal distributions as confirmed by normal probability plots, linear regression analyses (an R^2 value of 0.95 was considered acceptable), and coefficients of skewness and kurtosis in DATA DESK software (v6.2 for Mac OSX by Data Descriptions, Inc., Ithaca, NY). An *a priori* alpha level of 0.05 was used to determine statistical significance. Although these data lend themselves to repeated-measured ANOVAs, a three-way factorial ANOVA without repeated measures was used to identify statistical significance to account for missing data and unequal cell size.

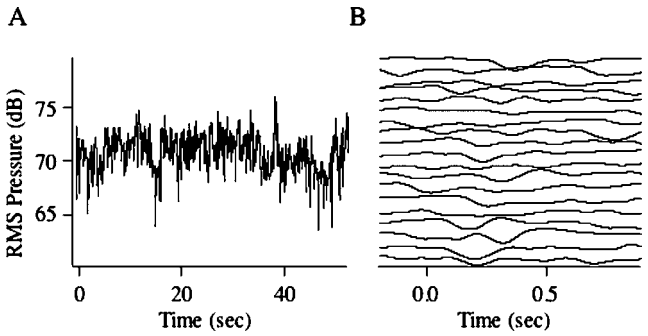


FIG. 1. (A) illustrates the voice rms pressure wave in dB across an entire recording session of ten vocalizations for one subject in one condition. (B) illustrates individual trials for upward loudness shifts for portions of the data in (A). The generally horizontal lines indicate there was little tendency for the subject to gradually reduce vocal loudness during the trials.

The microphone and headphones were both calibrated with a B&K sound-level meter (model 2203; weighting A), a B&K type 4131 1-in. microphone, and a B&K type 4152 artificial ear. The AKG boom microphone was calibrated within an IAC booth by presenting a triangle waveform (1 kHz from a function generator) through a free-field speaker. The microphone and sound-level meter were placed next to the speaker at the same distance. The function generator was adjusted to produce a 75-dB SPL amplitude signal as measured by the sound-level meter. The output of the microphone amplifier was recorded on the computer and converted to an rms pressure value (in volts). The rms voltage for a 75-dB SPL sound source was 0.323.

To calibrate the output of the headphones, the function generator was adjusted to produce an rms voltage of 0.323 at the output of the headphone amplifier and dB attenuator. This voltage source was presented directly to the input of the headphones (bypassing the microphone), which were placed

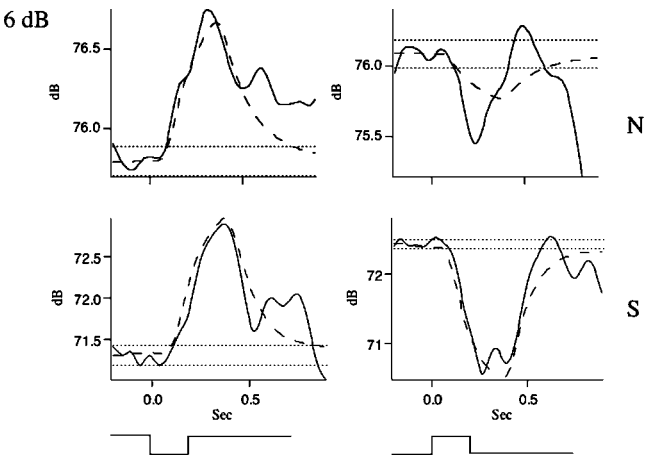


FIG. 2. Averaged responses for a representative subject for the 6-dB stimulus magnitude condition, upward and downward stimuli, and voice amplitude conditions. Description of panels: top row shows responses in the normal (N) voice condition, and bottom row shows responses in the soft (S) voice condition. Left column shows responses for downward stimuli, and right column shows responses for upward stimuli. Stimulus timing and direction are illustrated by square trace at bottom of panels. Solid curved line is averaged response. Dashed curved line is simulated response. Horizontal dotted lines indicates ±2 SDs of prestimulus mean. Vertical dashed lines indicate response latency.

TABLE I. Counts of compensatory (Comp.), following (Fol.), and nonresponses (NR) by voice amplitude condition.

	Normal	Soft	Total
Comp.	89	94	183
Fol.	2	3	5
NR	29	23	52
Total	120	120	240

on the artificial ear connected to the sound-level meter. The dB attenuators were then adjusted to achieve a 10-dB gain from the input of the microphone to the output of the headphones. This adjustment meant that a vocal level at the input to the microphone of 75 dB SPL was heard at the output of the headphones as 85 dB SPL.

III. RESULTS

Figure 1(a) displays a concatenated rms pressure waveform in dB consisting of ten consecutive 5-s vocalizations that comprised one of the average traces shown in Fig. 2. As can be seen in Fig. 1(a), there is considerable variability in the 50-s trace as voice amplitude occasionally drifts downward or upward. However, no consistent pattern appears to be present across vocalizations. Figure 1(b) displays the voice amplitude traces for each analysis window that was used to generate the average waveforms seen in Fig. 2. The traces for each trial (200-ms prestimulus baseline and 900-ms poststimulus response window) are essentially flat (except for the vocal compensation), with no discernible overall upward or downward drift across trials. These data are typical of all the subjects in the study. All subjects produced a lower amplitude voice for the soft condition (mean =70.2, ±1.5 dB) compared to normal (mean=76.1, ±2.3 dB; $F=502.3$, df=1,227, $p<0.0001$). However, subjective impressions of voice quality did not change as a result of the phonation amplitude.

All subjects responded to a loudness-shift stimulus, although not with equal frequency. Of the 240 total possible number of averaged responses (20 Ss × 2 voice amplitude levels × 2 stimulus directions × 3 stimulus magnitudes), 183 averaged responses (76%) compensated for the stimulus (response was in the opposite direction of the stimulus), 5 responses (2%) followed the stimulus (response in the same direction as the stimulus), and 52 responses (22%) did not meet the criteria for valid responses. Tables I–III show counts of compensatory, "following," and nonresponses as a function of voice amplitude, stimulus magnitude, and stimulus direction, respectively. While in most cases response types were evenly distributed across conditions, there were a

TABLE II. Counts of compensatory (Comp.), following (Fol.), and nonresponses (NR) by stimulus loudness.

	1 dB	3 dB	6 dB	Total
Comp.	48	74	61	183
Fol.	3	0	2	5
NR	29	6	17	52
Total	80	80	80	240

TABLE III. Counts of compensatory (Comp.), following (Fol.), and nonresponses (NR) by stimulus direction.

	Down	Up	Total
Comp.	88	95	183
Fol.	4	1	5
NR	28	24	52
Total	120	120	240

disproportionate number of nonresponses for the 1-dB stimulus condition compared with the 3-dB and 6-dB stimulus magnitudes as seen in Table II (chi square=23.61, df=4, $p<0.0001$).

Figure 2 illustrates representative responses for the 6-dB stimulus condition for the normal (top) and soft (bottom) voice amplitude conditions and for upward (right) and downward (left) loudness-shift stimuli. In all four cases, response latencies were close to 100 ms and compensated for the stimulus directions. The responses for the soft voice condition exceeded 1 dB, while those for the normal voice condition were less than 1 dB in magnitude. Similar trends may be seen for the 3-dB (Fig. 3) and 1-dB stimulus conditions (Fig. 4). However in Fig. 4, for the 1-dB normal voice condition, the changes in voice amplitude following the stimulus were quite variable and did not meet our criteria for valid responses. The curved dashed lines in Figs. 2–4 are simulations generated from the model shown in Fig. 7 (see below).

Statistical analysis using a $2\times2\times3$ ANOVA of the above magnitude-related trends across subjects revealed significant main effects for the voice amplitude, stimulus direction, and stimulus magnitude conditions. The voice amplitude condition resulted in significantly larger mean response magnitudes for the soft (0.99±0.50 dB SPL) compared with the normal (0.81±0.53 dB SPL) voice loudness condition ($F=9.6$, df 1,178, $p<0.03$). There was also a significant effect for stimulus direction with upward stimuli leading to larger mean response magnitudes (1.05±0.57 dB SPL) compared with downward stimuli (0.75±0.42 dB SPL; $F=19.6$, df=1,178, $p<0.0001$). There was also an overall significant effect of stimulus magnitude on response magnitude ($F=14.1$, df=2,178, $p<0.0001$). Bonferroni *posthoc* testing showed the 6-dB (0.98±0.27 dB SPL; $p<0.0001$) and 3-dB (0.96±0.26 dB SPL; $p<0.0001$) conditions produced significantly larger mean responses than the 1-dB condition (0.75±0.22 dB SPL). There were no significant interactions. Figure 5 (top row) illustrates the increase in response magnitude with stimulus magnitude as box plots for the normal (left) and soft (right) vocal conditions. However, when the same data are plotted as gain (response magnitude / stimulus magnitude) as shown in Fig. 5 (bottom row), an opposite effect is seen. In the gain plots, responses with the 1-dB stimuli approach unity (gain=1), whereas responses for the 3- and 6-dB stimulus magnitudes are less than 0.5.

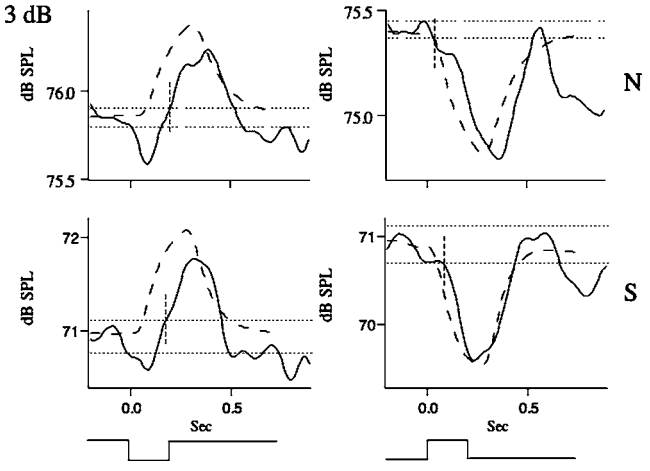


FIG. 3. Averaged responses and simulations for the same subject as in Fig. 1 for the 3-dB stimulus condition.

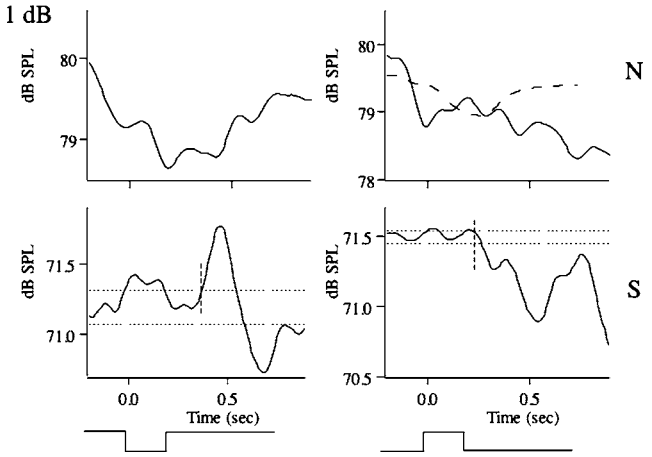


FIG. 4. Averaged responses and simulation for the same subject as in Figs. 1 and 2 for the 1-dB stimulus condition. Dotted lines representing the prestimulus mean loudness are not shown in the upper traces because these data did not meet the criteria of acceptable responses.

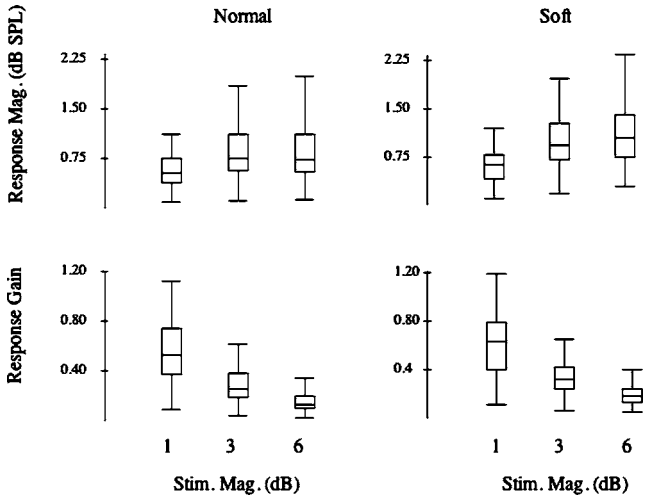


FIG. 5. Box plots illustrating response magnitude (dB SPL) (top row) and gain (bottom row) as a function of stimulus magnitude. Normal voice condition is on the left and soft voice condition is on the right. Box definitions: middle line is median, top and bottom of boxes are 75th and 25th percentiles, whiskers extend to limits of main body of data defined as high hinge +1.5 (high hinge—low hinge), and low hinge −1.5 (high hinge—low hinge). Points depicted by a circle extend beyond these limits, unless they exceed high hinge +3.0 (high hinge—low hinge) or low hinge −3.0 (high hinge—low hinge), in which case they are shown by an asterisk (DATA DESK; DATA DESCRIPTION).

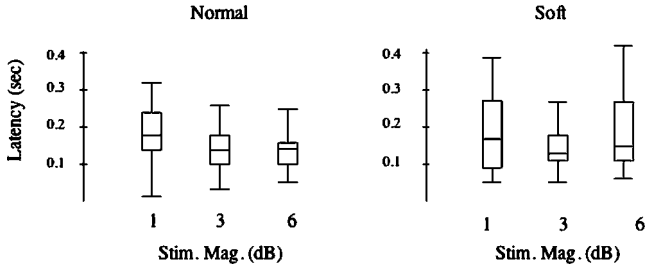


FIG. 6. Box plots illustrating response latencies (seconds) as a function of stimulus magnitude (dB). Data for normal voice condition are on the left and soft condition on the right.

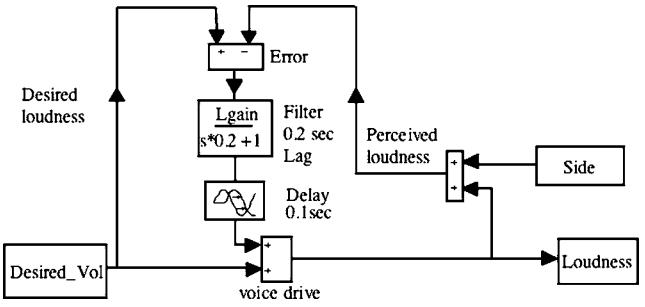


FIG. 7. Model of audio-vocal system producing the simulated traces depicted in Figs. 2–4. The variable representing desired loudness, *Desired_Vol*, is converted through a "black box" representing the entire central vocal production system (here just a summing junction labeled *voice drive*) into *loudness*. Loudness can be perturbed by adding a *side* input, creating *perceived loudness*. *Error* is computed by the difference between *desired loudness* and *perceived loudness*. *Error* is low-pass filtered in the element *filter* by scaling by L_{gain} and applying a lag with a time constant of 0.2 s. The filtered error signal is then passed through a delay of 0.1 s (to reproduce observed response latency), and added into the *voice drive* signal.

Statistical testing of response gain using a $2\times2\times3$ ANOVA also produced significant main effects for the voice amplitude, stimulus direction, and stimulus magnitude conditions. The voice amplitude condition produced a significantly larger gain for the soft (0.39 ± 0.30) than for the loud (0.32 ± 0.25) voice condition ($F=10.59$, df 1,173, $p<0.002$). An upward stimulus direction revealed a greater gain (0.41 ± 0.31) than downward stimuli (0.30 ± 0.23; $F=16.1$, df 1, 173, $p<0.0001$). Finally, stimulus magnitude produced the largest gain for the 1-dB condition (0.62 ± 0.34) compared with the 3-dB (0.33 ± 0.19) and 6-dB (0.17 ± 0.09) conditions ($F=81.8$, df 2,173, $p<0.0001$). Bonferroni *posthoc* comparisons demonstrated that the gain for the 1-dB condition were significantly larger than that for the 3-dB ($p<0.0001$) or 6-dB ($p<0.0001$) conditions, and the gain for the 3-dB condition was also significantly larger than that for the 6-dB condition ($p<0.0001$).

Response latencies were not affected by the experimental conditions. Although the latencies for 3-dB (mean =144 ms) stimulus magnitude were shorter than for the 1-dB (mean=185 ms) or 6-dB (mean=191) magnitude conditions ($F=4.42$, df=2,185, $p<0.014$), Bonferroni *posthoc* comparisons failed to yield significant comparisons. The overall mean latency was 157 ± 44 ms. Figure 6 illustrates latencies plotted against stimulus magnitude for the normal (left) and soft (right) voice conditions.

We simulated these responses using a simple negative feedback model shown in Fig. 7. This model is similar to that previously reported for stabilization of F_0 (Hain *et al.*, 2000). Desired loudness was compared to perceived loudness and computed as potential error. The error was filtered, delayed, and then used to adjust voice drive. For five randomly chosen subjects the scaling and delay parameters were fit using the optimization toolbox of MATLAB (version 14, Natick, MA). The time constant of the low-pass filter was set to 0.2 s based on the median of a preliminary fit where it was also allowed to vary. The variance accounted for (vaf) was used as an index of model performance. A perfect fit corresponded to a vaf of 1.0, and no correlation between the fit and data corresponded to a vaf of 0.

Eleven of 60 trials (12 conditions $\times$ 5 subjects) could not be fit with vaf's greater than 0.1. Six of these were for the lowest (1-dB) perturbation magnitude. As seen in Fig. 4, we were unable to fit the simulations for both "soft" conditions, and the "normal," down stimulus condition. Table IV shows the mean results of the fit in the remaining 49 trials for the five subjects.

There was a generally good fit of the model to the data and close correspondence between model parameters and experimentally measured values. Across the five modeled subjects, average L_{gain} (loudness scaling values from the model) ranged from 0.33 to 0.63. Across stimulus magnitude, L_{gain} decreased with increasing loudness (see Table IV). In this model, overall gain (change in loudness feedback)/(change in side tone) was a function both of L_{gain} and complex frequency (s). When L_{gain} was 0, or s was high, overall gain was 0. This can be seen from inspection of the filter element in the model (Fig. 7). When s goes to 0 (dc), neglecting the delays, overall gain was $[-L_{gain}/(L_{gain}+1)]$. The values shown in Table IV were for $s=0$.

The delay parameter, the model equivalent to experimental "latency" value, was substantially less (mean =90 ms) than the experimentally measured value (mean =157 ms). Average simulated delay across subjects ranged from 70 to 100 ms. Note that delay in the model was not an exact equivalent to experimental latency, as the delayed signals also passed through a lag, which contributed to the simulated latency. Mean L_{gain} was slightly larger for the upward stimulus direction (0.52) versus the downward direction (0.48), as well as much larger for the soft voice condition (0.61) versus the normal voice condition (0.32).

TABLE IV. Mean values of best fits for simulated variance (vaf), loudness scaling (L_{gain}), side-tone gain (S_{gain}), and delay across stimulus direction (up and down) for five representative subjects. n=number of conditions out of 12 for computation of the mean.

	1 dB Soft	1 dB Norm	3 dB Soft	3 dB Norm	6 dB Soft	6 dB Norm	Mean
n	6	8	8	10	9	8	
vaf	0.49	0.76	0.81	0.86	0.79	0.85	0.76
L_{gain}	0.97	0.33	0.56	0.42	0.30	0.23	0.47
S_{gain}	0.49	0.25	0.36	0.30	0.23	0.19	0.32
Delay (ms)	74	110	80	80	100	80	90

IV. DISCUSSION

The present study shows that brief perturbations in voice loudness feedback lead to compensatory responses in voice amplitude. The compensatory nature of these responses is similar to previous observations of side-tone amplification (Lane and Tranel, 1971; Siegel and Pick, Jr., 1974) and to perturbations in speech amplitude (Heinks-Maldonado and Houde, 2005). Furthermore, these data also suggest a mechanism of correcting for errors in amplitude production. Specifically, if voice amplitude, or more properly the perception of voice loudness, does not agree with the intended voice amplitude, the system makes automatic corrective responses in about 150 ms. Stimulus direction (up or down perturbations) or magnitude (1, 3, or 6 dB SPL) did not alter the response latency. Similarly, vocal task (soft or normal voice amplitude) did not change the response latency. However, the mechanism described in this study is not capable of complete compensation for errors. Complete compensation would imply a response magnitude equal to that of the perturbation (unity gain). The compensation responses described here rarely exceeded 1 dB SPL despite perturbations as great as 6 dB. An approximation of unity gain was only observed for the 1-dB SPL magnitude condition.

Responses to the 1-dB stimuli also differed from those to other stimuli by their greater numbers of measured nonresponses, which may indicate the 1-dB stimulus loudness is close to the threshold of the audio-vocal system for voice amplitude regulation. Although responses to 3- and 6-dB stimuli occur with much greater frequency than those to 1-dB stimuli (indicating these larger magnitude stimuli are well above the threshold), the audio-vocal system is not capable of fully compensating for feedback errors of these magnitudes. In a previous study of presenting loudness perturbations to vocalizing subjects, the magnitude of the stimuli were ±10 dB, and in that case there was almost a 100% response rate (Heinks-Maldonado and Houde, 2005). Thus, across the range of stimulus amplitudes from 1 to 10 dB, there is a relatively low response rate at 1 dB, and close to 100% at 10 dB, with 3 and 6 dB showing intermediate response rates. Therefore, the system seems less capable of recognizing small-magnitude stimuli compared to large magnitudes. Even though there was nearly a 100% response rate in the Heinks-Maldonado and Houde (2005) study, their response amplitudes also did not equal the stimulus magnitude. In their results, the mean response amplitude ranged from 0.61 to 1.32 dB, which is similar to results of the present study. It may be concluded then that even with very loud or soft stimuli, the system does not fully compensate for perturbations in voice loudness feedback. Similar results have been found also in studies of the pitch-shift response, where full compensation was observed for relatively small stimuli of 25 cents magnitude but not for larger magnitude stimuli of 100 or 200 cents (Burnett *et al.*, 1998; Larson *et al.*, 2000).

The fact that response amplitudes to both loudness and pitch-shifted feedback are limited in magnitude does not reduce their usefulness in stabilizing the voice. The vocal responses can begin within 150 ms of the onset of a loudness or frequency perturbation in voice pitch or loudness, and if the perturbation is small in magnitude (1 dB or 25 cents), the audio-vocal system could compensate for it and thus help to stabilize the voice. If the response does not completely nullify the perturbation, further automatic corrective responses could be triggered, or else voluntary mechanisms could intervene to compensate for the undesired perturbation. It is doubtful that systems as complex as the audio-vocal could rely on a single control mechanism to regulate output; rather, both voluntary and involuntary processes interact to stabilize the voice.

There are no unequivocal explanations for the failure of response magnitudes to equal the stimulus intensities; however, it was also previously noted that neither the Lombard response nor side-tone amplification effects had a gain of 1 (Lane and Tranel, 1971). In most cases the gain of the Lombard and side-tone amplification equaled about 0.5 across a much broader range of stimulus intensities than was used in the present study. One explanation for the response gain in the present study to be generally lower than 0.5 may be that subjects sustained nonspeech /u/ vowel sounds, whereas the studies of the Lombard effect and side-tone amplification were done on speech. The greater salience of the communication goal in speech compared to a sustained vowel without a specific communication goal may help explain why smaller response magnitudes were observed in the present study compared to the results of Lane and Tranel (1971). Future studies incorporating the techniques of the present study in a speech task may clarify this discrepancy.

A second explanation of the somewhat small response gain in the present study is that it may be reflexive, whereas observations of the Lombard and side-tone amplification were probably more of a voluntary response. The primary argument suggesting the responses may be reflexive relies on their latency. Responses that occur with a "short" latency are often described as being "reflexive," whereas responses with longer latencies are usually thought to be voluntary. Of critical importance is how one defines short. Typical short latency voluntary responses to a sound or visual stimulus by well-trained subjects in a reaction-time paradigm are on the order of 100–120 ms (Luschei *et al.*, 1967). However, these latencies only hold for "simple" reaction times of trained subjects. Reaction times involving a decision, "choice reaction time," are closer to 300 ms or more (Falkenstein *et al.*, 1993). In the present study, both brief upward and downward stimuli were presented, and untrained subjects generally produced compensatory responses, i.e., as if it were a choice reaction-time task with latencies around 150 ms. Therefore, on the basis of response latency, it is argued that the responses are reflexive in nature. Another argument bearing upon the reflexive nature of the responses is that subjects were indeed aware of some sort of feedback perturbation, but they did not know when or what type (upward or downward) of perturbation would be introduced into their auditory feedback. Given the randomized nature of perturbation onset, the brief perturbation duration of 200 ms, and randomization of perturbation type, subjects could not predict the onset time or type of stimulus and therefore could not adjust their vocalizations consciously in response to the perturbation. Never-

theless, the argument as to whether a response is voluntary or reflexive is not easily resolved, as most responses that a layperson would describe as being reflexive can be voluntarily modulated (Prochazka *et al.*, 2000). Our use of this term is meant to convey the idea that people respond to unexpected perturbations in voice loudness feedback without a conscious effort to do so, in other words automatically. It is possible that reflexive responses, of the type we are suggesting, have a lower gain than voluntary reactions.

If the reflexive responses described here are not capable of fully compensating for perturbations in voice loudness feedback, a reasonable question is what are their functions? Since the response magnitudes are rather small, it is suggested that this mechanism is largely responsible for stabilizing voice amplitude within a restricted range of 1 dB or less. Such a narrow range is one that would reduce the magnitudes of small fluctuations in vocal amplitude. It is also argued, as has been done for the pitch-shift reflex, that if the auditory-vocal system were capable of generating reflexive responses as large as those of the stimuli themselves, then environmental sounds, such as other voices, might cause large, unintended fluctuations in a person's vocal amplitude. Thus, a person would be unable to hold vocal amplitude steady in the presence of other sounds (Larson, 1998; Sivasankar *et al.*, 2005). Given these considerations, the system presumably relies on slower, consciously controlled voluntary mechanisms to meet the challenges of large drifts in the loudness of vocal feedback and relies on the automatic compensatory mechanisms for small perturbations.

In this study, we also observed that response magnitudes were larger for upward-shifted loudness feedback compared to downward shifts. We have no definitive explanation for this finding, but we suggest some possibilities. First, an increase in voice loudness may have greater salience (i.e., it may be more noticeable) than a decrease, and this may generate a larger response than a reduction in voice loudness. A second explanation may have to do with the reflexive nature of the response. If indeed the response is reflexive, the mechanisms for perception of voice loudness feedback change and/or generating the response may be more sensitive to an increase in voice amplitude than a decrease, which may lead to a stronger reaction. We have no explanation for such an asymmetry, but it could be related to a greater need to control for excessive voice loudness than reduced loudness. There may be no end to possible reasons for this, but it could be related to evolutionary pressures to guard against the possibility of attracting attention to oneself by being excessively loud. As a final note, it is interesting that control of voice loudness is a problem in people suffering from Parkinson's disease, and the reflexive mechanisms discussed here may not be adequately controlled in this disorder (Kiran and Larson, 2001).

In this study, we tested the hypothesis that control of voice amplitude requires greater reliance on auditory feedback for a soft voice compared to normal voice amplitude. The results supported this hypothesis by showing that response magnitudes were significantly greater under the soft voice production condition compared to the normal voice condition. Thus, it may be suggested that vocalizing, or speaking, with a relatively soft voice requires closer monitoring of auditory feedback than normal voice amplitude.

Our feedback model of loudness control generally reproduced the data well, showing that it is a feasible representation of internal circuitry that produces these responses. We modeled five subjects rather than the full dataset because these subjects are representative of the others. The purpose of the model was only to demonstrate a feasible and simple circuit that could account for the experimental results—a working hypothesis. The delay parameter was stable within subjects, ranging from 70 to 100 ms. Model delay was shorter than experimentally measured latency because of the model's low-pass filtering, which adds a time lag. This points out a pitfall—neglecting internal dynamics—that could have arisen had we assumed that latencies were derived from a simple internal delay process. As was found in the experimental data analysis, simulated side-tone gain decreased by a factor of about 2 with perturbation loudness and also was increased for the soft versus normal voice conditions. This considerable variability indicates that, at least for the moment, side-tone gain should not be considered "hard-wired."

V. CONCLUSION

In the present study, subjects were asked to sustain vowel sounds while hearing short perturbations (200 ms) in voice loudness feedback over headphones. Subjects responded to loudness perturbations with compensatory corrections in voice amplitude approximately 76% of the time, but full compensation was generally not achieved. Data analysis included only these compensatory responses. The remaining averages either "followed" the direction of amplitude perturbation (2%), or were classified as nonresponses (24%). Compensatory response magnitudes were further calculated as gain (response magnitude/stimulus magnitude) indicating responses to 1-dB stimuli approximated unity, while those to 3- or 6-dB stimuli were less than 0.5. These findings suggest that the system responsible for compensation is nearly optimal for correcting for small perturbations in voice amplitude, and hence may function to stabilize voice amplitude. However, full compensation was generally not achieved for larger magnitude stimuli. Response magnitudes were also compared across vocal task, indicating that the audio-vocal system relied on auditory feedback to a greater extent during a more difficult vocal task such as low voice intensity production compared to a less-challenging task such as conversational voice intensity production. Furthermore, the overall mean response latency across conditions was 157 ms, suggesting the responses were generated automatically. A relatively simple feedback model of the voice amplitude control system generated response simulations that captured the main features of the compensations, suggesting the model was a feasible representation of the actual internal mechanisms. Overall, the results suggested that auditory feedback was used to correct for small variations in voice loudness feedback and thereby helped stabilize voice amplitude.

ACKNOWLEDGMENTS

We would like to thank the reviewers for their insightful comments and helpful suggestions for improvement of the manuscript. This study was supported by NIH Grant No. DC006243-01A1.

Adams, S. G., and Lang, A. E. (**1992**). "Can the Lombard effect be used to improve low voice intensity in Parkinson's disease?," Eur. J. Disord. Commun. **27**(2), 121–127.

Alpert, M., Rosenberg, S. D., Pouget, E. R., and Shaw, R. J. (**2000**). "Prosody and lexical accuracy in flat affect schizophrenia," Psychiatry Res. **97**(2–3), 107–118.

Bauer, J. J. (**2004**). "Task dependent modulation of voice *F*0 responses elicited by perturbations in pitch of auditory feedback during English speech and sustained vowels," Ph.D. dissertation, Northwestern University.

Bauer, J. J., and Larson, C. R. (**2003**). "Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique," J. Acoust. Soc. Am. **114**(2), 1048–1054.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**(3), 1058–1063.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice *f*0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**(6), 3153–3161.

Falkenstein, M., Hohnsbein, J., and Hoormann, J. (**1993**). "Late visual and auditory ERP components and choice reaction time," Biol. Psychol. **35**, 201–224.

Ford, C. N., and Connor, N. P. (**2000**). "Phonatory effects of mass lesions," in *Voice Quality Measurement*, edited by R. D. Kent and M. J. Ball (Singular, San Diego), pp. 377–384.

Hain, T. C., Burnett, T. A., Larson, C. R., and Kiran, S. (**2001**). "Effects of delayed auditory feedback (DAF) on the pitch-shift reflex," J. Acoust. Soc. Am. **109**(5), 2146–2152.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–141.

Heinks-Maldonado, T. H., and Houde, J. F. (**2005**). "Compensatory responses to brief perturbations of speech amplitude," ARLO **6**(3), 131–137.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of f0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**(3), 1246–1251.

Jones, J. A., and Munhall, K. G. (**2002**). "The role of auditory feedback during phonation: Studies of Mandarin tone production," J. Phonetics **30**, 303–320.

Kawahara, H., and Williams, J. C. (**1996**). "Effects of auditory feedback on voice pitch trajectories: Characteristic responses to pitch perturbations," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, Sydney), pp. 263–278.

Kiran, S., and Larson, C. R. (**2001**). "Effect of duration of pitch-shifted feedback on vocal responses in Parkinson's disease patients and normal controls," J. Speech Lang. Hear. Res. **44**, 975–987.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Larson, C. R. (**1998**). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of *f*0 control," J. Commun. Disord. **31**, 489–503.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). "Effects of pitch-shift onset velocity on voice f0 responses," J. Acoust. Soc. Am. **107**(1), 559–564.

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., and Hain, T. C. (**2001**). "Comparisons of voice *f*0 responses to pitch-shift onset and offset conditions," J. Acoust. Soc. Am. **110**(6), 2845–2848.

Leentjens, A. F., Wielaert, S. M., van Harskamp, F., and Wilmink, F. W. (**1998**). "Disturbances of affective prosody in patients with schizophrenia; A cross sectional study," J. Neurol., Neurosurg. Psychiatry **64**(3), 375–378.

Luschei, E., Saslow, C., and Glickstein, M. (**1967**). "Muscle potentials in reaction time," Exp. Neurol. **18**, 429–442.

Murphy, D., and Cutting, J. (**1990**). "Prosodic comprehension and expression in schizophrenia," J. Neurol., Neurosurg. Psychiatry **53**(9), 727–730.

Natke, U., and Kalveram, K. T. (**2001**). "Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables," J. Speech Lang. Hear. Res. **44**, 577–584.

Natke, U., Donath, T. M., and Kalveram, K. T. (**2003**). "Control of voice fundamental frequency in speaking versus singing," J. Acoust. Soc. Am. **113**, 1587–1593.

Prochazka, A., Clarac, F., Loeb, G. E., Rothwell, J. C., and Wolpaw, J. R. (**2000**). "What do reflex and voluntary mean? Modern views on an ancient debate," Exp. Brain Res. **130**(4), 417–432.

Ramig, L. O. (**1994**). "Voice disorders," in *Introduction to Communication Sciences and Disorders*, edited by F. D. Minifie (Singular, San Diego), pp. 481–519.

Siegel, G., and Pick, H. L. Jr. (**1974**). "Auditory feedback in the regulation of voice," J. Acoust. Soc. Am. **56**(5), 1618–1624.

Siegel, G. M., and Kennard, K. L. (**1984**). "Lombard and sidetone amplification effects in normal and misarticulating children," J. Speech Hear. Res. **27**(1), 56–62.

Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (**2005**). "Voice responses to changes in pitch of voice or tone auditory feedback," J. Acoust. Soc. Am. **117**(2), 850–857.

Titze, I. R. (**1994**). *Principles of Voice Production* (Prentice-Hall, Englewood Cliffs, NJ).

Xu, Y., Larson, C., Bauer, J., and Hain, T. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**(2), 1168–1178.

Electropalatographic, acoustic, and perceptual data on adaptation to a palatal perturbation

Wendi A. Aasland and Shari R. Baum[a)]
School of Communication Sciences & Disorders, McGill University, Montreal, QC H3G 1A8, Canada

David H. McFarland
Ecole d'Orthophonie et d'Audiologie, Université de Montréal, Montreal, QC H3C 3J7, Canada

(Received 6 September 2005; revised 17 January 2006; accepted 17 January 2006)

Exploring the compensatory responses of the speech production system to perturbation has provided valuable insights into speech motor control. The present experiment was conducted to examine compensation for one such perturbation—a palatal perturbation in the production of the fricative /s/. Subjects wore a specially designed electropalatographic (EPG) appliance with a buildup of acrylic over the alveolar ridge as well as a normal EPG palate. In this way, compensatory tongue positioning could be assessed during a period of target specific and intense practice and compared to nonperturbed conditions. Electropalatographic, acoustic, and perceptual analyses of productions of /asa/ elicited from nine speakers over the course of a one-hour practice period were conducted. Acoustic and perceptual results confirmed earlier findings, which showed improvement in production with a thick artificial palate in place over the practice period; the EPG data showed overall increased maximum contact as well as increased medial and posterior contact for speakers with the thick palate in place, but little change over time. Negative aftereffects were observed in the productions with the thin palate, indicating recalibration of sensorimotor processes in the face of the oral-articulatory perturbation. Findings are discussed with regard to the nature of adaptive articulatory skills. [DOI: 10.1121/1.2173520]

PACS number(s): 43.70.Ep, 43.70.Bk [BHS] Pages: 2372–2381

I. INTRODUCTION

Exploring the compensatory responses of the speech production system to a variety of static and dynamic perturbations has provided valuable insights into a variety of fundamental processes, including central or predictive representations of speech movements and their interaction with sensory feedback (Löfqvist, 1997; Tremblay, Shiller, and Ostry, 2003). For the oral motor system, an adaptive response paradigm may provide an effective means of exploring many fundamental aspects of speech production, including the development of novel articulatory programs related to speech motor learning. We have completed a series of investigations designed to measure the adaptive responses of the speech motor control system to a variety of perturbations to oral form and function (Baum and McFarland, 2000; Baum, McFarland, and Diab, 1996; Baum and McFarland, 1997; McFarland, Baum, and Chabot, 1996). To induce a structural perturbation, we have employed an artificial palate with a substantial buildup of acrylic over the alveolar ridge, similar to that used initially by Hamlet and colleagues (Hamlet, Cullison, and Stone, 1979) and further modified for our experimental purposes (Baum and McFarland, 1997; McFarland, Baum, and Chabot, 1996). This artificial palate, when contrasted with one with a very thin acrylic covering and with no palate in place, significantly perturbs consonant production (in particular /s/ production) and may involve a rather lengthy adaptation period for normal articulatory profiles to be developed. Target intensive practice was found to accelerate the adaptation and make it reasonable to study compensatory processes within a short period of time (e.g., one hour) (Baum and McFarland, 1997). Results revealed significant improvements in /s/ production during the course of the adaptation period, with evidence of the formation of novel articulatory programs and sensorimotor recalibration of the articulatory space. Negative aftereffects were also observed, in which the accuracy of speech produced without the palate in place as well as with the thin appliance, was negatively impacted as a result of the target intensive practice with the modified palatal contour (Baum and McFarland, 1997).

These earlier investigations employed both acoustic and perceptual measures to make judgments about speech compensation and articulatory modification in response to oral articulatory perturbation. What was lacking was information regarding potential compensatory movements and contacts of the speech articulators. Such information is crucial for understanding motor control strategies related to the adaptation process. We have recently developed (Aasland, Baum, and McFarland, 2004a; b) a new technological modification of an existing electropalatographic system that allows for the tongue movement contact electrodes to be embedded in the modified palatal contour (Hamlet, Stone, and McCarty, 1978). In this way, compensatory tongue positioning can be recorded simultaneously with speech acoustics during the

a)Author to whom correspondence should be addressed: Shari R. Baum, Ph.D., School of Communication Sciences & Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec H3G 1A8, Canada. Electronic mail: shari.baum@mcgill.ca

course of adaptation to a significant perturbation to oral form. Such compensatory responses can be compared to standard EPG appliances that involve a very thin layer of acrylic that has been shown to be relatively less perturbing to speech production than the thick palatal appliance (Hamlet, Cullison, and Stone, 1979).

Electropalatography (EPG) is an extremely well-established measurement process that has been used both for understanding fundamental aspects of speech production and for clinical intervention to encourage appropriate tongue positioning for a variety of speech disorders, such as those associated with traumatic brain injury, velopharyngeal inadequacy/cleft palate, articulation disorders, deafness, and other communicative disorders (e.g., Flege, Fletcher, and Homiedan, 1988; Fletcher and Newman, 1991; Fletcher, McCutcheon, and Wolf, 1975; Gibbon, 2004; Gibbon and Wood, 2003; Goozée, Murdoch, and Theodoros, 2003; Parsloe, 1998). Tongue-palate contact patterns reflecting normal consonant and vowel articulation have been relatively well described using this technique and some studies have combined EPG analyses with more dynamic movement assessments such as by ultrasound (Stone, Faber, Raphael, and Shawker, 1992). Consequently, correlative data exist that allow for the referral of tongue-palate contact patterns to those related to tongue shape and cross-sectional area during speech sound production. The fricative /s/ (the target consonant of the current investigation) is articulated with an anterior tongue constriction and concave central groove that directs the airstream turbulence against the upper or lower incisors (Fletcher and Newman, 1991; Narayanan, Alwan, and Haker, 1995; Stevens, 1998; Stone *et al.*, 1992). Marginal lateral tongue contact and upward force vectors form a "lingualpalatal brace" for central groove configuration. Groove depth averages 5.6 mm but is highly speaker dependent (Stone *et al.*, 1992), as is the cross-sectional tongue configuration at the point of constriction, which varies from slit-like to more circular (Narayanan *et al.*, 1995). Midsagittal tongue shape is overall convex to concave, running anteriorly to posteriorly, but is also speaker dependent and is related to apical versus laminal articulation (Narayanan *et al.*, 1995).

The present investigation was designed to use the EPG measurement tool and these previous detailed analyses of normal fricative production to explore speech compensatory processes related to the presence of oral articulatory perturbations. Tongue-palate contact information was combined with detailed acoustic and perceptual measures (as in previous investigations) to provide a relatively comprehensive account of compensatory behaviors and potential individual differences in such behaviors. The overall goal is to understand speech motor control processes underlying the development and refinement of normal speech production. Such information is necessary for an increased understanding of a number of speech disorders and their potential remediation.

II. METHODS

A. Production experiment

1. Subjects

Nine young adult (19–27 years) native speakers of English served as participants. All reported a negative history of speech, language, and hearing impairments and none had any prior experience wearing a dental appliance. Subjects were screened for normal dental occlusion at the time of testing.

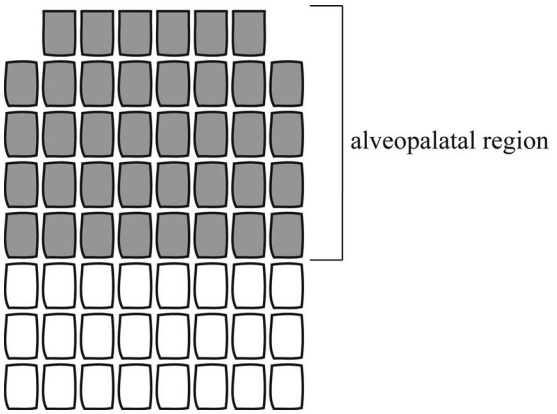


FIG. 1. Display of electrode array with the shaded portion illustrating the alveopalatal region. The anterior of the palate is at the top of the figure.

2. Stimuli and procedures

Two EPG artificial palates were constructed for each participant—one "normal" EPG palate, approximately 1 mm thick throughout ("thin" palate), with an array of 62 electrodes in keeping with the layout prescribed for Articulate Assistant EPG software (Wrench, 2003), and one "perturbed" EPG palate with a 6 mm buildup of acrylic at the alveolar ridge ("thick" palate), maintaining the same spatial distribution of the electrodes. These thick and thin[1] palates were identical in dimensions to those used in our previous investigations of compensatory responses to oral-articulatory perturbation and the reader is referred to these earlier studies for the artificial palate configurations (e.g., McFarland and Baum, 1995). The array includes seven rows of eight electrodes each, with the exception of the anterior-most row, which contains only six electrodes (see Fig. 1). Regions defining alveolar, palatal, and lateral borders are illustrated in the figure.

To assess the impact of the perturbations on compensatory tongue positioning, speakers were required to produce multiple repetitions of the bisyllable /asa/ under a variety of experimental conditions and at five time intervals. The target measured in each instance was the medial /s/. The target stimulus was first elicited[2] ten times with no palate in place to provide a baseline of normal /s/ production. Subsequently, ten repetitions of /asa/ were produced first with the thin palate in place, followed by ten repetitions with the thick palate in place. This entire procedure was then repeated for a total of 20 exemplars in each condition at Time 0. As in our previous investigations, stimuli were elicited every 15 min for a period of one hour during which speakers read /s/-laden passages with the thick palate in place (see, e.g., Baum and McFarland, 1997). At each 15 min interval (Times 15, 30, 45, 60), two blocks of ten repetitions of /asa/ were elicited with the thick palate in place, alternating with two blocks of ten repetitions with the thin palate in place. After Time 0, productions were elicited first with the thick palate, which was *already* in place, to ensure that speakers maintained any

compensatory adjustments developed during the practice interval. At Time 60, an additional two blocks of ten repetitions of /asa/ were elicited with no palate in place to examine any potential aftereffects of the 1 h palatal modification. Stimuli were recorded directly onto the computer using the Articulate Assistant software (Wrench, 2003) to ensure synchronization of the acoustic and EPG signals.

3. Acoustic analyses

Following earlier studies (Baum and McFarland, 1997; McFarland *et al.*, 1996), we determined three acoustic characteristics: (1) /s/ duration, measured from the onset to the offset of visible frication noise in the waveform display; (2) centroid frequency (or the first moment of the spectral distribution) at the /s/ midpoint, representing a weighted average of spectral peak frequencies; and (3) standard deviation of the spectrum (or the second moment of the spectral distribution).

4. EPG analyses

The goal of the EPG analyses was to provide information of compensatory tongue contact positioning in the face of oral-articulatory perturbation during /s/ production. Measures were based on previous assessments of positioning during normal and disordered articulation and were modified for the current experimental context. We began with a very large number of EPG contact measures computed at various points and in various palatal regions. As the majority of these measures were quite comparable, we report here only measures based on contact across the entire palate at the (temporal) point of maximum contact during /s/ production, as well as contact measures within the alveopalatal region at the point of maximum contact (when they differ from those for the whole palate). These two regions of interest were found to capture compensatory tongue positioning.

A measure of *whole contact* was computed, reflecting the percentage of sensors contacted averaged across all time frames within the maximum contact region. Similarly, a measure of *alveopalatal contact* was calculated, reflecting the percentage of sensors contacted within the five anterior-most rows of electrodes. A *variability index*, designed to reflect the stability of articulatory gestures, was calculated by computing the percent frequency of activation of each contact across relevant frames. Activation frequencies of 0% or 100% reflect stability and are associated with an index of 0; as contact frequency approaches 50% (i.e., highly inconsistent), a maximum variability index of 50 is assigned. A *center of gravity index* was also computed, indicating whether primary contact is more anterior or more posterior across the whole palate. Finally, a *laterality index* was computed, reflecting the degree to which average contact was more toward the midline or the outer sides of the palate; higher values indicate more lateral contact and provide an indication of groove width for /s/.

B. Perception experiment

1. Subjects

The listeners were ten native speakers of English (aged 20–29) who were phonetically naïve and exhibited no history of speech, language, or hearing impairment. None of the listeners had participated as speakers in Experiment 1.

2. Stimuli and procedure

Recordings from five speakers chosen at random from the production experiment served as stimuli for the perception experiment. (Use of the entire stimulus set from all speakers would have rendered the experiment too long.) For each speaker, a single clear production[3] of /asa/ was selected from those stimuli originally produced with no palate in place at Time 0 (i.e., before the start of the experimental manipulations) to serve as a "typical" exemplar. These typical exemplars were used as a point of comparison against which listeners were asked to rate the quality of /asa/ productions from the thin and thick palate conditions, as well as those from the no-palate condition recorded at Time 60.

The 90 tokens from each of the 5 speakers were included as follows: 10 tokens each (selected at random) from the thin and thick palate conditions at Times 0, 15, 30, 45, and 60, as well as 10 tokens from the no-palate condition at Time 60 (as noted above). Listeners were prompted prior to each trial with a visual "get ready" cue on a computer monitor, followed by a 100 silent interval. The typical exemplar for the individual speaker was then presented over closed headphones, followed after 750 ms by an experimental stimulus from that speaker. Listeners were instructed to rate the second stimulus, focusing solely on the /s/, on a scale from "1" (very poor quality /s/) to "5" (very good quality /s/, equivalent to typical exemplar). Five practice trials representing variations in stimulus quality (as judged by the experimenters) preceded the experiment. If the listener appeared not to have understood the instructions on the basis of results of the practice trials, instructions were repeated and the practice stimuli rerun. Trials were presented in a fixed random order, divided into three blocks to permit two breaks during the 45 min perception experiment.

III. RESULTS

A. Acoustic analyses

Mean centroid frequencies (collapsed across speakers) computed at each time period in the thin and thick palate conditions, as well as those calculated for productions with no palate in place at Times 0 and 60, are displayed in Fig. 2. As may be observed from the graph, upon first insertion of either palate, but notably the thick palate, centroid frequencies are substantially lower than in the no palate condition. However, within 15 min of practice with the thick palate in place, that pattern is reversed for the thick palate and remains that way throughout the practice period. These observations are confirmed by a Time (T0, T15, T30, T45, T60) $\times$ Palate (N,P) ANOVA, which yielded main effects of Time $[F(4,32)=5.195, p<0.002]$ and Palate $[F(1,8)=9.288, p<0.02]$, and a Time $\times$ Palate interaction $[F(4,32)$

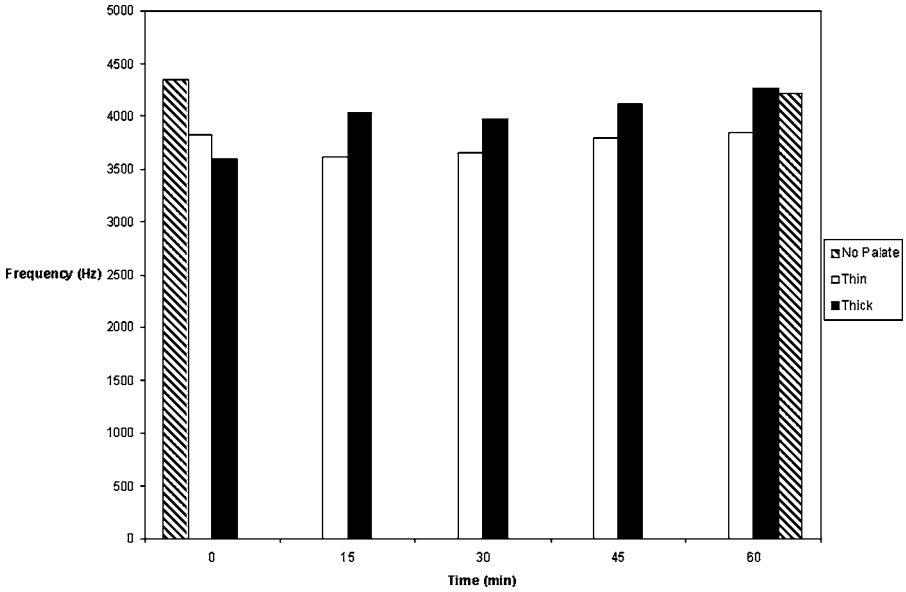


FIG. 2. Mean centroid frequencies over all time periods for the thin and thick palate conditions, and at times 0 and 60 min for the no-palate (NP) condition.

$=16.495, p<0.001$]. Post hoc analysis of the interaction using the Newman-Keuls procedure ($p<0.05$) confirmed that at Time 0, centroids in the thin condition were significantly higher than in the thick condition, whereas at all other time periods, centroids in the thick condition were significantly higher than in the thin condition. Comparing within conditions over time, centroids in the thin condition were significantly higher at Time 0 relative to Times 15 and 30 only. Within the thick palate condition, centroids at Time 0 were significantly lower than at Times 15, 30, and 45, which did not differ from one another; centroids at Time 60 were significantly higher than at all other times.

A separate ANOVA examined centroid frequencies in the no-palate and thin and thick conditions at Times 0 and 60 only. This Time $(0,60)\times$Palate (no palate, thin, thick) ANOVA revealed a significant main effect of Palate [$F(2,16)=16.734, p<0.001$] and a Time $\times$ Palate interaction [$F(2,16)=14.793, p<0.001$]. Post hoc analysis of the interaction yielded significant differences between the no-palate and both thin and thick palate conditions at Time 0; the thin and thick conditions did not differ from one another. In contrast, at Time 60, centroids in the thin condition were significantly lower than in both the thick and no-palate conditions, which did not differ from one another. (No differences were found in the no-palate condition over time.) However, an examination of the data for individual speakers at Time 60 revealed that one speaker in particular produced /s/ in the thick palate condition with very little change over time, suggesting that this individual failed to compensate for the perturbation despite the lengthy practice period. In addition, two of the speakers appear to have "overcompensated," producing centroids in the thick condition at Time 60 that were substantially higher than those in the no-palate condition.[4] Importantly, all speakers displayed the group pattern of lower centroid frequencies in the thin palate condition at the end of the practice period, as at the outset.

With respect to the measure of standard deviation [SD], mean SD measures in all conditions at all time periods are plotted in Fig. 3, illustrating a quasilinear increase in the standard deviation over time. A Time $\times$ Palate ANOVA yielded only a main effect of Time [$F(4,32)=8.552, p<0.001$], with significant increases in SD measures between Time 0 and Times 45 and 60, as well as between Time 15 and Times 45 and 60, and between Time 30 and Time 60. The Time (T0, T60) $\times$ Palate (no palate, thin, thick) ANOVA comparing no palate and palate conditions again revealed a main effect of Time only [$F(1,8)=7.593, p<0.05$], in keeping with the pattern described above. One possible interpretation of the SD measure is that it reflects the length of the constriction formed or the involvement of tongue blade (versus the tongue tip) in generating a constriction, with increased SD reflecting a longer (or more laminal) constriction (see, e.g., Stevens, Keywer, and Kawasaki, 1985). Despite the overall group pattern suggesting an increase in SDs over time, though, there was a great deal of variability in the patterns for individual speakers, rendering the group finding questionable.

Mean frication duration measures are plotted in Fig. 4. A Time $\times$ Palate ANOVA yielded only a main effect of Time [$F(4,32)=5.027, p<0.003$], with significantly longer durations at Time 60 relative to Times 0, 15, and 30; durations at Time 45 and Time 0 also differed significantly. An ANOVA comparing the no palate and palate conditions at Times 0 and 60 only revealed the main effects of Time [$F(2,16=4.945, p<0.05$] and Palate [$F(1,8)=9.657, p<0.02$], as well as a Time $\times$ Palate interaction [$F(2,16)=5.375, p<0.02$]. Post hoc analysis of the interaction demonstrated that at Time 0, fricative duration in the no-palate condition was significantly shorter than in both thin and thick conditions, which did not differ. This pattern was demonstrated by all but a single speaker—the same individual for whom centroid frequencies did not change much over the practice period. At Time 60, no duration differences between palate conditions remained (although there was a good deal of interspeaker variability); further, duration measures were longer at Time 60 than at Time 0 in all conditions, without concomitant increases in overall syllable durations.

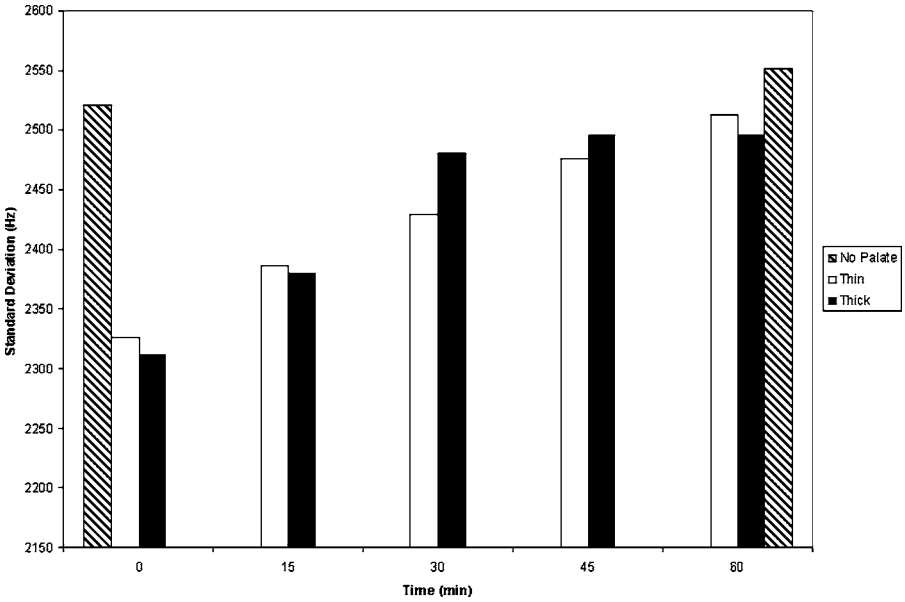


FIG. 3. Mean standard deviation (the second moment of the spectral distribution) at all time periods for the thin and thick palate conditions and at times 0 and 60 min for the no-palate (NP) condition.

B. EPG analyses

The average percentage of sensors contacted within the (temporal) region of maximum contact in each condition at each point in time is illustrated graphically in Fig. 5. A Time × Palate ANOVA yielded a Time × Palate interaction $[F(4,32)=4.303, p<0.01]$, confirming the patterns apparent in the graph; that is, at Time 0, contact patterns for both thin and thick palates were comparable (as confirmed by post hoc analyses), whereas at all other times, there was less overall contact in the thin relative to the thick condition. In an examination of individual data, these patterns held for the majority of speakers with a single exception; one speaker exhibited increased contact in the thin relative to the thick palate condition throughout the experiment.[5] EPG displays for a typical and the exceptional speaker are provided in Figs. 6(a) and 6(b). The contact patterns for the exceptional speaker show the surprisingly increased contact in the thin palate condition, with relatively little change from Time 0 to Time 60.

Further, within the thin condition, contact values were significantly higher at Time 0 than at all other times, with the exception of Time 60. In contrast, within the thick palate condition, the extent of contact did not differ significantly over time. Analyses of contact patterns in the alveopalatal region revealed a largely similar pattern, with a single important exception. At Time 0, the extent of contact was substantially (although not significantly) lower in the thin compared to the thick palate condition and remained that way throughout the 60 min practice interval. This pattern is illustrated in Fig. 7 (for all speakers) and was exhibited by all but two individuals.

In terms of the indices computed using the Articulate Assistant software (Wrench, 2003), analysis of the Centre of Gravity (CoG) index (reflecting contact weighting in the anterior-posterior dimension) demonstrated that the thin condition yielded higher CoG measures than the thick condition at all time intervals, indicating a more anterior contact distribution. A Time × Palate ANOVA yielded only a significant

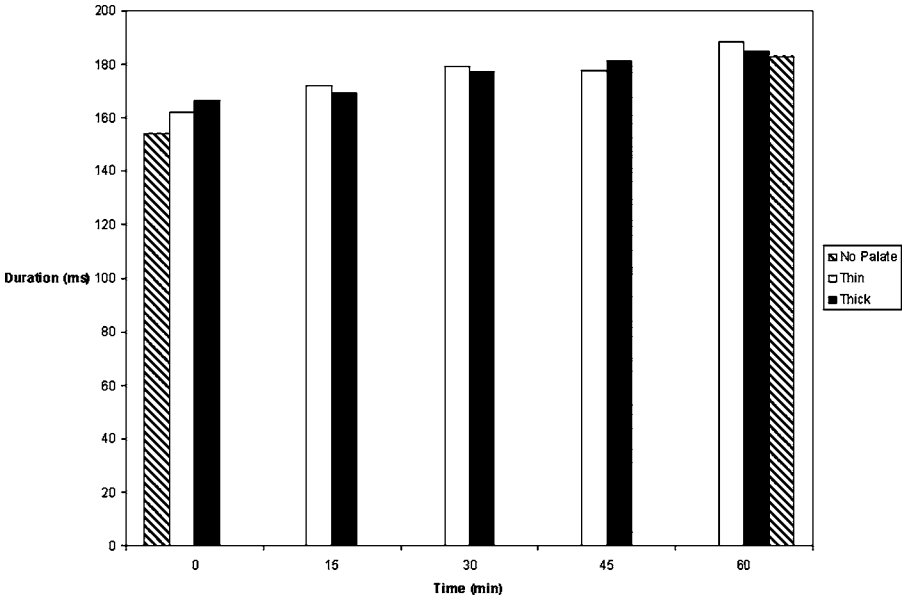


FIG. 4. Mean duration of frication in the /asa/ productions at all time periods for the thin and thick palate conditions, and at times 0 and 60 min for the no-palate (NP) condition.

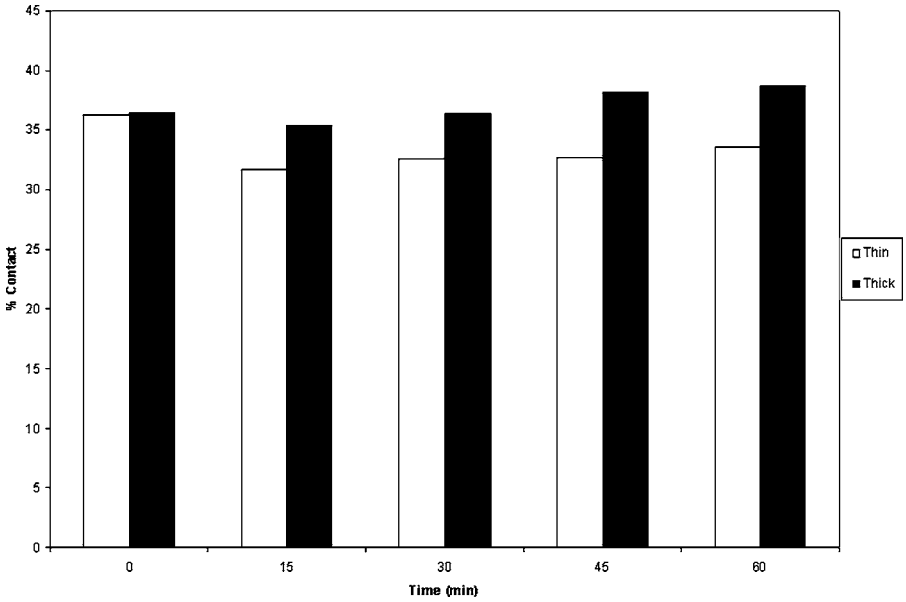


FIG. 5. Mean percentage of sensors contacted during the portion of the /s/ production with maximum tongue-palate contact.

main effect of Palate $[F(1,8)=5.316, p<0.05]$. In examining CoG indices for individual speakers, all but a single speaker (different from the individual "outlier" noted in the description of maximum contact patterns) generally conformed to the group pattern; this speaker displayed a consistently reversed pattern, with CoG measures in the thin condition lower than in the thick condition at all time intervals.

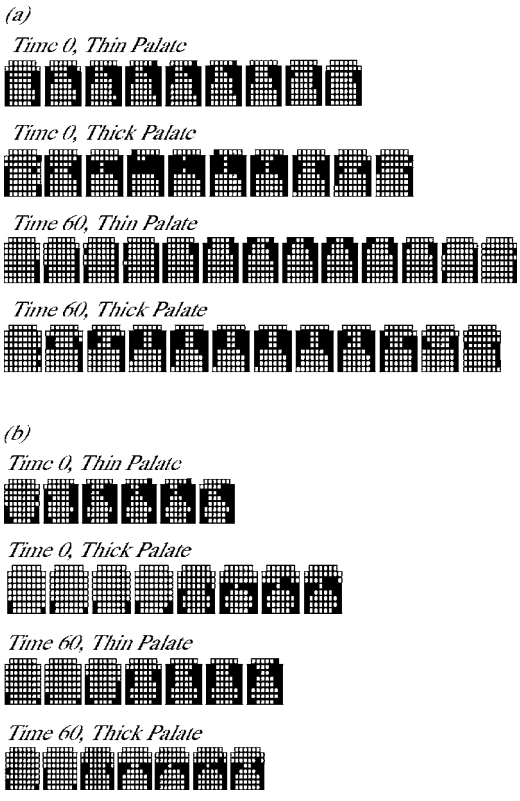


FIG. 6. Contact patterns for exemplar tokens produced by two individual speakers at Times 0 and 60 in both thin and thick palate conditions. The displays represent fricative contact, as defined by the onset and offset of acoustic evidence of frication noise in the waveform; contact patterns are represented in 20 ms intervals. The first four displays (a) are taken from a speaker who produced typical patterns of contact; the remaining displays (b) are taken from a speaker who produced aberrant patterns.

An analysis of the Laterality index also showed consistently higher values (indicating more lateral contact or a wider groove width) for the thin compared to the thick condition at all time intervals; this pattern held for all individual speakers. The Time × Palate ANOVA yielded main effects of both Time $[F(4,32)=4.873, p<0.005]$ and Palate $[F(1,8)=169.479, p<0.001]$. Post hoc analysis of the Time main effect revealed significantly higher Laterality index values at Times 15 and 30 relative to Time 0, collapsed across palate conditions.

Finally, a Time × Palate ANOVA on the Variability (V) index measures revealed only a Time × Palate interaction $[F(4,32)=3.844, p<0.02]$. Post hoc analyses showed that, although the patterns of variability in the two palate conditions changed somewhat over time (as illustrated in Fig. 8), none of the differences reached statistical significance.

C. Perceptual analyses

Mean perceptual quality ratings in all conditions at all time periods are illustrated in Fig. 9. It is quite clear that initial quality ratings for productions from the thick condition were substantially lower than those for the thin condition. In keeping with the acoustic data, that pattern was reversed at Time 15; after 60 min of practice, quality ratings for both thin and thick conditions were comparable and higher than at Time 0, but still lower than ratings for stimuli produced with no palate in place. A Time × Palate ANOVA revealed a main effect of Time $[F(3,27)=26.048, p<0.001]$ and a Time × Palate interaction $[F(3,27)=55.099, p<0.001]$. Post hoc analysis of the interaction demonstrated that at Time 0, ratings for the thin palate condition were significantly higher than for the thick palate condition and that pattern was reversed at Time 15. At other time periods, the two palate conditions did not differ. Within the thin palate condition, ratings got significantly worse from Time 0 to Time 15, whereas within the thick palate condition, ratings improved significantly from Time 0 to Time 15. In the thin palate condition, from Time 15 to Time 30, ratings

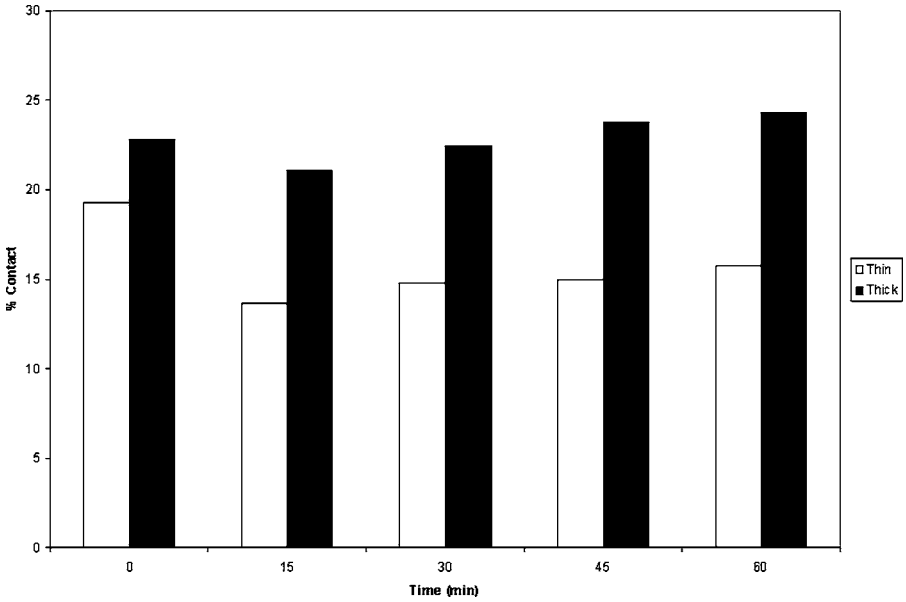


FIG. 7. Percentage of contact in the alveopalatal region (the five most anterior rows of sensors) during the period of maximum contact.

improved significantly. A second ANOVA examining the three palate conditions at Time 60 alone revealed a significant main effect of Palate $[F(2,18)=56.113.p<0.001]$. Post hoc analysis confirmed significantly higher ratings in the no-palate condition relative to both thick thin palate conditions, which did not differ from one another. An examination of individual data revealed that all listeners exhibited all of the group patterns.

IV. DISCUSSION

The present investigation was designed to explore speech compensatory processes related to the presence of a rather significant perturbation of oral form using a triangulation of measures involving acoustic, perceptual, and electropalatographic analyses. Electropalatographic electrodes were embedded in both thin and thick palatal appliances to measure compensatory tongue positioning during a period of intense, target-specific practice.

A. Acoustic and perceptual findings

Target-specific practice appeared to facilitate the development of novel articulatory programs in the face of the perturbation, as revealed by both the acoustic and perceptual profiles. Consistent with our previous findings (Baum and McFarland, 1997; McFarland, Baum and Chabot, 1996) and those of others (Hamlet, Cullison, and Stone, 1979), the thick artificial palate was highly perturbing to /s/ production, with centroid frequencies and quality ratings of /s/ that were significantly lower for this palatal condition in the initial test interval when contrasted with the no-palate and thin palate conditions. (Perturbing effects of the thin palate were also evident, as discussed elsewhere.) Acoustic and perceptual accuracy of the /s/ production with the thick appliance in place increased over the practice interval, and after 60 min no significant differences were found between the thick and no-palate conditions in /s/ centroid values. This appeared to be related to two factors: an increase in mean centroid values

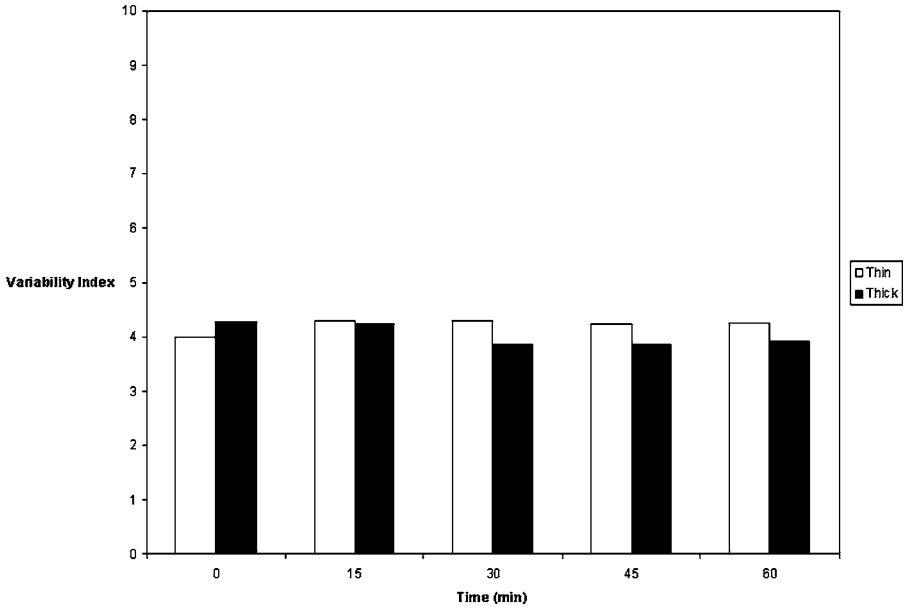


FIG. 8. Mean variability index (ranging from 0–50) during the period of maximum contact. (Higher values indicate greater variability within productions.)

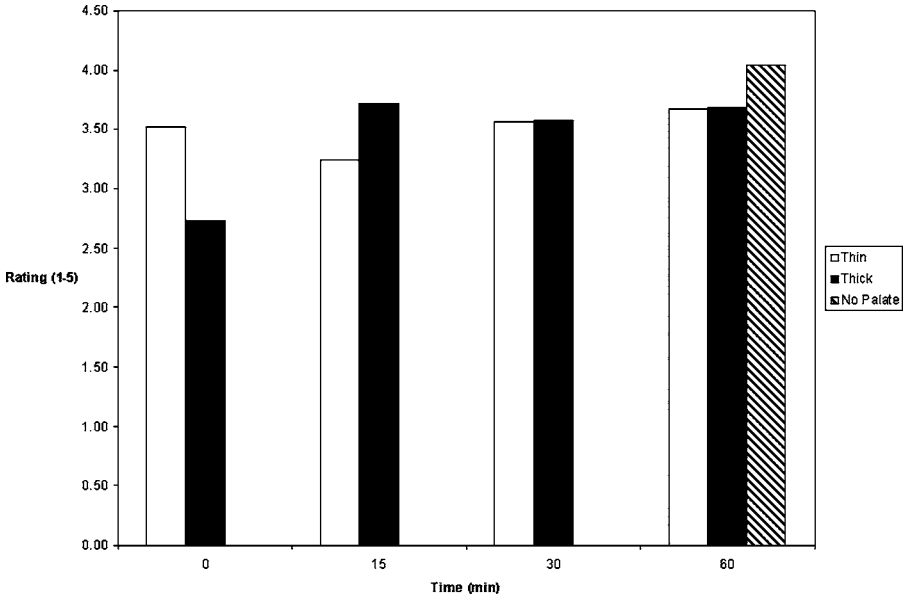


FIG. 9. Mean rating of /s/ quality by listeners, on a scale from 1 (very poor quality /s/) to 5 (very good quality /s/).

associated with the thick palate over time, and a (nonsignificant) decrease in centroid values with no palate in place. Similarly, longer /s/ durations were observed across palatal conditions, including the no-palate condition after 60 min of practice when contrasted to initial measurement periods. The decrease in mean centroid values and increased utterance durations associated with the no-palate condition might be indicative of the overall recalibration of the articulatory profiles associated with /s/ production and distributed effects of the novel articulatory profiles appropriate to the change in oral form associated with the thick palate. In other words, these changes in production with no palate in place may have represented negative aftereffects of the target-specific practice with the thick palate in place. Additional evidence of negative aftereffects comes from the thin palate data. After 15 min of practice with the thick appliance in place, mean centroid values and quality ratings were lower for the thin palate than those computed during the initial measurement period. Negative aftereffects are indicative of the development of new speech articulatory profiles that are appropriate for the change in oral form and function (Jones and Munhall, 2000). Similar effects have been observed in other motor systems, and have been particularly well studied in visual-motor control (e.g., van der Kamp, Bennett, Savelsbergh, and Davids, 1999) and gait control systems (e.g., Anstis, 1995). The fact that such effects can be observed in compensatory speech posturing suggests that similar motor control processes are operating in rather diverse sensorimotor systems.

B. EPG findings

Several inter-related measures of tongue contact were made that supplemented the acoustic and perceptual results highlighted above. Increased maximum contact, posterior tongue positioning, and more medial contact patterns were observed for the thick as contrasted to the thin palatal appliance, although individual differences were apparent. These tongue contact patterns indicate potential tongue overshoot and flattening in response to the buildup of acrylic over the alveolar ridge (see Narayanan *et al.*, 1995; Stone *et al.*, 1992). In turn, there is likely to be a narrower and decreased depth of tongue groove due to the presence of the buildup of acrylic over the alveolar ridge. This in turn seems likely to have influenced turbulence generation and consonant spectral characteristics. Consequently, /s/ production under these conditions yielded lower centroid frequencies and lower quality ratings. Previous combined EPG and ultrasound studies have revealed that /s/ is normally associated with relatively marginal lateral lingual palatal contact (when contrasted with /š/, for example) (Stone *et al.*, 1992). Increased medial tongue contact under conditions of perturbation may be related to mechanical compression against the acrylic buildup (Stone *et al.*, 1992). It might also represent a compensatory posturing of the tongue in response to the change in oral form. Previous kinematic studies have revealed the relatively marginal lateral tongue palate contact is importantly involved in the generation of "lingualpalatal bracing" necessary for appropriate tongue grooving (Flege *et al.*, 1988; Stone *et al.*, 1992). Perhaps increased medial contact allowed for compensatory changes in tongue shape and the redirection of the airstream, which resulted in overall improvements in fricative spectral characteristics and quality over the course of the practice interval. That is, there may have been other changes in tongue shape—for example, midsagittal groove dimensions (Fitzpatrick and Ní Chasaide, 2002)—supported by the compensatory tongue positioning, that would not have been revealed by our EPG measures and would require direct assessments of tongue shape related to groove configuration such as those provided by ultrasound, MRI or EMA (e.g., Cheng, Goozée, and Murdoch, 2005; Fitzpatrick and Ní Chasaide, 2002; Hoole, Nguyen-Trong, and Hardcastle, 1993; Narayanan *et al.*, 1995; Stone *et al.*, 1992). Decreased tongue contact was observed with the thin palatal appliance over time, suggesting rather rapid compensation to this relatively small perturbation to an oral-articulatory structure. Similar findings of rapid compensation were observed in our previous investigations of the thin versus thick appliances using non-EPG measurement proce-

dures (McFarland, Baum, and Chabot, 1996). The subtle changes observed in the acoustic and perceptual data indicating negative aftereffects on the thin palate condition were less apparent in the EPG patterns, again arguing for an additional measurement of tongue shape or movement.

Overall, the combined measurement systems used in the present investigation revealed evidence of the development of compensatory articulatory programs after an intense, but relatively short duration target-specific practice. We have previously suggested that such compensation may represent a sensorimotor recalibration and the development of novel speech motor programming appropriate for the change in oral form and function (Baum and McFarland, 1997). This interpretation is supported by additional evidence of negative aftereffects, where the new articulatory profiles were apparently applied to nonperturbed productions and resulted in disrupted speech production, even without the palate in place. It seems logical to assume that sensory feedback would be relatively crucial in the development of these novel programs and may include information related to tongue positioning/tongue contact as well as the acoustic (and perceptual) consequences of the novel articulatory gestures. The fact that speech acoustic parameters associated with /s/ production improved significantly during the practice interval, while tongue contact patterns remained relatively constant, may be interpreted as suggesting that speakers are relying on auditory or perceptual goals as contrasted to those related to kinesthetic feedback. Alternatively, it could be argued that kinesthetic feedback was involved in modulating articulatory gestures and in particular groove shape or tongue configuration, but these were not apparent to our levels of analysis.

The fact that learning appears to occur in a relatively brief period with target-specific practice, despite the highly perturbing nature of the oral articulatory appliance, supports the utility of this type of experimental paradigm to explore these types of fundamental aspects of speech production. As in our previous investigations, there was a good deal of individual variability in the degree of adaptation, with some speakers exhibiting relatively quick compensatory adjustments and others (at least one) who appeared to show no evidence of compensation to the presence of the artificial palates, even after the 1 h practice interval. In our previous investigations, this one-hour practice interval was sufficient to allow at least some speakers to adapt to the presence of an artificial palate (Baum and McFarland, 1997; 2000). Several factors may account for these individual differences, including differences in palatal morphology, tongue motility, and speakers' perceptual sensitivity. With regard to the latter factor, Perkell and colleagues (2004) have recently provided data suggesting that the production strategies utilized by individual speakers depend to a great extent on their discriminative sensitivity for the specific speech contrast under investigation. In future studies, it will be interesting to examine the perceptual abilities of individual speakers in relation to their speech adaptation patterns as well as potential anatomical constraints on compensatory abilities. Clearly, individual compensatory strategies are of great interest in understanding motor control strategies related to the development and perhaps refinement of oral articulatory profiles (Flege *et al.*, 1988). The fact that there are apparently sometimes rather large differences in articulatory contact and configuration in response to oral perturbation might have important clinical implications as well, particularly given the fact that EPG measures are often used to remediate deviant vocal tract form and function such as that related to velopharyngeal insufficiency. Comparing subjects to idealized "normal" contact patterns may not be the most appropriate technique and may in fact undermine individual compensatory strategies. In fact, recent EPG investigations involving remediation of speech-language disorders have emphasized the importance of considering individual characteristics in speech production abilities as a predictive function influencing treatment success (e.g., Carter and Edwards, 2004; Gibbon, 2004).

The findings of the present investigation confirm and extend those of our previous investigations (Baum and McFarland, 1997; McFarland, Baum, and Chabot, 1996), indicating substantial flexibility within the speech production system, permitting the majority of speakers to adapt to a significant oral-articulatory perturbation with (relatively) minimal practice. Interestingly, in contrast to the acoustic and perceptual results, the EPG data revealed few changes in tongue-palate contact over time, suggesting that compensatory adjustments for /s/ production were made by tongue shape changes in regions that did not contact the palate. Future investigations using methods that capture tongue contour more directly will elucidate precisely what those adaptive strategies may be and how they may vary across individual speakers.

ACKNOWLEDGMENTS

This work was supported by a grant from the Natural Sciences & Engineering Research Council of Canada (NSERC) to the second author, and a scholarship from the Fonds de la Recherche en Santé du Québec (FRSQ) to the first author. We are indebted to Meg Grant for her assistance in data analysis and manuscript preparation.

[1]The thin palate presumably reflected an unperturbed state and was required to record EPG contact patterns.

[2]Prior to any experimental recording, subjects inserted the thin palate and produced the phrase "I said 'ahh' doctor" five times. This was done to ensure that during normal vowel production for these speakers, no electrodes were consistently contacted; had any electrodes been contacted during /a/, they would have been removed from analyses of contact patterns for /s/. This proved unnecessary.

[3]Any production produced with no palate in place at Time 0 could be included unless precluded by noise in the recording or hesitation on the part of the speaker.

[4]Importantly, means computed excluding these three individuals do not differ substantially from those for the group as a whole, as illustrated in the figure. In the no-palate condition, with the exclusion of these speakers, mean centroids were 4439 and 4189 Hz at Times 0 and 60, respectively; in the thin palate condition, values were 3893 and 3812 Hz at Times 0 and 60; in the thick palate condition, centroids were 3723 and 4172 Hz at Times 0 and 60, excluding these three speakers.

[5]Interestingly, this was not the same speaker who appeared to be an "outlier" based on the acoustic analyses.

Aasland, W., Baum, S., and McFarland, D. (2004a). "Adaptation to structural modifications of the vocal tract during speech: Electropalatographic and acoustic measures," paper presented at *From Sound to Sense: Fifty +Years of Discoveries in Speech Communication*, Cambridge, MA.

Aasland, W., Baum, S., and McFarland, D. (2004b). "Adaptation to structural modifications of the human vocal tract during speech: Electropalatographic measures," paper presented at *Acoustical Society of America*, New York, NY.

Anstis, S. (**1995**). "Aftereffects from jogging," Exp. Brain Res. **103**, 476–478.

Baum, S., and McFarland, D. (**1997**). "The development of speech adaptation to an artificial palate," J. Acoust. Soc. Am. **102**, 2353–2359.

Baum, S., and McFarland, D. (**2000**). "Individual differences in speech adaptation to an artificial palate," J. Acoust. Soc. Am. **107**, 3572–3575.

Baum, S., McFarland, D., and Diab, M. (**1996**). "Compensation to articulatory perturbation: Perceptual data," J. Acoust. Soc. Am. **99**, 3791–3794.

Carter, P., and Edwards, S. (**2004**). "EPG therapy for children with long-standing speech disorders: predictions and outcomes," Clin. Ling. & Phon. **18**, 359–372.

Cheng, H. Y., Goozée, J. V., and Murdoch, B. E. (**2005**). "Analysis of articulatory dynamics in dysarthria following brain injury in childhood using electromagnetic articulography and electropalatography," J. Med. Sp.-Lang. Path **13**, 15–35.

Fitzpatrick, L. and Ní Chasaide, A. (**2002**). "Estimating lingual constriction location in high vowels: a comparison of EMA- and EPG-based measures," J. Phonetics **30**, 397–415.

Flege, J. E., Fletcher, S. G., and Homiedan, A. (**1988**). "Compensating for /s/ and /ʃ/ production: Palatographic, acoustic and perceptual data," J. Acoust. Soc. Am. **83**, 212–228.

Fletcher, S. G., and Newman, D. G. (**1991**). "[s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width," J. Acoust. Soc. Am. **89**, 850–858.

Fletcher, S. G., McCutcheon, M. J., and Wolf, M. B. (**1975**). "Dynamic palatometry," J. Speech Hear. Res. **18**, 812–819.

Gibbon, F. E. (**2004**). "Abnormal patterns of tongue-palate contact in the speech of individuals with cleft palate," Clin. Ling. & Phon. **18**, 285–311.

Gibbon, F. E., and Wood, S. E. (**2003**). "Using electropalatography (EPG) to diagnose and treat articulation disorders associated with mild cerebral palsy: a case study," Clin. Ling. & Phon. **17**, 365–374.

Goozée, J. V., Murdoch, B. E., and Theodoros, D. (**2003**). Electropalatographic assessment of tongue-to-palate contacts exhibited in dysarthria following traumatic brain injury: Spatial characteristics," J. Med. Sp. Lang. Path. **11**, 115–129.

Hamlet, S., Cullison, B., and Stone, M. (**1979**). "Physiological control of sibilant duration: Insights afforded by speech compensation to dental prostheses," J. Acoust. Soc. Am. **65**, 1276–1285.

Hamlet, S., Stone, M., and McCarty, T. (**1978**). "Conditioning prostheses viewed from the standpoint of speech adaptation," J. Prosthet. Dent. **40**, 60–66.

Hoole, P., Nguyen-Trong, and Hardcastle, W. (**1993**). "A comparative investigation of coarticulation in fricatives: Electropalatographic, electromagnetic, and acoustic data," Lang Speech **36**, 235–260.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Löfqvist, A. (**1997**). "Theories and models of speech production," in *The Handbook of Phonetic Sciences*, edited by W. J. Hardcastle and J. Laver (Blackwell Publishers, Inc., Cambridge, MA).

McFarland, D., and Baum, S. (**1995**). "Incomplete compensation to articulatory perturbation," J. Acoust. Soc. Am. **97**, 1865–1873.

McFarland, D., Baum, S., and Chabot, C. (**1996**). "Speech compensation to structural modifications of the oral cavity," J. Acoust. Soc. Am. **100**, 1093–1104.

Narayanan, S., Alwan, A., and Haker, K. (**1995**). "An articulatory study of fricative consonants through magnetic resonance imaging," J. Acoust. Soc. Am. **98**, 1325–1347.

Parsloe, R. (**1998**). "Use of speech pattern audiometer and the electropalatograph to explore the speech production/perception relationship in a profoundly deaf child," Int. J. Lang Commun. Disord. **33**, 109–121.

Perkell, J., Matthies, M., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., and Guenther, F. (**2004**). "The distinctiveness of speakers' /s/-/š/ contrast is related to their auditory discrimination and use of an articulatory saturation effect," J. Speech Lang. Hear. Res. **47**, 1259–1269.

Stevens, K. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, MA).

Stevens, K., Keywer, S., and Kawasaki, H. (**1985**). "Toward a phonetic and phonological theory of redundant features," in *Invariance and Variability in Speech Processes*, edited by J. Perkell and D. Klatt (Lawrence Erlbaum Associates, Hillsdale, NJ).

Stone, M., Faber, A., Raphael, L. J., and Shawker, T. H. (**1992**). "Cross-sectional tongue shape and linguopalatal contact patterns in [s], [], and [l]," J. Phonetics **20**, 253–270.

Tremblay, S., Shiller, D. M., and Ostry, D. J. (**2003**). "Somatosensory basis of speech production," Nature **423**, 866–869.

van der Kamp, J., Bennett, S. J., Savelsbergh, G. J., and Davids, K. (**1999**). "Timing a one-handed catch. II. Adaptation to telestereoscopic viewing," Exp. Brain Res. **129**, 369–377.

Wrench, A. (**2003**). *Articulate Assistant* (Articulate Instruments Ltd., Edinburgh).

On phonetic convergence during conversational interaction

Jennifer S. Pardo[a)]
Psychology Department, Barnard College, 3009 Broadway, New York, New York 10027

(Received 6 May 2005; revised 28 January 2006; accepted 30 January 2006)

Following research that found imitation in single-word shadowing, this study examines the degree to which interacting talkers increase similarity in phonetic repertoire during conversational interaction. Between-talker repetitions of the same lexical items produced in a conversational task were examined for phonetic convergence by asking a separate set of listeners to detect similarity in pronunciation across items in a perceptual task. In general, a listener judged a repeated item spoken by one talker in the task to be more similar to a sample production spoken by the talker's partner than corresponding pre- and postinteraction utterances. Both the role of a participant in the task and the sex of the pair of talkers affected the degree of convergence. These results suggest that talkers in conversational settings are susceptible to phonetic convergence, which can mark nonlinguistic functions in social discourse and can form the basis for phenomena such as accent change and dialect formation. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2178720]

PACS number(s): 43.70.Gr, 43.71.Bp, 43.70.Bk, 43.71.An [ARB] Pages: 2382–2393

I. INTRODUCTION

Speech is variable in its realization, both within and between talkers, despite apparent consistency in perception. Somehow, a listener is able to overcome the phonetically disparate productions of phonemes to arrive at what a talker intends to say. Recent studies of within and between-talker imitation have attempted to resolve the ongoing debate over whether speech perception yields acoustic, auditory, articulatory, gestural, or more abstract phonological parameters by positing a close connection between perception and production. In particular, such accounts propose that speech perception yields linguistically significant gestural parameters that automatically drive production, leading inevitably to imitation (such gestures are synergies of articulators; see Browman and Goldstein, 1991; Fowler *et al.*, 2003; Goldinger, 1998; Sancier and Fowler, 1997). A recent paper by Pickering and Garrod (2004) likewise proposes an automatic priming account for lexical, syntactic, and schematic parity in language use. The current study attempts to constrain such proposals by estimating the extent to which imitation in conversational settings is exact and automatic.

If the function that yields communicative parity is one in which perception of a linguistic unit automatically drives production of that unit, via priming or another cognitive mechanism, then imitative productions ought to match their models to some tolerance. There are at least two problems with such accounts. First, the question of matching depends crucially on which attributes are assessed, and in particular, the level of abstraction of the attributes. For example, when two talkers use the same word to designate an ambiguous figure, there is a match in relation to the lexicon; however, it is unlikely that two productions of the same word match at an acoustic-phonetic level. Indeed, phonetic imitation may be impossible to achieve in purely acoustic terms (Krauss and Pardo, 2004; see also Vallabha and Tuller, 2003, and Viechnicki, 2002). Even for a single talker, no two productions of the same phonetic segment are acoustically identical, nor do articulatory patterns match for the same phonemes under different speaking conditions (e.g., Perkell *et al.*, 2002). Therefore, at a phonetic level imitation is likelier to be graded and inexact, rather than perfectly compliant with the acoustic, articulatory, and phonetic detail of the spoken model for imitation. A satisfactory account of communicative parity must specify the particular dimensions that are relevant as well as the range of tolerance for mismatches to exact parity.

Second, an automatic mechanism of the kind discussed in the literature is characterized as an all-or-none obligatory function, with no processes intervening between perception and production. According to Chartrand and Bargh (1999), a "perception-behavior link posits the existence of a natural and nonconscious connection between the act of perceiving and the act of behaving, such that perceiving an action being done by another makes one more likely to engage in that same behavior" (p. 900). In the domain of spoken language, Pickering and Garrod (2004) claim that "as dialogue proceeds, interlocutors come to align their linguistic representations at many levels ranging from the phonological to the syntactic to the semantic. This interactive alignment process is automatic and only depends on simple priming mechanisms that operate at the different levels, together with an assumption of parity of representation for production and comprehension" (p. 188). That is, if perception yields parameters that automatically drive production, the straight line from perception to production is not modulated by intervening processes. Whatever parameters perception resolves, whether categorical or continuous, production must follow suit. In order to account for discrepancies between two instances of the same utterance (when they occur), such a model must assert that the differences occur as a result of error or noise in perception and/or production, and do not pattern according to other functions. At this point, such proposals do not adequately address the influence of processes

[a)]Electronic mail: jsp2003@columbia.edu

 0001-4966/2006/119(4)/2382/12/$22.50

outside of perception/production, the effects of which the current study attempted to evaluate. The next section reviews evidence of a close connection between perception and production, before moving to the main question addressed by this study: Whether ordinary circumstances of language use evoke phonetic convergence.

A. Perception-production link and imitation

Evidence for a close connection between speech perception and production spans a number of different time scales, from immediate shadowing of nonsense syllables through sentence syntax. At the briefest time scale, studies of simple versus choice response times in speech shadowing tasks show that speech perception yields targets for speech production very quickly (e.g., Fowler *et al.* 2003; Porter and Castellanos, 1980). In typical experimental investigations of response time, a participant must respond to a signal by performing a single response (in the simple condition) or by performing one of several alternative responses (in the choice condition). Fowler *et al.* note that choice response times exceed simple response times in similar tasks by 100 to 150 ms on average, and this difference is presumed to be due to an additional decision-making process in the choice condition. In studies of speech production response time, listeners hear a series of VCV utterances spoken by another talker that varies in the identity of the medial consonant. On each trial, a listener shadows the initial portion of the VCV utterance, and then responds to the consonant transition with either a single CV, or the CV that was produced by the model. Shadowers respond to the consonant transition in the choice condition almost as quickly as the simple condition (from 26 to 50 ms difference across different experiments; Fowler *et al.* 2003; Porter and Castellanos, 1980), presumably because the signal itself provides information for its articulation, obviating the need for a stage of decision-making requiring choice among abstract phoneme categories preliminary to articulating the response on each trial.

Fowler *et al.* (2003) took this paradigm a step further by showing that shadowers track subcategorical variability in consonant voicing by adjusting voice onset time (VOT) in their own shadowing responses toward those of the spoken models. However, it is important to note that shadowers did not match model VOTs precisely. Fowler *et al.* attribute this discrepancy to an influence of habitual action in speech production, and not to processes intervening between or superordinate to perception and production. A similar explanation was offered by Sancier and Fowler (1997) for a bilingual talker who varied her VOTs in both languages in the direction of her most recent language environment, but did not match the distribution in her second language. Accordingly, speech perception specifies gestural actions, and the direct link to speech production yields an imitative response with a moderate degree of fidelity. In contrast, Vallabha and Tuller (2004) explicitly asked talkers to imitate their own isolated steady-state vowels and measured the acoustic discrepancy between the sample and imitated versions. They found systematic biases (within each individual talker) in the discrepant repetitions that were not accounted for by models of random noise in perception or production. They pointed tentatively to dialectal differences among their talkers, which are not characterized by generic perceptual and productive models, as a likely source for the observed idiosyncrasies in the systematically biased imitations.

Although speech perception appears to specify gestural parameters that support rapid detailed shadowing, these studies find that talkers never match input signal properties, whether such signals were produced by others or by the talkers themselves. The shadowing paradigm is conducive to imitation, yet the acoustic output reflects both perceptual/productive limitations on fine-grain accuracy and the influence of other factors that induce directional biases in the discrepancies. As of yet, there is no account that explains how such factors intrude upon a presumably automatic and direct perception-production link.

At a slightly greater time scale, shadowing tasks also provide evidence for imitation of lexical items (Goldinger, 1998; Namy, Nygaard, and Sauerteig, 2002). In these studies, talkers were recorded producing words prompted from a list, and these items were then used as models in a shadowing task with different talkers. In order to determine whether shadowers imitated the models in their use of phonetic variants, independent listeners were recruited to provide perceptual judgments of imitative fidelity in the shadowed utterances. The motivation behind this methodological innovation is that perception integrates across multiple acoustic-phonetic dimensions, thereby providing a more configural assessment of imitation than a selected acoustic measure would. The listeners performed an AXB task in which they heard three versions of the same lexical items and judged which item produced by the shadowing talker, A or B (taken from the pretask and shadowed sessions), sounded like a better imitation of (Goldinger) or was more similar to (Namy *et al.*) the model's sample item, X. Listeners chose a shadowed item more often than a previously produced item as a better imitation of a sample item; however, performance was variable, perhaps indicating inconsistent degrees of imitative fidelity or inconsistencies in perceptual judgments of imitation. Some of the variability was accounted for by factors related to episodic memory (Goldinger) or to talker sex (Namy *et al.*).

As these studies demonstrate, a linked perceptual-productive system might produce convergence, if not a perfect imitative match, in the acoustic-phonetic and sublexical domains. Studies of language use in broader settings and time scales support the proposed link, but also hint at other influences on convergence that are not readily encompassed by strict amalgamation of perceptual/productive mechanisms. A central phenomenon identified by such approaches is an increase in similarity among linguistic components, a process variously termed *convergence, accumulating common ground*, or *alignment*. Over longer stretches of speech, interlocutors are known to converge in speaking rate (Giles, Coupland, and Coupland, 1991), subvocal frequency/amplitude contour (Gregory, 1990), and vocal intensity (Natale, 1975); to establish and increase common ground to the exclusion of over-hearers (Schober and Clark, 1989); and to align description schemes (Garrod and Doherty, 1994) and

syntactic constructions (Branigan, Pickering, and Cleland, 2000). In addition, a talker will attenuate or accentuate regional dialect expression in response to an interviewer's expressed attitude toward regional dialect or use of formal or regional dialect in British English, but phonetic similarity between talkers in these settings was not assessed directly (Bourhis and Giles, 1977; Giles, 1973). Considering these phenomena theoretically, Labov (1976, 1984) suggests that dialect formation and change result largely from opportunities for direct social contact among talkers and are influenced by social relationships between interacting talkers. Missing from this literature is an assessment of how conversational settings promote or hinder a putatively ineluctable tendency to imitate at a fine phonetic grain.

B. Measuring phonetic convergence

In order to examine phonetic variability in social interaction, the current study collected a conversational speech corpus and performed perceptual measures of phonetic convergence. The goal of this experiment was to determine whether pairs of talkers converged in phonetic repertoire over the course of a single conversational interaction. Phonetic convergence is an increase in segmental and suprasegmental similarity of the speech of one talker to another. Although sublexical imitation/similarity increases when a talker simply shadows another's speech at short latency in the laboratory (Fowler *et al.*, 2003; Goldinger, 1998; Namy *et al.*, 2002), the current experiment examined a situation that evoked this process in more natural communicative contexts.

The first part of the experiment collected samples of speech before, during, and after pairs of talkers interacted in a conversational task. To permit assessment of phonetic convergence, the task had to elicit the same lexical items spoken across partners. Moreover, these items had to be identified prior to the conversation in order to collect preinteractive tokens. These constraints were satisfied by the map task, which was developed by the Human Communication Research Center at the Universities of Glasgow and Edinburgh, Scotland (Anderson *et al.*, 1991; see Appendixes A and B for a sample pair of map task maps). The map task uses paired schematic maps that contain labeled illustrated landmarks: One member's map includes a path drawn from a starting point, around various landmarks, to a finishing point, and the companion's map contains only a starting point and landmarks. The goal of the task is for the talkers to communicate effectively enough so that the path on the first map, which cannot be seen by the holder of the pathless map, can be duplicated on the second map. Completion of the task requires active involvement of both participants, and spoken samples of the landmark labels can be collected both before and after the conversational interaction, to compare to those that are produced by both participants over the course of conversation. In addition, the task permits assignment of different social roles—one member is the instruction giver and the other is the instruction receiver.

The repetition of landmark label phrases from the map task between talkers enables assessment of phonetic convergence in an AXB perceptual test. In addition to using a conversational corpus to provide materials, the AXB test for convergence differed from previous research on two other points. First, this study restricted perceptual judgments to assessments of similarity in pronunciation. Pronunciation was chosen as a precise and readily accessible concept for untrained listeners to use. The listeners were encouraged to focus specifically on the way that the talkers were articulating the consonants and vowels, and to decide which item sounded more similar to the sample item in pronunciation. This specification was introduced because a listener who is asked explicitly to judge imitation or simply to judge similarity may focus on other nonphonetic attributes, such as the melodiousness of the vocal quality, or the apparent emotionality in the voices, or any idiosyncratic dimension; and this study is particularly concerned with changes in a talker's phonetic repertoire while interacting with another talker, regardless of a conscious intention to imitate. Second, prior assessments of the persistence of imitation used items produced after a delay of a few seconds from sample presentation (Goldinger, 1998). The current study used items produced immediately after the conversational interaction and not directly prompted by another talker's utterance. Although this method induced a longer interval between a sample and its repetition, the talkers in the current study interacted in a coordinated social setting, which may have produced more robust and persistent convergence (see Pickering and Garrod, 2004).

Evidence for phonetic convergence in this experiment would consist of finding that a talker's speech became more similar in pronunciation to the partner's speech than it was before the interaction. If a talker converged toward his/her partner, then the landmark label phrases spoken in response to a partner's utterance should sound more similar to that utterance than an item produced by the talker before or after the task. If phonetic variation is not tied to a particular setting, then all variants should sound equally like or unlike that of a conversational partner's utterances. This study also investigated the time course of convergence by measuring convergence early and late in the conversational session, and by assessing whether talkers persisted in convergence into the post-task session. In addition, Giles' communication accommodation theory (Giles, Coupland, and Coupland, 1991; Shepard *et al.*, 2001) predicts that talkers may have varied in degree of convergence depending on conversational role. For example, a giver's more dominant role could have led to greater convergence on the part of a receiver. It is also worth considering that the sex of the pair of talkers might have influenced the degree of phonetic convergence, such that female talkers might have converged more than male talkers, as found by Namy *et al.* (2002).

II. METHOD

A. Materials

1. Corpus elicitation

For the conversational task, each participant received a packet of five 8.5 by 11-in. sheets of paper printed with map task maps. The instruction giver had a set of five map task

maps with paths, and the instruction receiver had a similar set of maps without paths and used a pencil to draw the path on his/her map. Some of the map task landmark labels were modified from the original set to adjust for differences between British and American-English naming conventions; however, none of the iconic drawings or paths was changed from the originals. The pre- and post-task sessions' items were prompted with a packet of printed sheets listing the map task landmark labels and other lexical items in the following order: (1) A numbered set of map task landmark labels each to be spoken in the carrier phrase, "Number_is the_;" (2) Five randomized repetitions of the complete American English vowel set embedded in hVt context and filler words, each to be spoken in the carrier phrase, "Say_again;" and (3) A second numbered set of map task landmark labels each to be spoken in the carrier phrase, "Number_is the_." In both the pre- and post-task sessions, the talkers were encouraged to produce the sentences fluently in a normal speaking style. The vowel samples were collected for a separate study and served as fillers between the sets of landmark label phrases.

2. Convergence assessment

From the set of map task landmark labels that was repeated between both members of a pair, the listening test used samples of four items that were repeated across all six pairs of talkers—*abandoned monastery, green bay, walled city*, and *wheat field*. The selection of this set of items was guided by four design constraints: (1) Use of between-talker repetitions of conversational items that occurred within a relatively short period of time; (2) Exclusion of the first mentions of the items in the discourse, which have been found to be of longest duration, among other distinctions (Bard *et al.* 1991; Catchpole and Pardo, 2003; Fowler, 1988; Fowler and Housum, 1987; Fowler, Levy, and Brown, 1997; Krauss and Weinheimer, 1964); (3) Use of conversational items that were produced in clause- or sentence-final position, where the end of a clause was defined as the point where a breath or full pause occurred, in order to match the sentence-final position of the items produced in the pre- and post-task sessions; and (4) Use of landmark label phrases for items that were shared across the giver and receiver maps. As a result of these constraints, the items varied with respect to whether they were second or later mentions in the discourse (the factors distinguishing first from second and later mentions do not distinguish second from later mentions), and the items also varied in their discourse function (sometimes the repetitions were of the same type, sometimes of different types), but none of these considerations was confounded with the experimental variables considered in the analyses.

To test for differences in degree of convergence over the course of the interaction, for each pair, two of the repeated items were taken from early in the conversation and two were taken from later in the conversation. Items designated as early occurred prior to the halfway point in each pair's interaction, and late items occurred after that point. To assess effects of talker role, for each pair, two of the items were repeated from giver to receiver (GX repeated by receiver), and the other two were repeated from receiver to giver (RX repeated by giver; and also not confounded with timing). Finally, corresponding productions of the items from the pre- and post-task sessions of the appropriate talker served as competitors for the task repetitions in the AXB similarity test. In both nontask sessions, each talker produced the set of landmark labels twice—those items taken from the pretask session came from the second iteration of the set to use the most fluent and reduced productions possible, and those items taken from the post-task session came from the first iteration of the set to use productions from the briefest interval.

B. Procedure

1. Corpus elicitation

To provide speech samples, each talker sat at a desk in a sound-attenuated booth approximately 18 in. away from a desk-mounted dynamic microphone. All utterances were recorded onto analog cassette tapes via a Denon stereo cassette tape deck, which operated outside the booth. The utterances and biographical information collected in the pretask session were analyzed to determine pairings for the map task session. First, to avoid social dominance phenomena associated with mixed-sex pairs (see Bilous and Krauss, 1988; Namy *et al.*, 2002), this study employed same-sex pairs. Second, measures of average *F*0 in the hVd/t items were clustered to select pairs whose *F*0's were not exactly the same, but were proximal to each other in average *F*0. Goldinger's (1998) talkers were first compared in multidimensional scaling analyses of similarity ratings in order to maximize variability in his talker set. Although this method may be ideal for ensuring an evenly distributed set of talkers, the current study is concerned with more natural conversational settings, in which talker characteristics vary more freely. Therefore, the approach used here ensured that talkers differed in an acoustic attribute prior to contact, increasing the likelihood that some measurable difference would result without imposing a strict criterion on talker variability. Third, analyses of the filler words in the list verified that all talkers differentiated among vowels in the word sets, *marry/Mary* and *merry, cot* and *caught*, and *pen* and *pin*; however, the biographical information provided by the participants indicated that they were drawn from varied regional backgrounds. Thus, the talker set did not exhibit a homogeneous dialect; none of the talkers exhibited a strong regional accent; and, the pairs incidentally comprised individuals from different dialect regions. Finally, in none of the pairs were the members acquainted with each other prior to participating in the map task session.

The talkers returned between one and two weeks after the pretask session in same-sex pairs to participate in the map task session and to provide post-task speech samples. For the map task session, the talkers sat at identical desks in the same sound-attenuated booth and were separated by a plywood divider that prevented them from seeing each others' maps, bodies, and faces. To permit measurements of between-talker repetition latency, one talker's microphone recorded onto the left channel of a tape, and the other recorded onto the right channel. Prior to beginning, each talker

in a pair was assigned a role for the duration of the map task session—one talker was designated the instruction giver, and the other talker was the instruction receiver. However, both participants were encouraged to converse in order to complete the maps as accurately as possible. The ordering of the maps was varied across pairs. Once a session began, a pair of talkers performed the map task pairs in order until they completed all five; most pairs spent 30–60 min on the task session. At the end of a map task session, the talkers provided speech samples for the post-task session separately.

2. Convergence assessment

Because the perceptual measure of phonetic convergence requires lexical repetition, the recordings were first coded for timing of each instance of use of a map task label phrase by each talker. In addition, the accuracy and duration of completion of each map in the map task were measured. To determine accuracy, a similar technique was used as that of Anderson *et al.* (1991): A transparency of each of the giver's maps was created with a 1- by 1-cm grid superimposed, and then placed over the appropriate receiver's map. Accuracy was tallied as the proportion of the number of grid cells that a receiver's path duplicated a giver's path to the total number of grid cells on a giver's path. Descriptive analyses of these data from the map task session were conducted to confirm that there was effective communication during performance of the map task.

All items were digitized (44 kHz, 16 bit) from audiotape using a Denon stereo tape deck connected to a Power Macintosh 6100/60AV computer running SOUNDEDIT 16 (by Macromedia, Inc.). Minimal digital editing was required to remove all items from their contexts, either in running conversational speech or in sentences from the pre- and post-task sessions, and to remove infrequent artifacts produced by talker movements. Such noises were brief and were usually caused by the talker bumping into the desk or microphone. They were removed by excising the noise at zero crossings so that the editing was unnoticeable in the final tokens. The listening tests were conducted in quiet testing rooms and presented over Sennheiser HD280 Pro circumaural headphones connected to Macintosh G3 computers running PSYSCOPE (Cohen *et al.* 1993).

As shown in Fig. 1, the AXB similarity test consisted of a series of trials in which a listener heard three repetitions of the same landmark label phrase, and the phrase varied across trials. On each trial, a task sample item produced by either a giver or receiver (X) was flanked by two items at 200-ms ISI from the corresponding partner (A and B). In order to assess convergence within the conversation, the task repetitions were compared to an item from the pre- or post-task session. The post-task comparison condition was included to assess the persistence of convergence. In addition, another measure of persistence was taken by comparing pre- and post-task items directly to task sample items. If phonetic convergence does occur and extends beyond the conversation, post-task items ought to sound more similar to the task sample items than pretask items.

The pre- and post-task direct comparison condition also alleviates a potential confound in the other two comparison types. When a listener chooses the task repetition as more similar to the task sample item than the pre- or post-task items, the decision could be based on putative differences between spontaneous and prompted speech. However, most studies demonstrating accurate perceptual classification of read and spontaneous speech have used stretches of discourse longer than the short phrases used here. Blaauw (1994) cites an unpublished finding that classification is above chance for items that are 4 to 6 syllables long, but also notes that performance is much lower than for longer stretches of speech. Even for longer stretches of speech, Blaauw found that classification of "the spontaneous-read distinction is not strictly categorical, but seems to be of a more continuous character" (p. 366). Therefore, read and spontaneous speech samples vary in the degree to which they support perceptual classification as such, and the ability to make the distinction is susceptible to the duration of the sample used to elicit the judgment. At this point, a listener's ability to judge whether a short phrase was read or spontaneously produced is uncertain, and the current protocol attempted to minimize the kinds of prosodic markers that have been found to support perception of the difference by encouraging fluent production of the items in the prompted sessions, and by using items sampled from the second iteration of the map task phrase list in the pretask session. Furthermore, the AXB comparison condition that compared pretask and post-task items directly is not susceptible to a spontaneous-read distinction because both items were prompted from lists.

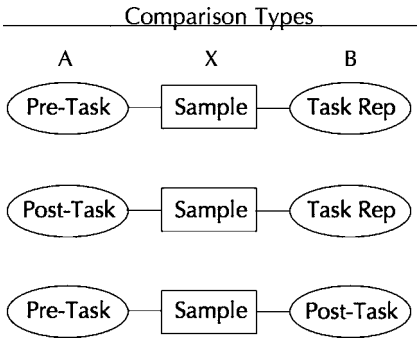


FIG. 1. AXB trial structure. On each trial, a listener heard the same landmark label phrase three times. The first and last items were to be compared to the middle item for similarity in pronunciation. The sample item was a giver's or receiver's task session production, and the task repetition was the corresponding partner's repetition of the same phrase. One-third of the time, the phrase was compared to the partner's production of the item from the pretask session, or to an item from the post-task session, or a pretask and a post-task item were compared. Half the trials used items that were repeated from giver to receiver, and the other half, from receiver to giver.

For each triplet, a listener decided as quickly as possible which item, the first or the last, sounded more similar to the middle item in its pronunciation. Listener responses were collected via the number 1 (first item) and 0 (last item) keys on the keyboard. Each successive trial began 1000 ms after a listener indicated a response. The order of presentation of flanking items was counterbalanced, each trial type was presented three times in mixed random order, and the effects of *timing, persistence/comparison type, talker role*, and *pair sex* were all tested within subject (blocked by *talker role*, to keep

the speaker of the task sample item the same throughout a block).

C. Participants

The talkers were six men and six women from the undergraduate population of Yale University who were paid for their participation. All participants reported that they were native speakers of American English with no speech or hearing disorders.

A total of 30 listeners participated in the AXB similarity test. All were native American English speakers from the student population of Columbia University who reported normal hearing and received Introductory Psychology course credit or were paid for participation.

III. RESULTS

A. Map task performance

The accuracy of path duplication in the map task (percentage of path grid cells on a giver's map that were duplicated on a receiver's map) was 85% overall, and there were no significant differences in performance across the dataset. The average time spent on each map was 8.84 min, and the average total amount of time spent on all five maps was 44.22 min (47.30 min for female pairs and 41.13 min for male pairs, with overlapping ranges). Overall, performance on the map task was good with a moderate amount of conversation, indicating that talkers were communicating effectively throughout the map task session.

The conversations also yielded enough repetitions across partners to compose the AXB test of phonetic convergence. From the total set of 24 pairs of items used in the AXB test, nine of the repetitions occurred less than 4 s after the sample, ten of the repetitions occurred between 4 and 11 s after the sample, and the remaining five repetitions occurred between 19 and 83 s after the sample. The length of time between repetitions was not confounded with any experimental factor. Although Goldinger (1998) found that shadowed tokens produced after a 3-s delay did not sound like imitations of their samples, there were not enough items in this corpus to satisfy both the current constraints and such a short interitem repetition interval. Given that a finding of convergence despite longer interitem intervals would constitute a more conservative test, the timing constraint here was more relaxed than Goldinger's findings would prescribe.

B. Perceptual assessment of convergence

Responses in the AXB test were scored as the percentage of trials on which a task repetition was chosen as more similar to the task sample item than a pre- or post-task item, or a post-task item was chosen as more similar to a task sample than a pretask item. The data were submitted to a repeated measures ANOVA to test for the effects of *timing* (early vs late), *persistence/comparison type* (pretask versus task, post-task versus task, and pretask vs post-task comparison conditions), *talker role* (giver X repeated by receiver vs receiver X repeated by giver), and *pair sex* (females vs males).

TABLE I. Interaction between comparison type and role.

	R to GX	G to RX
Pretask vs task[a]	62	68
Post-task vs task	56	57
Pretask vs post-task[a]	58	65

[a]95% confidence intervals verified that means comparisons differed across these two rows and all measures were different from chance.

In the AXB similarity test, listeners detected increased similarity in pronunciation between talkers during conversational interaction (percent task repetition chosen vs pretask items, 65%, vs post-task items, 57%, and pretask vs post-task, 62%. Unless otherwise noted, for all percentages reported in the text or in the table, 95% confidence intervals confirmed that performance was significantly above chance, which was 50%). This similarity was detected for items produced early in the conversation (59%) and was greater for items produced later in the conversation [63%; main effect of *timing*, $F(1,29)=9.88$, $p<0.004$]. In addition, the similarity persisted beyond the conversation, as indicated by the finding that post-task items were judged more similar to task sample items than the pretask items (62%). Persistence of convergence was also reflected in the difference between the pretask vs task and post-task vs task comparison conditions [65% vs 57%; main effect of *persistence/comparison type*, $F(2,58)=27.24$, $p<0.001$]. Presumably the persistence of phonetic convergence in the post-task session was strong enough to allow a listener to resolve convergence in the task session items.

In contrast, the effects of *talker role* and *pair sex* went in the opposite directions than predicted: task repetitions produced by receivers were less similar to givers' task sample items than givers' task repetitions were to receivers' task sample items (GX 59% < RX 63%); and female pairs' task repetitions were less similar to task sample items than male pairs' task repetitions [females 55% < males 68%; main effect of *talker role*, $F(1,29)=12.92$, $p<0.001$, main effect of *pair sex*, $F(1,29)=63.55$, $p<0.001$].

Table I shows the interaction between *persistence/comparison type* and *talker role*. The effect of *talker role* was significant for those trials in which the task repetitions were compared to the pretask productions, but not for trials in which the task repetitions were compared to the post-task productions [interaction between *persistence/comparison type* and *talker role*, $F(2,58)=5.48$, $p<0.007$; 95% confidence intervals were used to establish the differences for the means comparisons in the first and third rows of the table]. Therefore, givers converged to receivers more than receivers converged to givers, but only when considered against the items produced before the interaction. After the interaction, the residual convergence during the post-task session neutralized the talker role asymmetry displayed during the task session, arguably because the post-task items reflected a similar degree of convergence that was evoked during the conversational setting. The data in the bottom row show that the role-governed asymmetry in similarity was detected in the trials comparing a talker's pretask and post-task items to their partner's task items. In this condition, both comparison

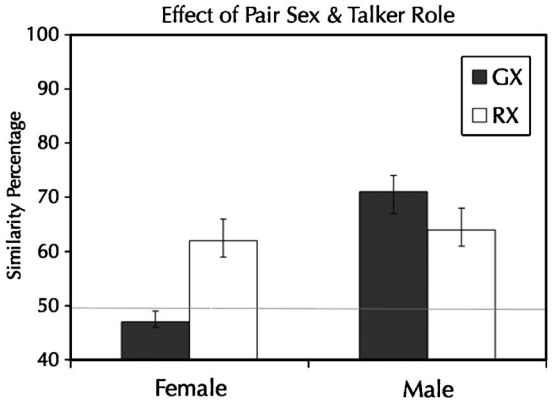


FIG. 2. AXB test interaction between *pair sex* and *talker role* with 95% confidence intervals. The dark GX bars correspond to the convergence of receivers' items to givers' sample items, and vice versa for the RX bars.

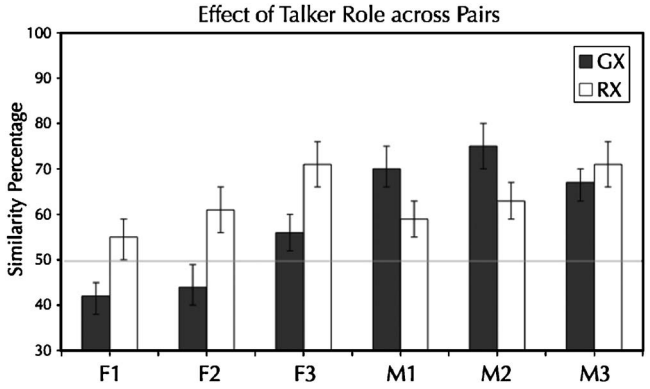


FIG. 3. AXB test interaction between *pair* and *talker role* with 95% confidence intervals. The dark GX bars correspond to the convergence of receivers' items to givers' sample items, and vice versa for the RX bars.

items from the talker were prompted by a list, yet the post-task items were produced immediately after the conversational setting, and exhibited persistent phonetic convergence in the same pattern of role-governed asymmetry that was found to be present in the task repetitions. An appeal to a distinction between spontaneous and read speech does not apply to this condition, yet parallel results obtained. This assertion was confirmed by a separate analysis of variance on the pretask vs post-task comparison trials' data: The main effects of *talker role* and *pair sex* were significant, as well as the interaction between *talker role* and *pair sex*, all showing similar patterns to the full dataset [*talker Role*: GX 58% $<$RX 65%, $F(1,29)=17.21$, $p<0.001$; *pair sex*: females 57% $<$males 65%, $F(1,29)=18.93$, $p<0.001$; *interaction*: females GX 50% and RX 65%, males GX 66% and RX 65% $F(1,29)=18.20$, $p<0.001$].

Figure 2 shows the interaction between *pair sex* and *talker role*, collapsing across all comparison conditions. Female talkers exhibited the overall pattern, greater similarity of givers to receivers, while male talkers exhibited the opposite pattern, greater similarity of receivers to givers. Comparing across the whole data set, the similarity of givers to receivers was comparable, but the receivers were not similar to givers for the female pairs, while the receivers were more similar to the givers for the male pairs [$F(1,29)=118.48$, $p<0.001$; error bars depict 95% confidence intervals]. It appears that the male talkers followed giver-dominated convergence, and the female talkers exhibited receiver-dominated convergence. Figure 3 shows the effect of *talker role* across individual pairs of talkers. The group behavior appears to be more consistent across the female than the male pairs of talkers [$F(2,58)=5.80$, $p<0.005$; error bars depict 95% confidence intervals].

Overall, the AXB similarity data indicate that phonetic convergence occurred during the map task conversation, carried through to speech produced immediately after the conversation, and was greater when a receiver provided the sample utterance that a giver repeated.

IV. DISCUSSION

This study found robust phonetic convergence between conversational participants. Despite the fact that partners were unacquainted prior to interaction, there was evidence for early convergence that increased over the course of conversation and persisted beyond the conclusion of the interaction. In the AXB listening tests, there were effects of the sex of the pair and the role of the talker in the conversation. Overall, male talkers converged more than females, and givers converged more than receivers. In female pairs, givers exhibited convergence to receivers, but receivers did not converge to givers. In male pairs, the opposite pattern was found—male receivers converged to male givers more than the reverse. Therefore, this study establishes the existence of a relatively rapid process of phonetic convergence between interacting talkers influenced by a talker's role and sex, and persisting beyond the conversation that induces it.

The effects of talker sex and role did not follow the predictions suggested by previous research on accommodation, in which female talkers converged more than male talkers and less dominant talkers converged more than more dominant talkers. In a study of accommodation that used a lexical shadowing task, Namy *et al.* (2002) found that female shadowers converged to their models more than male shadowers, and that female listeners detected convergence more readily than male listeners. Citing earlier research finding that women were more accurate in identifying talkers (Nygaard and Queen, 2000), Namy *et al.* suggested that female listeners detected convergence more readily due to greater perceptual sensitivity or attention to indexical features of talkers. If female listeners are generally more sensitive to indexical features of talkers, then the female talkers in the current study should have resolved their partner's phonetic forms in more detail than the males, leading to greater convergence in response to their partner's speech. Because the current study found greater convergence on the part of male talkers, it is more likely that attention rather than absolute perceptual sensitivity is moderating these effects. Men and women might sustain habitual attentional sets that vary across different circumstances, leading to differences in the grain of perceptual analysis and subsequent shadowing or conversational convergence.

Previous research on the influence of status or dominance on accommodation phenomena found that the implied dominance relationship among members of the pairs did not

always determine the direction of convergence (Bilous and Krauss, 1988; Bourhis and Giles, 1977; Giles, 1973; Gregory and Webster, 1996). In the current study, female receivers did not converge to female givers, but male receivers converged more than male givers, indicating that the interpretation of dominance is not a straightforward function of nominal role. The interaction with talker sex hints that men and women may interpret role labels differently, leading to distinct patterns of convergence. If such interpretations are not mandated by a talker's sex *per se*, it should be possible to influence the direction of the effect in different settings or instructional conditions. These findings do not support an interpretation based solely on differences in perceptual sensitivity; rather, functions outside the domain of perception appear to be influencing the degree of phonetic convergence.

Although perceptual sensitivity to indexical or phonetic features is necessary for accommodation, it is not sufficient to evoke phonetic convergence. Attention may adjust the grain of perceptual resolution, but additional processes influence a talker's phonetic form. The current study cannot provide a clear explanation for the direction of the observed effects of talker sex and role, but the reliability of these effects challenges an account of convergence based solely on a direct link between speech perception and speech production. If automatic priming is the mechanism by which phonetic convergence operates, then the sex-and role-governed patterns found here must be due to processes within speech perception and/or production that are susceptible to extralinguistic factors. There is currently no account of priming that incorporates the sex or role of a talker as a modulator of the degree of priming. Priming is mainly influenced by factors internal to lexical representation, such as semantic relatedness, word frequency, and lexical neighborhood density (see Luce and Pisoni, 1998). In the current experiment, the differences in phonetic convergence do not hinge on such factors because the same lexical items were counterbalanced across all talker role and sex conditions.

If speech perception and production are not labile to superordinate influences—if they are modular (Fodor, 1983; Liberman and Mattingly, 1985)—then these findings suggest three alternative possibilities: (1) a loosening of the tie between perception and production; (2) a process intervening prior to speech perception; or (3) a process intervening between speech production planning and execution. It could be the case that a talker's role or sex induces a more or less focused attentional set, which in turn specifies an appropriately detailed phonetic representation that drives production. A report by Goldinger and Azuma (2003) provides evidence that talkers and listeners can be biased to produce and to respond to relatively finer- or coarser-grained aspects of speech signals, depending on a simple instructional manipulation. Therefore, processes outside speech perception and production enter the system at some point, but a detailed elaboration of the mechanisms and acoustic-phonetic attributes awaits further investigation. The patterned variability in phonetic convergence demonstrates that convergence is not a result of automatic priming—the social setting of language use modulates the degree to which a talker's phonetic repertoire converges on that of a conversational partner. How does a talker incorporate aspects of an interacting partner's phonetic repertoire?

A. Episodic memory

Goldinger (1998) provided compelling evidence that lexical imitation is a natural consequence of an episodic memory system. Indeed, some of the factors that influenced the degree of lexical imitation found in his shadowing experiments coincided with predictions derived from an episodic memory model, namely Hintzman's (1986) MINERVA2. This model begins with an assumption that every perceptual episode leaves a trace in long-term memory, and warrants that every time a new episode is encountered, all traces that are similar to the original episode are activated and compared directly to the sample token. Based on the outcome of comparison, the episodic system generates a composite representation, an echo, of the activated traces and the sample token. Goldinger proposed that this composite serves as the model for the shadowed productions. Because the multitude of traces integrated into an echo has different effects depending on the number of exemplars that are available, word-frequency effects on imitation were predicted and found both in perceptual judgments and MINERVA2-based modeling. Common words have more traces in memory that attenuate the prominence of distinctive aspects of a new token in the echo. Echoes from rarer words, in contrast, will integrate a recent occurrence with fewer competing traces, leading to greater imitative fidelity.

Crucially, for most items repeated after a 3-s delay in Goldinger's study, imitation was not detected. The absence of delayed imitation was attributed to dilution of distinctive aspects of a sample token while holding the item in working memory: Analogous to effects found for common words, as an item is rehearsed in working memory, its distinctive attributes are lost through reiterative comparison with traces in long-term memory. Therefore, the relative influence of long-term memory on echo-based imitation increases with both greater numbers of similar traces and greater amounts of time between a sample and its repetition, leading to a decrease in imitative fidelity.

The current study found evidence for a form of delayed imitation: Phonetic convergence in conversational settings persisted into a post-task session. An episodic memory system containing detailed lexical episodes could not form the basis for this effect because the delay would increase the influence of long-term memory traces on repeated words, leading to reduced phonetic convergence. Moreover, the observed increase in phonetic convergence over the course of conversational interaction occurred across different lexical items, providing evidence that the change in phonetic repertoire was not tied to specific lexical forms. At a much broader time scale, Sancier and Fowler (1997) found that a bilingual talker shifted pronunciation of consonants in both languages as a result of recent language experience—even consonant VOTs in the unused language were affected. This finding is unlikely to be based on lexical episodes, as the episodes in question did not match the vocabulary of the

affected language. Finally, any account relying on episodic memory must be mute with respect to the effects of talker sex and role—both talkers in a given pair heard the same lexical items over the course of the conversational setting, yet one talker produced more convergent forms. Like an automatic priming mechanism, an episodic memory system fails to explain the effect of a transient social factor on phonetic convergence. From this evidence, it appears that the functional circumstances of language use induce a kind of phonetic convergence that is not found in individual echoes of mere exposures. A more appropriate conceptualization of these findings might be found in the literature on entrainment, which incorporates relative dominance in dynamical systems without appeal to automatic priming or episodic memory traces.

B. Entrainment

The principles of entrainment were initially identified in von Holst's (1937/1973) early research on endogeneous rhythmicity, in which the complex motions of oscillating fish fins were readily described in terms of the superimposition of sinusoidal functions. Later, von Holst's principles were found to scale up to the dynamics of more glamorous organisms, like pairs of humans swinging legs or wrist pendulums (see Turvey, 1990). Accordingly, the *magnet effect* is the tendency for a more dominant or stable oscillator to pull a less dominant oscillator into synchrony. Absolute coordination or entrainment is a rare phenomenon in which both the phase relationship and the frequency of oscillation match, and only occurs with rigid coupling of systems that have identical intrinsic dynamics (Schmidt and Turvey, 1989). Lacking rigid coupling, interpersonal entrainment typically exhibits only relative coordination. Relative coordination demonstrates another principle of entrainment, the *maintenance tendency*. Despite the pull to entrain to a coupled oscillator, the manifest pattern exhibits a latent influence of the original intrinsic dynamics, presumably because the external oscillator's pattern is superimposed onto the internal oscillator's pattern rather than supplanting it.

These properties of entrainment in coordinated dynamical systems provide a ready model of the integration of internal and external forces in human behavior. Beek, Turvey, and Schmidt (1992) framed the relation between internal organization of coordinated activities and information from external sources in terms of dynamical systems theory, proposing that external information is an embedded forcing function on internal dynamics. For example, externally derived information might influence speech production by first indirectly participating in a separate perceptual-memory system or by directly specifying phonetic forms. Beek *et al.* propose that external information acts as an embedded forcing function on internal dynamics, inducing changes in the overall pattern of activity that push the activity to different values in its intrinsic range. Research on self-regulation of speech production, in particular the Lombard sign (Lane and Tranel, 1971) and perceptual-productive adaptation of vowel formants, speaking fundamental frequency, and consonant spectra (Houde and Jordan, 2002; Jones and Munhall, 2000, 2003) shows that talkers can incorporate auditory feedback of their own productions to adjust subtle aspects of speech at short latencies. If perception of another talker's speech yields detailed phonetic forms, such forms could influence subsequent production under circumstances, such as the demands of conversational interaction, that promote coupling between talkers.

For the kinds of phenomena examined in the current study, the direction and form of phonetic convergence is difficult to predict. Viewed one way, the giver is dominant by providing the information to be copied, and the task requires the receiver to comply; on the other hand, the receiver must ensure that the giver provides adequate instructions, therefore, the receiver might set the tone for interaction. Either organization is feasible, and potentially idiosyncratic to different pairs. In future investigations, it will be useful to manipulate relative dominance by explicitly instructing one talker to imitate the other talker. Perhaps the instruction to imitate will override the nominal effects of talker role in setting up the dominance hierarchy. More significant, however, dominance is irrelevant for entrainment if the systems are not coupled. With looser coupling, there is likely to be less convergence, as is generally the case with informationally coupled systems, such as interacting talkers (see Schmidt and Turvey, 1989). With respect to the current findings, an account that acknowledges varied degrees of coupling and dominance in between-talker interaction is a better fit to the findings than an autonomous account of automatic priming from speech perception to speech production.

As an instance of relative entrainment, phonetic convergence may be analogous to other forms of alignment phenomena between talkers. Pickering and Garrod (2004) claim that discrepancies from alignment are attributed to indirect secondary processes that monitor comprehension and adjust output when comprehension fails. However, sometimes a talker will diverge from an interlocutor without a failure of comprehension (Bilous and Krauss, 1988; Bourhis and Giles, 1977). Although speech perception resolves the detailed aspects of phonetic form that would be necessary for phonetic convergence with any degree of fidelity, a talker is not automatically driven to imitate those forms. What sorts of nonlinguistic functions might phonetic convergence or divergence serve for a talker?

C. Convergence and social function

Currently, accounts of language use in social interaction emphasize the social situation in which speech occurs, as opposed to individual factors in speech production. Thus, language is not produced by isolated individuals in order "to generate grammatical strings," (Krauss, 1987, p. 97); rather, in addition to the production of linguistic forms, speech projects social categories (Giles, Scherer, and Taylor, 1979; Shepard, Giles, and Le Poire, 2001), is used to accomplish mutual goals (Clark, 1996; Clark and Wilkes-Gibbs, 1986), and/or to align representations (Garrod and Doherty, 1994; Pickering and Garrod, 2004). Communication is more than a matching process; a talker expresses more than a sequence of phonemes, and a listener uses the speech signal to under-

stand the talker as well as the message. If some part of phonetic variability is communicative, i.e., it serves some purpose for the talker, then a listener who resolves the phonetic detail can project this into a perception of the talker's connotation, apart from lexical access. Among interacting talkers, phonetic convergence might contribute to mutual comprehension and/or rapport through a decrease in social distance (Shepard, Giles, and LePoire, 2001).

The current study establishes the existence of a process of rapid phonetic convergence that emerges in conversational settings, providing a link between the laboratory studies of nonsocial shadowing imitation and community-level linguistic change. However, the constraints on convergence in conversational settings appear to differ somewhat from those of shadowing studies. Labov (1974, 1986) suggested that linguistic change is motivated by the need to add emphasis to expression, and that new forms are adopted as a result of interactive conversation. At the same time that talkers share a common phonetic ground, each talker maintains some distinction through novelty, perhaps as Labov (1974) suggests, "to signal a stronger meaning than the older form; to display the speaker's membership in a local group; and to demonstrate greater intimacy than an older form." (p. 253) The current study also provides some evidence that interactive changes persist beyond a particular social situation, perhaps to be carried to the next interaction.

Garrod and Doherty (1994) provided insight into the development of conventions in paired and community-level conversational interactions. They proposed that greater overall stability might arise across a community as opposed to individual pairs because of the greater initial variability of viewpoints afforded by communities. This variability induces competition among different concepts, leading to a greater likelihood that a more stable form will survive across a community of talkers. In the context of the current findings, it may be possible to trace phonetic convergence among a community of talkers to determine more precisely the kinds of changes that are durable. If the observed effects of talker role serve a pragmatic purpose, they ought to endure community interaction.

Although these findings extend convergence in conversational interaction to the phonetic domain, interacting talkers do not match on all acoustic-phonetic dimensions. If perception yields goals that drive production, then an adequate account of this relation requires an explanation of the lack of perfect correspondence (Pardo and Remez, in press). Because some of the disparities between talkers in this study patterned according to talker role and sex, it is likely that the discrepancy is due to more than noisy perceptual resolution or productive output of gestural goals (see also Vallabha and Tuller, 2004). Individual talkers in social settings have communicative goals that go beyond mere intelligibility (but see Lindblom, 1990), and these goals must be an integral part of the perceptual and productive system that creates the spoken message (see also Bradlow, 2002). Perhaps such goals act as weights on embedded forcing functions that incorporate perceived phonetic structures into produced phonetic forms.

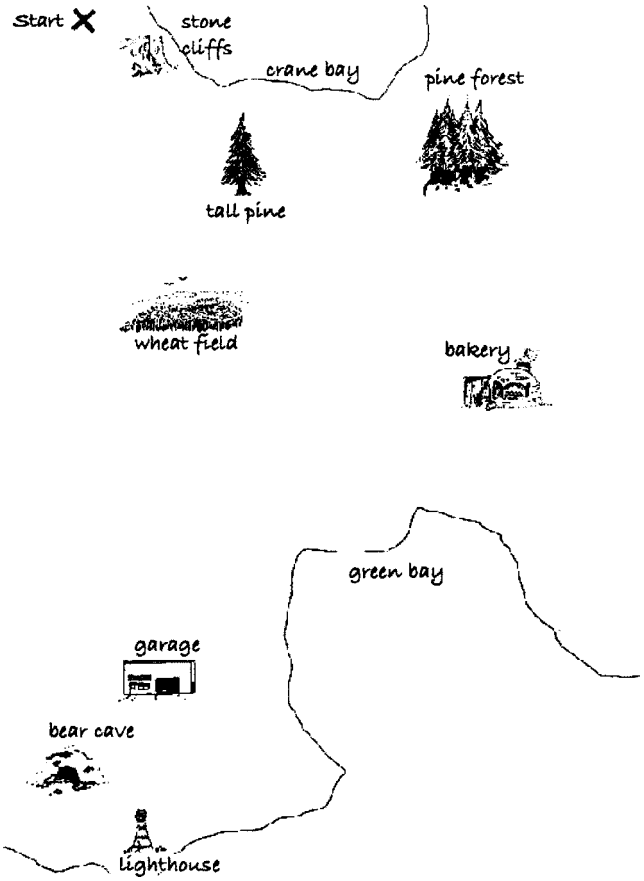
This study attempts to broaden an understanding of speech perception and production to include social function in language use. The link between perception and production in spoken communication is not automatic; it is subject to situational constraints that influence the direction and magnitude of phonetic convergence in conversational interaction. Future investigations might question whether these effects are durable enough to extend across community interactions, perhaps because of a broader cooperative function in social discourse. Many theorists propose some form of cooperative principle in social interaction (e.g., Clark 1996), yet few have examined the operation of the principle at the level of phonetic variability, and fewer still have attempted a rigorous exposition of the likely structural factors that evoke or attenuate the cooperative principle (e.g., Giles, Coupland, and Coupland, 1991). Although phonetic convergence varies across talkers in social interaction, an individual is not ruled completely by circumstance, but each implements convergence as the situation warrants.

ACKNOWLEDGMENTS

This research was supported in part by an NRSA training grant from the National Institutes of Mental Health to Jennifer Pardo and Robert Krauss at Columbia University. Portions of this project were conducted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to J. S. Pardo at Yale University. The author thanks Robert Crowder, Carol Fowler, Wendel Garner, Raquel Gardner, Robert Krauss, Robert Remez, George Ton, Rebecca Treiman, and many anonymous reviewers for their help in conceptualizing and completing this project.

APPENDIX A: SAMPLE MAP TASK MAP FOR A GIVER

APPENDIX B: SAMPLE MAP TASK MAP FOR A RECEIVER.



Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (**1991**). "The HCRC Map Task corpus," Lang Speech **34**, 351–366.

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.

Beek, P. J., Turvey, M. T., and Schmidt, R. C. (**1992**). "Autonomous and nonautonomous dynamics of coordinated rhythmic movements," Ecological Psychol. 4, 65–95.

Bilous, F. R., and Krauss, R. M. (**1988**). "Dominance and accommodation in the conversational behaviors of same- or mixed-gender dyads," Lang. and Commun. **8**, 183–194.

Blaauw, E. (**1994**). "The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech," Speech Commun. **14**, 359–375.

Bourhis, R. Y., and Giles, H. (**1977**). "The language of intergroup distinctiveness," in *Language, Ethnicity and Intergroup Relations*, edited by H. Giles (Academic, London), pp. 119–135.

Bradlow, A. R. (**2002**). "Confluent talker- and listener-oriented forces in clear speech production," in *Laboratory Phonology 7*, edited by C. Gussenhoven and N. Warner (Mouton de Gruyter, New York), pp. 241–273.

Branigan, H. P., Pickering, M. J., and Cleland, A. A. (**2000**). "Syntactic co-ordination in dialogue," Cognition **75**, B13–25.

Browman, C. P., and Goldstein, L. (**1991**). "Gestural structures: Distinctiveness, phonological processes, and historical change," in *Modularity and the Motor Theory of Speech Perception*, edited by I. G. Mattingly and M. Studdert-Kennedy (Erlbaum, Hillsdale, NJ), pp. 313–338.

Catchpole, C., and Pardo, J. S. (**2004**). "Articulatory shortening in repeated noun phrases is affected by participant role," Architectural Mechanisms for Language Processing, Universite de Provence, 16–18, September, 2004.

Chartrand, T. L., and Bargh, J. A. (**1999**). "The chameleon effect: The perception-behavior link and social interaction," J. Pers Soc. Psychol. **76**, 893–910.

Clark, H. H. (**1996**). *Using Language* (Cambridge University Press, Cambridge).

Clark, H. H., and Wilkes-Gibbs, D. (**1986**). "Referring as a collaborative process," Cognition **22**, 1–39.

Cohen, J., MacWhinney, B., Flatt, M., and Provost, J. (**1993**). "PSYSCOPE: An interactive graphical system for designing and controlling experiments in the psychology laboratory using Macintosh computers," Behav. Res. Methods Instrum. Comput. **25**, 257–271.

Fodor, J. A. (**1983**). *The Modularity of Mind* (MIT Press).

Fowler, C. A. (**1988**). "Differential shortening of repeated content words produced in various communicative contexts," Lang Speech **28**, 47–56.

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.

Fowler, C. A., Levy, E., and Brown, J. (**1997**). "Reductions of spoken words in certain discourse contexts," J. Mem. Lang. **37**, 24–40.

Fowler, C. A., Brown, J., Sabadini, L., and Weihing, J. (**2003**). "Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks," J. Mem. Lang. **49**, 396–413.

Garrod, S., and Doherty, G. (**1994**). "Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions," Cognition **53**, 181–215.

Giles, H. (1973). "Accent mobility: A model and some data," Anthropological Linguistics **15**, 87–109.

Giles, H., Coupland, J., and Coupland, N. (**1991**). *Contexts of Accommodation: Developments in Applied Sociolinguistics* (Cambridge University Press, Cambridge).

Giles, H., Scherer, K. R., and Taylor, D. M. (**1979**). "Speech markers in social interaction," in *Social Markers in Speech*, edited by K. R. Scherer and H. Giles (Cambridge University Press, Cambridge, England), pp. 343–375.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Goldinger, S. D., and Azuma, T. (**2003**). "Puzzle-solving science: The quixotic quests for units in speech perception," J. Phonetics **31**, 305–320.

Gregory, D., and Webster, S. (**1996**). "A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status predictions," J. Pers Soc. Psychol. **70**, 1231–1240.

Gregory, S. W. (1990). "Analysis of fundamental frequency reveals covariation in interview partners' speech," J. Nonverbal Beh. **14**, 237–251.

Hintzman, D. L. (**1986**). "'Schema abstraction' in a multiple trace memory model," Psychol. Rev. **93**, 411–428.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech. I. Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295–310.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of *F*0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246–1251.

Jones, J. A., and Munhall, K. G. (**2003**). "Learning to produce speech with an altered vocal tract: The role of auditory feedback," J. Acoust. Soc. Am. **113**, 532–543.

Krauss, R. M. (1987). "The role of the listener: Addressee influences on message formulation," J. Lang. Soc. Psychol. **6**, 81–98.

Krauss, R. M., and Pardo, J. S. (**2004**). "Is alignment always the result of priming?," Behav. Brain Sci. **27**, 203–204.

Krauss, R. M., and Weinheimer, S. (**1964**). "Changes in the length of reference phrases as a function of social interaction: A preliminary study," Psychonomic Sci. **1**, 113–114.

Labov, W. (**1974**). "Linguistic change as a form of communication," in *Human Communication: Theoretical Explorations*, edited by A. Silverstein (Erlbaum, Hillsdale, NJ), pp. 221–256.

Labov, W. (**1986**). "Sources of inherent variation in the speech process," in *Invariance and Variability in the Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 402–425.

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**, 677–709.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

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 Academic, Dordrecht, Netherlands), pp. 403–439.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The

neighborhood activation model," Ear Hear. **19**, 1–36.
Namy, L. L., Nygaard, L. C., and Sauerteig, D. (**2002**). "Gender differences in vocal accommodation: The role of perception," J. Lang. Soc. Psychol. **21**, 422–432.
Natale, M. (**1975**). "Convergence of mean vocal intensity in dyadic communication as a function of social desirability," J. Pers Soc. Psychol. **32**, 790–804.
Nygaard L. C., and Queen J. S. (**2000**). The role of sentential prosody in learning voices. Paper presented at the meeting of the Acoustical Society of America, Atlanta, GA.
Pardo, J. S., and Remez, R. E., "The perception of speech," in *The Handbook of Psycholinguistics*, 2nd ed., edited by M. Traxler and M. Gernsbacher (Elsevier, Cambridge, MA, in press).
Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (**2002**). "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," J. Acoust. Soc. Am. **112**, 1627–1641.
Pickering, M. J., and Garrod, S. (**2004**). "Toward a mechanistic psychology of dialogue," Behav. Brain Sci. **27**, 169–190.
Porter, R., and Castellanos, F. (**1980**). "Speech production measures of speech perception: Rapid shadowing of VCV syllables," J. Acoust. Soc. Am. **67**, 1349–1356.
Sancier, M. L., and Fowler, C. A. (**1997**). "Gestural drift in a bilingual speaker of Brazilian Portuguese and English," J. Phonetics **25**, 421–436.
Schmidt, R. C., and Turvey, M. T. (**1989**). "Absolute coordination: An ecological perspective," in *Perspectives on the Coordination of Movement*, edited by S. A. Wallace (Elsevier Science, New York), pp. 123–156.
Schober, M. F., and Clark, H. H. (**1989**). "Understanding by addressees and overhearers," Cogn. Psychol. **21**, 211–232.
Shepard, C. A., Giles, H., and Le Poire, B. A. (**2001**). "Communication accommodation theory," in *The New Handbook of Language and Social Psychology*, edited by W. P. Robinson and H. Giles (Wiley, New York), pp. 33–56.
Turvey M. T. (**1990**). Coordination. American Psychologist **45**, 938–953
Vallabha, G. K., and Tuller, B. (**2004**). "Perceptuomotor bias in the imitation of steady-state vowels," J. Acoust. Soc. Am. **116**, 1184–1197.
Viechnicki, P. D. (**2002**). "Composition and granularity of vowel production targets," Diss. Abstr. Int., C **63**(4-A), 1320, UMI.
von Holst, E. (**1937/1973**). "On the nature of order in the central nervous system," in *The Behavioral Physiology of Animal and Man: The Collected Papers of Erich von Holst* (University of Miami Press, Miami, FL).

Perceptual invariance of coarticulated vowels over variations in speaking rate

Janet W. Stack
Communication Disorders Program, Curry School of Education, University of Virginia, 2205 Fontaine Avenue, Suite 202, Charlottesville, Virginia 22903

Winifred Strange
Ph.D. Program, Speech and Hearing Sciences, The Graduate Center, CUNY, 365 Fifth Avenue, New York, New York 10016

James J. Jenkins, William D. Clarke III, and Sonja A. Trent
Department of Psychology, University of South Florida, 4200 Fowler Avenue, Tampa, Florida 33620

(Received 9 September 2004; revised 8 January 2006; accepted 11 January 2006)

This study examined the perception and acoustics of a large corpus of vowels spoken in consonant-vowel-consonant syllables produced in citation-form (lists) and spoken in sentences at normal and rapid rates by a female adult. Listeners correctly categorized the speaking rate of sentence materials as normal or rapid (2% errors) but did not accurately classify the speaking rate of the syllables when they were excised from the sentences (25% errors). In contrast, listeners accurately identified the vowels produced in sentences spoken at both rates when presented the sentences and when presented the excised syllables blocked by speaking rate or randomized. Acoustical analysis showed that formant frequencies at syllable midpoint for vowels in sentence materials showed "target undershoot" relative to citation-form values, but little change over speech rate. Syllable durations varied systematically with vowel identity, speaking rate, and voicing of final consonant. Vowel-inherent-spectral-change was invariant in direction of change over rate and context for most vowels. The temporal location of maximum F1 frequency further differentiated spectrally adjacent lax and tense vowels. It was concluded that listeners were able to utilize these rate- and context-independent dynamic spectrotemporal parameters to identify coarticulated vowels, even when sentential information about speaking rate was not available. [DOI: 10.1121/1.2171837]

PACS number(s): 43.71.An, 43.71.Es [PFA] Pages: 2394–2405

I. INTRODUCTION

Many studies of speech perception have found significant effects of the rate of speech on the identification of both natural and synthetic speech tokens (see the extensive and thoughtful review by Miller, 1981, as well as a variety of more recent articles, e.g., Cummins, 1999, Gopal, 1990, Moon and Lindblom, 1994, and Smith, 2001). Remarkably, only a small portion of this research has focused on the perception of vowels. Further, when vowels have been studied, the investigation has frequently been limited to a small number of specific vowel contrasts. For example, Piterman (2000) is devoted to a single contrast, /ɛ-a/, in a single intravowel context in French, and Gottfried *et al.* (1990) is limited to three contrasts of synthesized English syllables, /i-ɪ/, /ɛ-æ/, and /ɛ-i/. As Gottfried and his colleagues point out, one result of this selectivity is that one may find large perceptual effects of speaking rate for some particular contrast where duration is particularly important, while no such differences are detected for other contrasts, which are more dependent on distinctive spectral parameters. In addition, it is important to recognize that the task of identifying a stimulus, natural or synthetic, as one of two possible vowels in a closed-set comparison is radically different from identifying a vowel when all the vowels in the language are possible candidates.

An additional problem in evaluating previous research on the effects of speaking rate on vowel perception is that most previous studies, including some of our own, have been limited to only one or two consonant environments although it is well known that consonantal context has major effects on the acoustics of coarticulated vowels (Hillenbrand *et al.*, 2001; Stevens and House, 1963; Strange, 1989). A further complication is that most of the studies examining the effects of speaking rate on vowel perception have used a "mismatch" paradigm in which rate cues in speech surrounding the test syllables were manipulated to determine the extent to which such rate mismatches affected the perceived identity of the vowel. For example, Johnson and Strange (1982) found that such mismatches had little or no effect on the identification of vowels in /tVt/ syllables spoken at a normal rate spliced into rapid sentences (accurate identification was 95% or better). However, when syllables from rapidly spoken sentences were spliced into normal rate sentences, or presented in isolation as if they had been produced in citation form, errors on intrinsically long vowels increased significantly, while perception of intrinsically short vowels remained relatively accurate. In general, the errors were confusions of spectrally similar long-short vowel pairs indicating that the rapidly spoken long vowels were identi-

 0001-4966/2006/119(4)/2394/12/$22.50

fied as their short counterparts. Verbrugge and Shankweiler (1977) reported the same findings with both stressed and destressed vowels in /pVp/ syllables.

Other research with this experimental paradigm showed that the results could be altered by giving the listeners information about the generation of the stimuli. When Johnson and Strange told their listeners in a second experiment how the sentences were produced, the error rate for rapidly spoken long vowels embedded in normal rate sentences was greatly reduced. It is clearly one thing to show that listeners' identification of vowels can be altered by presenting stimuli in varied context, but it is quite another thing to conclude that the listener is obliged to use the context under all circumstances. On the positive side, the fact that vowels in syllables spoken at normal speaking rates were little affected by being embedded in rapid sentences, argues for the importance of information intrinsic to the syllable itself. When such information for vowel identity is well specified, it has sufficient weight to overcome contextual manipulation of perceived rate of speech.

The literature lacks studies that examine a broad sample of vowels in American English, produced at different speech rates in natural or quasi-natural sentential contexts, with the vowels in a variety of phonetic environments. In short, a study is needed to determine what the fully informed listener *can* do in identifying vowels spoken at different rates as a preliminary to examining what kinds of information the listener could be using.

The present research takes a step toward such a general study of the effect of speaking rate on vowel perception by focusing on the identification of a large corpus of American English vowels (excluding the true diphthongs and rhotic vowels) embedded in six different stop consonant environments and spoken in sentences at normal and fast speaking rates. The paper reports two perceptual experiments in which these sentences and consonant-vowel-consonant (CVC) syllables excised from them were presented to independent groups of listeners who were asked either to indicate the rate of speech of the (original) utterances, or to identify the vowels.

The research is reported in two main sections. The first section (Secs. III and IV) deals with perceptual questions; the second section (Sec. V) reports the acoustical analysis of the materials. In the first section, the following questions were addressed:

(1) Can listeners identify the rate at which the sentences were spoken?
(2) Can listeners identify the speaking rate when given only the CVC syllables excised from the sentences?
    These two questions were raised to determine the extent to which explicit cognitive information concerning rate could be a moderating factor in the identification of vowels. The judgment of rate was expected to be excellent when full sentences were presented. Whether the same was true of excised syllables was a question to be determined by the experiment.
(3) Can listeners identify the vowels when syllables are presented in original sentence context randomly mixed over rate?
    This condition evaluated the listeners' ability to identify vowels when all the contextual cues for rate were present in each stimulus, but rate of speech was variable from stimulus to stimulus. The expectation was that vowel identification would be excellent.
(4) Can listeners identify the vowels in syllables excised from the sentences when the stimuli are blocked by rate in presentation?
(5) Can listeners identify the vowels in the excised syllables when the stimuli are presented randomly mixed over rate?

The first of these two conditions removes any uncertainty about the rate of speech prior to presenting a syllable for vowel identification. The second condition embodies uncertainty as to speech rate but leaves syllable-intrinsic information for the vowel intact.

II. SPEECH MATERIALS

The speech materials were produced by an adult female who spoke a General American dialect. The speaker was an experienced speech-language pathologist who normally spoke at a relatively rapid rate. The corpus of speech materials was recorded in an IAC chamber on a reel-to-reel Tascam tape recorder using a Shure Omnidyne microphone and Ampex 631, 1.5 mil polyester tape. The stimuli consisted of 11 American English vowels /i, ɪ, e, ɛ, æ, ɑ, ɔ, ʌ, o, ʊ, u/ produced in different consonant contexts in citation form disyllable utterances and in sentences spoken at a self-selected normal rate and at a rapid speaking rate (i.e., as rapidly as possible without any distortions or deletions). Three sets of materials were recorded.

A. Variable-consonant sentence context (Context corpus)

The 11 vowels were produced in /həCVC/ disyllables embedded in the phrase "I hear the sound of /həCVC/ some more." The six contexts used were /bVd, dVd, gVd, bVt, dVt, and gVt/. There were two productions of each vowel in each context, and all were produced at both normal and rapid speaking rates for a total of 264 sentences. These utterances were used as stimuli in the perceptual experiments.

B. Neutral-consonant sentence context (Neutral corpus)

The 11 vowels were produced in the disyllables /əhVd/ and /əhVt/ embedded in the same carrier phrase "I hear the sound of /əhVC/ some more." These sentences were also produced at the speaker's normal and rapid speaking rates. There were two productions of each utterance at each rate for a total of 88 utterances (11 vowels $\times$ 2 contexts $\times$ 2 rates $\times$ 2 repetitions). These were produced for the purpose of comparing vowel targets, formant trajectories, and durations in syllables in which the vowel was coarticulated with an initial stop consonant (Context corpus) and those produced in the /hVC/ context which is often regarded as neutral (Hillenbrand, *et al.*, 2001; Stevens and House, 1963).

C. Citation-form /əhVC/ disyllables (Citation corpus)

The 11 vowels were produced in the same disyllables as in the Neutral corpus, /əhVd/ and /əhVt/. The speaker preceded each utterance with a number announcement; "number one," "number two," etc., followed by a pause. Thus, the disyllables were spoken with utterance-final intonation, with the target vowel stressed. For this condition, utterances were produced only at the self-selected normal rate. There were two productions of each token for a total of 44 tokens. These productions were taken as defining the traditional canonical vowel targets for this speaker.

The entire recorded corpus was digitized on a Digital Equipment Corporation VAX Station II/GPX at a sampling rate of 20 kHz, 12-bit resolution, after low-pass filtering at 9.8 kHz and 6 dB/octave pre-emphasis above 1 kHz. Audiotapes were then generated for the conditions of the perceptual experiments.

The purpose of Experiment 1 was to examine listeners' judgments of speaking rate and to determine whether those judgments differed in accuracy when listeners heard sentences versus when they heard the CVC syllables excised from the sentences. The purpose of Experiment 2 was to examine listeners' ability to identify vowels in normal- and rapid-rate sentences and in the CVC syllables excised from those sentences.

III. EXPERIMENT 1: RATE IDENTIFICATION

A. Method

1. Stimuli

The stimuli for the Sentence Condition consisted of the sentences of the Context corpus, with normal- and rapid-rate sentences randomly intermixed with a 4 s interstimulus interval (ISI) and an 8 s interval after every 11 trials. Two independent randomizations were created using the two tokens of each type. Each randomization consisted of 132 sentences (11 vowels $\times$ 6 contexts $\times$ 2 rates) for a total of 264 trials in all. The stimuli for the Syllable Condition consisted of the same two random sequences of the CVC syllables excised from the normal and rapid rate sentences of the Context corpus. The interblock interval remained 8 s, but the ISI was increased to 5 s.

2. Participants

Twenty phonetically untrained listeners participated, ten in each condition. Participants were volunteers from undergraduate psychology classes who were awarded extra credit points for participating. All were monolingual speakers of American English with the exception of one participant in the Sentence Condition who also spoke some Spanish. All participants reported their hearing as normal.

3. Procedures

Participants listened to stimuli presented over headphones in a sound insulated booth. They were tested in groups of one to three. For the Sentence Condition, the experimenter explained how the stimuli were produced. Listeners were then given response forms on which they were instructed to circle N (for normal-rate) or R (for rapid-rate) for each stimulus. They were familiarized with the test stimuli by hearing 22 sentences spoken at normal and rapid rate in mixed order without responding. This was followed by the test. There was a short break between the two sets of 132 sentences. The order of presentation of the randomizations was counterbalanced across listeners. For the Syllable Condition, participants were informed that one-half of the utterances they would hear had been excised from sentences spoken at a normal speaking rate and half from sentences spoken at a rapid rate. They were familiarized with the test stimuli by listening to six sentences, three at each rate, in mixed order. They then heard (without responding) 22 randomly ordered stimuli to provide familiarization with the excised stimuli. The test procedure was otherwise exactly parallel to the Sentence Condition.

B. Results

It was assumed that speaking rate information would be available to listeners when they heard sentence length utterances, but less available in syllables that had been excised from the sentences. This was found to be the case. In judging the rate of sentences, listeners made only 1.6% errors on average. In judging the rate of excised syllables, however, listeners made an average of 25.1% errors. A two-way (2×2) analysis of variance of the errors with stimulus condition (sentence versus syllable) as a between-subjects variable, and speaking rate (normal versus rapid) as a within-subjects variable, revealed a significant main effect of condition [$F(1,18)=281, p<0.001$]. There was no significant effect of speaking rate and no significant interaction between condition and rate. Analyses were also made to determine whether intrinsic vowel duration affected the accuracy with which listeners judged the rate of sentences and excised syllables. Vowels were grouped as intrinsically long /e, æ, ɑ, ɔ, o/, midlength /i, u/, and short /ɪ, ɛ, ʌ, ʊ/. For the Sentence Condition, a two-way (2×3) within-subjects analysis of variance revealed that neither of these variables affected rate judgments. Listeners judged the speaking rate of sentences with high accuracy irrespective of intrinsic vowel duration or speaking rate.

The analysis for the Syllable Condition similarly failed to produce significant main effects, but revealed a significant crossover interaction [$F(2,18)=89.6, p<0.001$]. For syllables containing long vowels, the speaking rate was misjudged more often when they were spoken at a rapid rate (41% errors) than when they were spoken at a normal rate (10% errors). For syllables containing short vowels, the speaking rate was misjudged more often when they were spoken at a normal rate (35% errors) than when they were spoken at a rapid rate (14% errors). Syllables containing midlength vowels showed the same pattern as short vowels (35% errors on normal-rate syllables and 16% errors on rapid-rate syllables).

These results suggest that speaking-rate judgments of isolated syllables were affected by a combination of speaking rate and intrinsic vowel length. To examine the extent to which syllable duration alone might account for the rate judgments, syllable duration (measured from the wave form) was correlated with the number of normal-rate judgments,

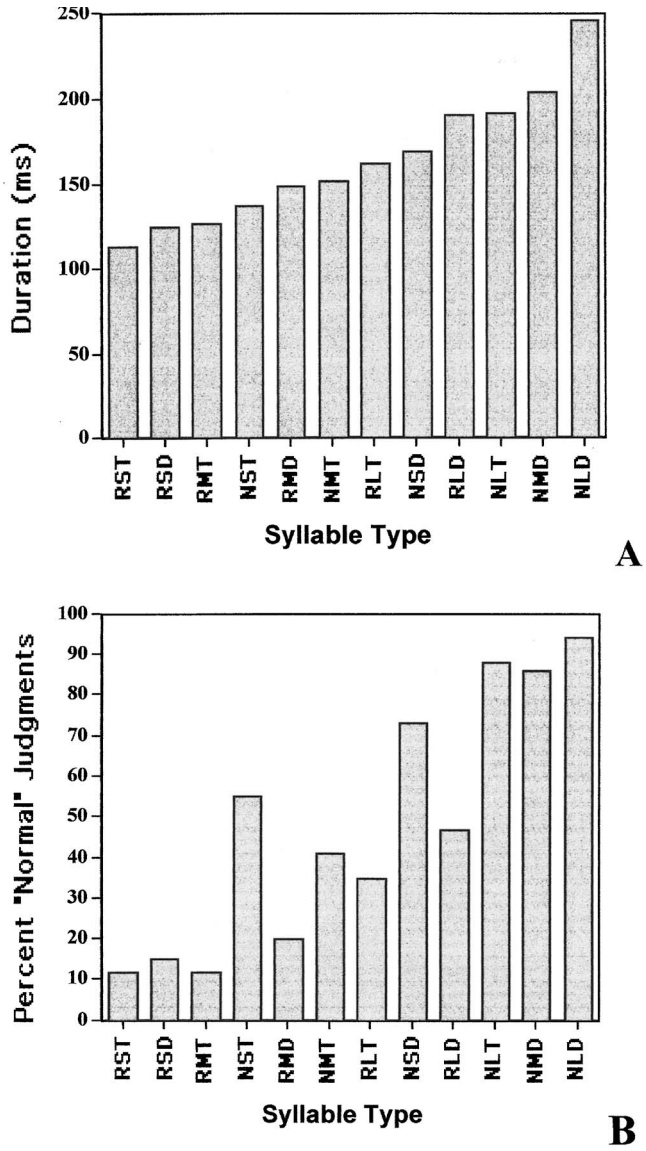


FIG. 1. Relationships between syllable duration (A) and judged speaking rate (B) for subsets of stimuli. (A) Average syllable duration (in ms). Sets of stimuli grouped by rate (N/R), intrinsic vowel duration (L/M/S), and final consonant (D/T) and arranged in order of increasing duration. (B) Percent of "normal rate" judgments for the subsets of stimuli arranged in the same order as in (A).

(averaged over the two presentations of each type) in each of its six consonantal contexts and two rates (132 points). The Pearson product-moment correlation coefficient was +0.77. Thus, measured duration of the syllable was associated with approximately 59% of the variance in rate judgments in this condition.

In order to investigate further, the percentage of "normal" responses for the Syllable Condition was examined on subgroups of syllables grouped by three factors known to affect syllable duration: Speaking rate, intrinsic vowel duration, and voicing of the final consonant. Figure 1(a) shows the average syllable duration of each of these subgroups. As would be expected, the shortest duration syllables are the rapidly spoken (R), short vowel (S), voiceless final consonant (T) syllables and the longest duration syllables are the normal-rate (N), long vowel (L), and voiced final consonant (D) syllables.

If syllable duration were the only criterion for rate judgments, then the proportion of normal responses, in comparison to "rapid" responses, should steadily increase as the syllable duration increased. Figure 1(b) shows the syllable subgroups displayed in the same order as in Figure 1(a) (from shortest to longest duration) but with the ordinate indicating the percent of normal judgments made for each syllable subgroup. It is obvious that there were several marked deviations from a monotonic increase in the perception of normal rate based solely on syllable duration. Those syllable types that were judged as normal *less often* than would be predicted by duration alone included rapid-rate midlength vowels and rapid-rate long vowels (columns 3, 5, 7, and 9). Those syllable types that were judged as normal *more often* than would be predicted by duration alone were the normal-rate short vowels (Columns 4 and 8). These results are consistent with the notion that part of the listener's judgment of rate is relative to the intrinsic duration of the vowel in the syllable. Thus, syllables containing intrinsically mid and long vowels spoken rapidly were heard as being rapidly produced in spite of their greater relative duration, and syllables with intrinsically short vowels produced at normal rate were heard as normal rate in spite of their shorter relative duration. These data show clearly that rate judgments were modulated by vowel identity.

IV. EXPERIMENT 2: VOWEL IDENTIFICATION

A. Method

1. Stimuli

The audio tapes (for Sentences and Syllable Conditions) from Experiment 1 were used for the first two conditions of Experiment 2. New audiotapes were prepared for a third condition, Blocked Syllables. These consisted of the excised syllables of the Syllable Condition but normal-rate syllables and rapid-rate syllables were presented separately, that is, the stimuli were blocked by rate. Thus, the stimulus materials for Experiment 2 consisted of the Sentence Condition (mixed normal- and rapid-rate sentences); the Syllable Condition (mixed normal- and rapid-rate excised syllables); and Blocked Syllables (excised syllables blocked by speaking rate).

2. Participants

Thirty phonetically naive listeners participated; ten were randomly assigned to each condition. Participants were volunteers who were awarded extra credit points for their participation. All were monolingual speakers of American English and reported normal hearing. (Nine additional listeners participated in the experiment, but their data were discarded because they did not meet the task familiarization criteria described below.)

3. Procedures

Stimuli were produced and delivered by the equipment described in Experiment 1. Response forms contained 264 entries on 24 pages (11 per page). Listeners identified each vowel by making a mark through a one-syllable word that corresponded to the vowel perceived. Words representing the

11 vowels were *eek, if, ape, (h)eck, as, ah, (h)awk, up, oh, (h)ook*, and *ooze*. For all conditions, listeners were familiarized with the task and response form as follows: The experimenter described the stimuli as CVC syllables embedded in the carrier sentence "I hear the sound of /həCVC/ some more." Listeners were told that three initial consonants (/b, d, g/) and two final consonants (/d, t/) would be used in the syllables. The experimenter read aloud the 11 vowels and explained the orthographic system used on the answer sheet. Each participant, in turn, repeated the 11 vowels as the experimenter noted the dialect of the speaker. The experimenter then read aloud 11 sentences containing the 11 vowels and listeners marked 11 responses on the first page of their answer booklets. The experimenter gave immediate feedback after each entry and attempted to clarify any confusions. Next, over earphones, listeners heard 11 recorded sentences spoken by a male voice, and marked responses. The experimenter then read the correct vowels for the 11 sentences. This was repeated for page 3. Finally, listeners heard 22 sentences and marked pages 4 and 5. They were then told the correct vowels for these 22 sentences. Only data from those listeners who missed no more than three vowels, excluding /ɑ-ɔ/ confusions,[1] and no more than one token of the same vowel were retained for further analysis.

For the Random Sentence Condition, participants were told that they would listen to the same vowels in sentences as in familiarization, but spoken by a female speaker. They listened to 22 sentences from the test corpus (11 normal rate and 11 rapid rate) in random order without responding. Listeners were then presented with the test stimuli and were asked to mark their responses in the answer booklet. For the Random Syllable Condition, stimulus familiarization was modified to accommodate the syllables in place of the sentences. For orientation regarding the normal and rapid rates, listeners first heard three normal and three rapid sentences spoken by the female speaker, in mixed order. From the test corpus, they then heard 22 excised syllables (11 normal and 11 rapid) in mixed order. The test was identical to that in the Random Sentence Condition except excised syllables were substituted for sentences. For the Blocked Syllable Condition, one-half of the participants listened without responding to three normal-rate sentences spoken by the female speaker, followed by 11 normal-rate excised syllables. They were then presented with 132 test stimuli of normal-rate syllables and marked their answer booklets. These listeners were then familiarized with the rapid-rate stimuli by listening to a corresponding set of three rapid-rate sentences followed by 11 rapid-rate excised syllables. For the test, they heard 132 test stimuli of rapid-rate syllables and responded in the answer booklet. The other half of the participants heard rapid-rate stimuli first, followed by normal-rate stimuli.

B. Results

The percent of errors in vowel identification was tabulated for the three conditions; (/ɑ-ɔ/) confusions were not counted as errors. Error rates are given for each vowel in each condition in Table I. The columns give the results for each experimental condition, subdivided by speaking rate of the stimuli. The rows show the vowels grouped by intrinsic vowel duration as short, mid, and long duration vowels. Listeners made only 2% errors overall in Randomized Sentences, 7% errors in the Randomized Syllables, and 4% errors in the Blocked Syllables. An analysis of variance revealed only a significant main effect of condition $[F(2,63)=5.6, p<0.01]$. A Fisher's Protected LSD test indicated a significant difference only between the Randomized Sentence Condition and the Randomized Syllable Condition. There was no significant difference between the Blocked Syllable Condition and either of the other two conditions. The remarkably low rate of errors in the Random Syllable Condition and the fact that the error rate did not differ significantly from the Blocked Syllable Condition suggest that explicit information concerning rate of speaking is not critical to vowel identification although there is still a small reliable advantage to having the entire original sentence when identifying vowels.

TABLE I. Errors in vowel identification (as a percentage of opportunities) for all stimulus conditions.

Vowel type	Sentences mixed		Syllables mixed		Syllables blocked	
	Normal	Rapid	Normal	Rapid	Normal	Rapid
			Short vowels			
ɪ	1	1	3	7	0	2
ɛ	2	3	10	6	1	3
ʌ	3	4	12	5	5	8
ʊ	12	8	23	18	11	14
			Midlength vowels			
i	0	1	0	3	0	3
u	0	2	8	7	5	8
			Long vowels			
e	1	0	1	12	0	4
æ	0	1	8	13	13	9
ɑ	1	5	1	3	0	0
ɔ	1	2	2	2	1	3
o	0	1	4	3	0	0
Total	2	2	6	7	3	5

When vowel errors were analyzed by intrinsic vowel duration, (short, mid, and long), no significant effects of rate were found in the ANOVA in any of the three conditions. It should be recalled that previous studies by Johnson and Strange (1982) and Verbrugge and Shankweiler (1977) found that rapidly spoken long vowels were significantly more often misclassified as spectrally similar short vowels when presented in isolation. However, in Johnson and Strange's Experiment II, the error rate for rapid long vowels was greatly reduced when listeners were informed concerning the way the stimuli had been produced and were given some examples of rapid sentences. The more extensive information and examples given to listeners in the present study may well account for much of the overall reduction in error rate. The improvement in performance, of course, resulted in ceiling effects, despite the fact that the consonantal context varied.

A comparison of the results of Experiments 1 and 2 reveals that when listeners were presented only the excised syllables, errors in identification of speaking rate were far

more frequent than errors in vowel identification. Further, the errors in judged speaking rate did not appear to be closely related to the identification of vowels in those conditions. Vowel identity was well maintained over variability in speaking rate even when extra-syllable information was not available.

Because listeners had little problem in identifying vowels in syllables excised from speech at different rates, despite variability due to consonantal contexts, one must ask what acoustic information enabled them to perform that task. What distinctive information for vowel identity was available at both rates of speech in the coarticulated syllables themselves? The next section addresses those questions.

V. ACOUSTICAL ANALYSES

Previous research implicates several sources of information for vowel identity available in coarticulated CVC syllables. First, the *vocalic nuclei* provides information about acoustic *vowel targets*, usually represented as the relative frequencies of the first two formants measured within a single spectral cross section at the acoustic midpoint of the syllable or at formant maxima/minima (van Son and Pols, 1990). However, target formant patterns of different vowels are often acoustically overlapping even when the productions of a single speaker are examined; that is, variability due to intraspeaker speech style, rate, and phonetic context render some target information ambiguous. Studies of the effects of speaking rate and style on the variability of vowel targets have often yielded conflicting results. According to an early model proposed by Lindblom (1963) and Lindblom and Studdert-Kennedy (1967), the amount of acoustic target undershoot was a simple function of syllable duration. More recently, however, it has been shown that the amount of target undershoot varies as a function of many variables: Individual differences in articulatory effort (Gay, 1981; Kuehn and Moll, 1976), the style of speech as "clear" versus "casual" (Moon and Lindblom, 1994), and whether syllable shortening is due to a change in lexical stress or speaking rate (Gay, 1978). On the other hand, some studies have reported little or no differences in midsyllable formant frequencies for stressed vowels spoken at normal and rapid rates in sentence materials (van Son and Pols, 1990, 1992; Verbrugge and Shankweiler, 1977). Targets for specific vowels spoken in multisyllabic utterances at normal and rapid rates often are similar to each other but vary significantly from "canonical" values, due to the effects of coarticulation with consonants (Hillenbrand *et al.*, 2001; Stevens and House, 1963; Strange, 1989).

Second, *intrinsic vowel duration* is important in the differentiation of coarticulated English vowels. Systematic differences in syllable durations were first documented by Peterson and Lehiste (1960) who classified English "single-target" vowels as intrinsically long, mid, or short. Perceptual studies of both natural and synthetic speech materials have verified that listeners utilize this information to disambiguate spectrally similar vowels (Gottfried *et al.*, 1990; Hillenbrand *et al.*, 2000; Huang, 1985; Strange, 1989).

Acoustic studies of coarticulated vowels produced in read sentences or continuous speech have usually shown that syllable durations decrease systematically with increases in speaking rate. However, results of various studies disagree with respect to the manner in which this compression is achieved. Several studies have reported proportional invariance of vowel durations over variations in stress and rate (Crystal and House, 1988, 1990; Gay, 1978; Johnson and Strange, 1982, Port, 1976), while other studies fail to find proportional compression (Caisse, 1982, Gay, 1981). Gopal (1990) argued that constant ratios are found only when durations of short and long vowels are averaged, and when the amount of change in speaking rate is relatively small (10 to 20%). Cummins (1999), Gopal (1990), Smith (2001), and Stack (1992) reported individual variations across speakers in the effects of speaking rate on vowel durations.

Third, the *direction and extent of movement of the first and second formants* (F1 and F2) within vocalic nuclei often differentiate English vowels that have similar spectral targets. Nearey and his colleagues (Andruski and Nearey, 1992; Nearey, 1989; Nearey and Assmann, 1986) have described distinctive patterns of *vowel-inherent spectral change (VISC)* even for so-called monophthongal Canadian English vowels spoken in isolation. More recently, Watson and Harrington (1999) have reported somewhat similar results for vowels in Australian English. (For further studies of the effects of VISC on perception, see Assmann and Katz, 2000, Hillenbrand *et al.*, 2001; and Hillenbrand and Nearey, 1999).

Fourth, *temporal formant trajectories* of CVC syllables have also been shown to provide perceptually relevant information for the identification of vowels. DiBenedetto (1989a, b) described perceptually relevant systematic differences in F1 temporal trajectories for front vowels differing in tongue height. She hypothesized that the temporal location of the maximum frequency of F1 (F1 Max), proportional to the total syllable duration, provided the relevant information. The *symmetry of temporal trajectories* has also been shown to influence vowel perception. Huang (1985) demonstrated that perception of synthetic vowel continua contrasting lax and tense vowels shifted as a function of whether the temporal trajectories were asymmetrical (longer final transitions than initial transitions) or symmetrical (equal duration of initial and final transitions). Watson and Harrington (1999) analogously reported for Australian English that the relative time to reach vowel target values was useful in discriminating tense and lax vowels that were spectrally similar. Strange and Bohn (1998), offer a similar discussion of North German vowels.

In a detailed study, van Son and Pols (1992) reported a comparison of normal and rapid rate in reading of a single text by an experienced newscaster. They analyzed vowel durations and formant frequencies measured at 16 points throughout the vocalic nuclei of Dutch vowels (normalized for overall duration). Durations of both long and short (nonreduced) vowels in stressed and unstressed sentence positions decreased systematically with an increase in speaking rate. However, vowel compression (15% on average) was less than the overall compression of the passage (25%), suggesting that the vowels were more resistant to durational compression than other phonemes, at least at the rates of speech used by their speaker. A most important finding in

relation to the present study was that there was little effect of speaking rate on formant frequencies, except for a small but consistent increase in F1 (about 20 Hz) in the fast-rate tokens of four of the vowels.

In the present analysis, the main question of interest was how these several kinds of information varied as a function of speaking rate and consonantal context in the Context corpus, and which, if any, of the acoustic parameters suggested above should be candidates for further examination as invariant information over rate and context.

A. Method

To examine the effects on target formant frequencies of speaking style (citation versus sentence), speaking rate (normal versus rapid) and initial consonantal context (glottal, labial, alveolar, velar), midsyllable formant frequencies of vowels in the Citation and Neutral corpora and the Context corpus were compared. The formant frequencies of the vowels produced in neutral context disyllables spoken in citation form (Citation corpus) were used as traditional measures of the speaker's "canonical vowel targets" (Peterson and Barney, 1952; Stevens and House, 1963; Hillenbrand *et al.*, 2001). Target formant frequencies for the same disyllables produced at normal and rapid rates in sentences (Neutral corpus) were compared to these canonical values to determine the relative contribution of speech style and speech rate to variability of vowel targets. Finally, comparisons of midsyllable formant values for vowels in the Neutral and Context corpora established the effects of coarticulation with the initial stop consonants on vowel targets.

For these measures involving formant frequencies at different temporal locations in the syllables (relative to total duration), and for comparisons of syllable durations, syllable onsets were defined as beginning with the burst of the initial stop consonant, or the onset of the voiced portion following /h/. Syllable offsets were defined as the beginning of closure of the final /d/ or /t/ of the syllable (as indicated on digital wave forms and spectrograms by the cessation of all but low-frequency voice murmur for /d/ and by silence for /t/. Spectral analysis was accomplished using both linear predictive coding (LPC) and fast fourier transforms (FFT). For LPC analysis, 14-pole analysis using a 7 ms Hamming window updated every 1 ms was utilized The windows that began at or nearest the beginning of a pitch period were then used to extract the frequencies and amplitudes of the first three formants at 1/3, 1/2, and 2/3 temporal points in the syllables.[2] While previous studies used the 1/4, 1/2, 3/4 points to analyze trajectories (e.g. Huang, 1985), this was not possible in our corpus because the 1/4 points for some of the shortest syllables were not within the voiced portion of the target syllable. In addition, we wanted to analyze the portion of the syllable least affected by the consonantal transitions in order to capture formant movement that was inherent to the vowel (Nearey, 1989).

FFT spectra were also computed at the same points for all three vowel corpora. These spectra served to check the validity of formant extraction from LPC parameters and to provide missing information when the LPC analysis had failed to detect a formant. Out of the total of 2376 points measured, there were 79 (3.3%) that LPC analysis failed to compute correctly. These were handcorrected from the observed spectral peaks on the FFTs and verified by examination of the spectrograms.

B. Results

The results of comparisons of formant targets for the three corpora (Citation, Neutral, and Context) are presented first, followed by a discussion of the variability of syllable durations as a function of speaking rate and consonantal context. Finally, VISC and F1 max parameters for the Context corpus used in the perceptual studies were examined.

1. Formant targets

Target formant frequencies for F1 and F2, measured at the temporal midpoints of the syllables, were compared across the Context, Citation, and Neutral corpora to investigate the variation of vowel targets as a function of speaking style, speaking rate, and consonant context. F1 and F2 values of syllables spoken at normal and rapid rates with both /t/ and /d/ final consonants in the Context corpus were examined. Results indicated very small and nonsystematic variation in vowel targets due to speaking rate and final consonant. In all three initial consonantal contexts, a comparison of normal versus rapid rate utterances resulted in only minor shifts in F1 and F2 values. With the exception of the F1 for /u/ in /d-/ context, average target values for normal and rapid syllables were within 1 Bark of each other on both F1 and F2. In agreement with van Son and Pols (1992), there was a tendency for the F1 of some of the high and midvowels to be slightly higher in rapid rate syllables than in normal rate syllables. Differences in average F2 values for normal and rapid rate syllables were even smaller. Given these small differences in target values across speaking rate in the Context corpus (and in the Neutral corpus, upon inspection of those data), the normal and rapid syllables with both /t/ and /d/ final consonants were pooled and averaged for the comparison of effects of speech style and initial consonant context.[3] Figure 2 presents the canonical F1/F2 data for the Citation corpus and both sentence corpora (Neutral and Context), with initial consonant context plotted separately for the Context materials.

When syllables in the Context and Citation corpora are compared, large discrepancies in F1 and F2 target values can be seen for several vowels. In general, there was a shrinking of the F1/F2 vowel space for the vowels produced in sentence materials, relative to the Citation-form disyllables. The F1 values of the vowels /æ, ɑ, ɔ/ in all three initial stop consonant contexts were more than 1 Bark lower than their canonical values as determined from Citation disyllables. In the case of /æ/, and /ɑ/, and this resulted in formant targets quite near to adjacent (short) vowels, /ɛ/, and /ʌ/, respectively. In general, the F1 and F2 values for disyllables of the Neutral corpus were more similar to the values for the disyllables of the Context corpus than to the values of the Citation corpus. The only exception is that /ɑ/ in the Neutral sentence condition was similar to its citation form value.

In summary, the comparisons presented in Fig. 2 show that there were large differences in formant targets for sev-

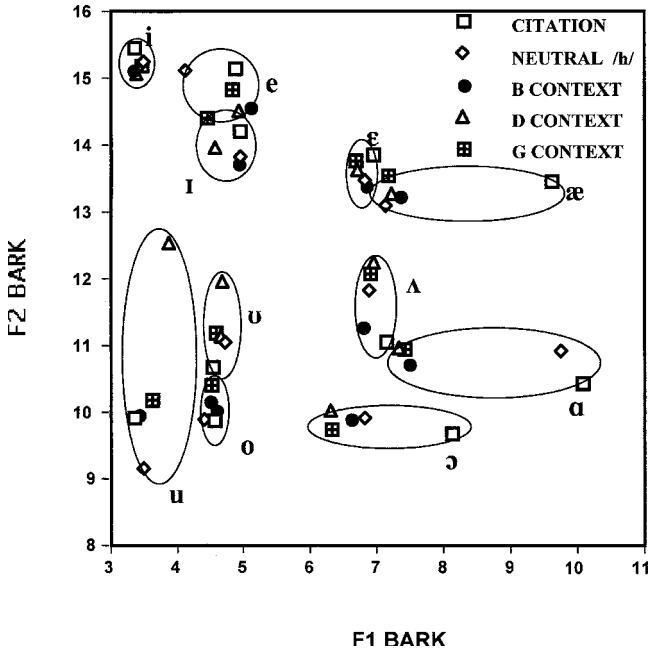


FIG. 2. Average F1, F2 target frequencies (Bark) for vowels in the Context, Citation, and Neutral corpora, collapsed over speaking rates and final consonant. Hand-drawn ellipses indicate instances of each vowel.

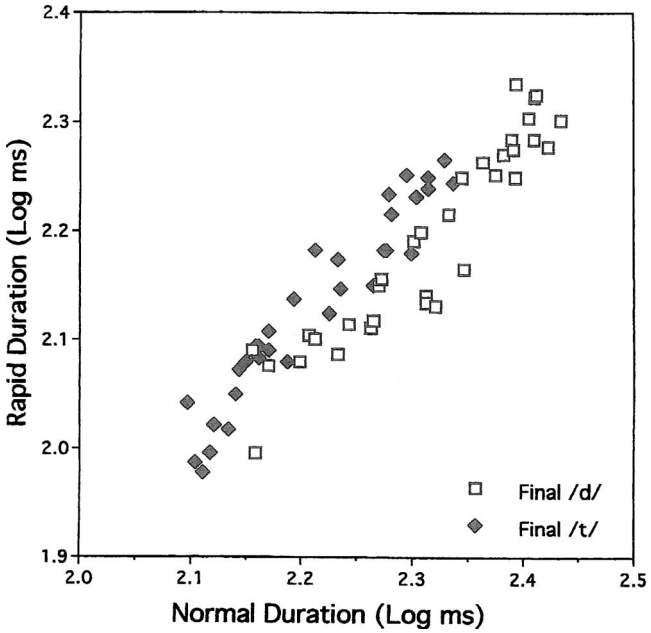


FIG. 3. Scatterplot showing the correlation between the durations (log ms) of Context corpus syllables spoken at normal and rapid rate. (CVd syllables—solid diamonds; $y=-0.164+1.018x$; CVt syllables—open squares; $y=-0.251+1.077x$).

eral vowels produced in disyllables in sentence context relative to canonical values, although the vowels varied as to which formant differed and the degree to which formants differed. In contrast, there were only minor differences in the midsyllable formant frequencies between rapid and normal rates of speech in sentence context. We conclude tentatively then that the largest perturbation of vowel targets was due to speaking style. For the sentence materials, there was an overall shrinking of the vowel space and target formant frequencies for four vowel pairs /ɪ-e, ɛ-æ, a-ʌ, o-ʊ/ overlapped (note that all four of these pairs contrast in intrinsic vowel duration). Target information for these contrasts can be considered ambiguous; indicating that listeners must employ other parameters, such as intrinsic duration or dynamic spectral information to identify the vowels correctly.

2. Syllable duration

The 264 sentences of the Context corpus showed systematic temporal effects of an increase in rate from the speaker's self-selected normal rate to a rapid rate. Rapid sentences were, on average, 69% as long as normal sentences. However, temporal compression was not uniform across the entire sentence, but decreased from the first stress foot to the second stress foot to the target syllable. The smaller amount of compression of target syllables was probably due to their being in the focal position of the sentence.

Study of the duration data showed that the absolute amount of shortening of syllables in rapid speech differed as a function of the voicing of the final consonant. This is to be expected if syllable durations maintain a *constant ratio* over changes in speaking rate (/-d/ syllables were longer than /-t/ syllables). To assess whether proportional durations remained constant over consonantal context, durations of each vowel in each context (averaged over the two repetitions) were converted to compression ratios (rapid/normal). The compression ratios did not vary systematically as a function of the initial consonant (80%, 79%, and 79% for /b/, /d/, and /g/, respectively). However, syllables ending in /t/ did not compress proportionally as much as syllables ending in /d/ (83% versus 73%, respectively). A two-way analysis of variance of these data revealed that this difference in mean compression ratio was statistically reliable [$F(1,60)=33.7, p<0.001$].

As reported earlier, Gopal (1990) suggested that averaging over long and short camouflages variability in vowel duration relationships, which is apparent when one inspects particular short/long vowel pairs in specific consonantal contexts. In order to assess how much of the variability of syllable duration was accounted for by the constant-ratio relationship between normal and rapid syllables, Pearson product-moment correlations were computed for syllables ending in /d/ and those ending in /t/, separately. Again, syllable durations were the average value for the two repetitions of each vowel in each context at each rate. On the assumption of a constant ratio, the correlation between the log duration of normal rate syllables and the log duration of rapid syllables should be linear with a slope of 1.0. Figure 3 plots the syllable durations for each normal/rapid vowel pair with syllables ending in /d/ and /t/ indicated by open and closed symbols, respectively. Both regression lines yielded slopes slightly greater than 1.0; correlation coefficients are $r=+0.956$ and $r=+0.938$ for /-t/ and /-d/ syllables, respectively. It is apparent that about 90% of the variance in syllable duration in a particular final consonant context was accounted for by the constant-ratio relationship between rapid and normal rate syllables.

3. Dynamic spectral information

To explore the effects of changes in speaking rate and consonantal context on formant trajectories, the direction and

extent of movement over the middle third of each syllable of the Context corpus were compared. To increase the reliability of formant measurements, values at 1/3, 1/2, and 2/3 points of each syllable in each context at each rate were averaged over the two repetitions. Inspection of the resulting trajectories indicated only small and unsystematic differences for syllables ending in /d/ and /t/. Therefore, formant trajectories were averaged over final consonants as well (four syllables/trajectory). Figure 4 shows these average trajectories for long vowels (left) and short vowels (right) for each consonantal context (labial top, alveolar middle, velar bottom). Normal and rapid rates are superimposed in each figure (normal—open arrows; rapid—stippled arrows).

With few exceptions, trajectories for particular vowels varied only slightly in direction and extent as a function of speaking rate. Differences in *extent* of movement from 1/3 to 2/3 points for normal and rapid rate trajectories of greater than 0.5 Bark for either F1 or F2 occurred in only 7 of 66 cases. Differences across rate in *direction of F1* movement were apparent in only 4 out of 66 cases; differences in *direction of F2* movement were found for only 2 out of 66 cases.

These figures illustrate that formant movement throughout the middle third of syllables containing short vowels tended to be proportionally greater than for syllables containing long vowels. When one considers the fact that their absolute durations were significantly shorter, it becomes clear that the *rate* of midsyllable formant movement for short vowels was strikingly greater than for most long vowels. This reflects the fact that short vowels produced in CVC syllables at normal and rapid rates by this speaker had no quasi-steady-state formants, nor even much decrease in the rate of movement of formants throughout the syllable nucleus. In contrast, for syllables containing long vowels, formant movement "slowed down" and, in some contexts at least, appeared to maintain quasi-steady-state values in the middle portion of the syllables.

Given that the patterns of formant movement appeared to be rather stable across changes in speaking rate, it was possible to examine the extent to which VISC remained invariant over (initial) consonantal context. For each vowel, 12 trajectories were available for analysis: 2 final consonants $\times 2$ rates $\times 3$ initial consonants. Formant trajectories, defined as the movement of F1 and F2 within the vocalic nucleus of each syllable, were examined for the Context corpus materials to determine the effects of both speaking rate and consonantal context on VISC and on F1 temporal patterns. With respect to VISC, the question of interest was whether the direction and/or extent of formant movement remained invariant over changes in speaking rate and over differences in the place of articulation of the initial consonant. Estimates of the relative position of F1 Max were used to determine the variability or invariance of this temporal parameter across speaking rate and consonantal context.

Patterns of VISC emerged as invariant over speaking rate in direction, but not always in extent. VISC for the long vowels /e, æ, o, u/ was toward the extremes of the vowel space. The extent of movement, as well as the starting and ending points of trajectories varied greatly with consonantal context for /u, o/. For /e, æ/ both direction and extent were similar across contexts. Trajectories for /i/ had very little movement in any context and trajectories for /ɑ, ɔ/ showed relatively small movements in varying directions as a function of consonantal context. For short vowels, the general direction of VISC was toward increasing values for F1 and toward the center of the vowel space for F2. The direction of F1 movement was consistent over the three contexts for all four vowels, although the extent of movement varied for /ɪ/. For F2, the direction was consistent across contexts for /ɪ, ɛ/ although the extent varied somewhat. In contrast, the diverging directions of F2 movement across contexts for /ʌ/, and to a lesser extent for /ʊ/, showed clear coarticulatory effects of the initial consonant.

These trajectory patterns indicate that the *direction* of formant movement in the middle third of the syllable for 6 of the 11 vowels produced by this speaker was invariant over rate and context, despite perturbations in the extent of that movement caused by coarticulatory influences of the initial consonant. (Of course, because the place of articulation of the final consonant was constant across all syllables in this corpus, we cannot be certain as to what extent the apparent invariance in direction of F2 movement was due to coarticulatory influences of the final consonant.) Of the remaining five vowels, /i/ showed little movement and indeed little perturbation of "target" values. The differences in direction of F2 movement for /ɑ, ɔ, ʌ, ʊ/ suggest that coarticulatory effects overrode any movement patterns that were intrinsic to these vowels. Movements of both formants (when present) were not inherent to the vowel, but rather varied as a function of contextual influences.

Turning now to temporal trajectories (F1 Max), the values of F1 at the 1/3, 1/2, and 2/3 points in the Context syllables were used to estimate this parameter. Since there were no systematic differences in F1 Max for a particular vowel across rate or consonantal context, the 12 instances of each vowel were compared to see whether there were consistent differences across vowels in F1 Max. No relationship between vowel height and F1 Max was found in the present data for either the front vowels or for the back vowels. However, there were systematic differences in the pattern of F1 Max for long and short vowels. All four short vowels were characterized by F1 patterns which reached their highest frequency in the second half of the syllable (modal value of F1 Max was at the 2/3 position). Furthermore, this pattern was consistent across the 12 syllables. To evaluate the reliability of these patterns statistically, the distribution for each vowel was assessed against the null hypothesis of a flat distribution (4, 4, 4) by the Chi Square test. The obtained distributions for all four short vowels were reliably different from the predicted chance distribution ($p < 0.01$). In contrast, the long vowels, except for /æ/, showed inconsistent patterns of F1 Max (/i, e, ɑ, ɔ/) or distributions with modal values early in the syllable (/o, u/). Short vowels could be differentiated from their spectrally adjacent long vowels, on average, by the temporal pattern of F1 for all pairs except for /i, ɪ/ and /ɛ-æ/. It is interesting to note that the long vowel /æ/ had an F1 Max pattern very similar to the short vowels (modal value at 2/3, significantly different from flat, $p < 0.01$). In English, this vowel is long in intrinsic duration, but is classified as

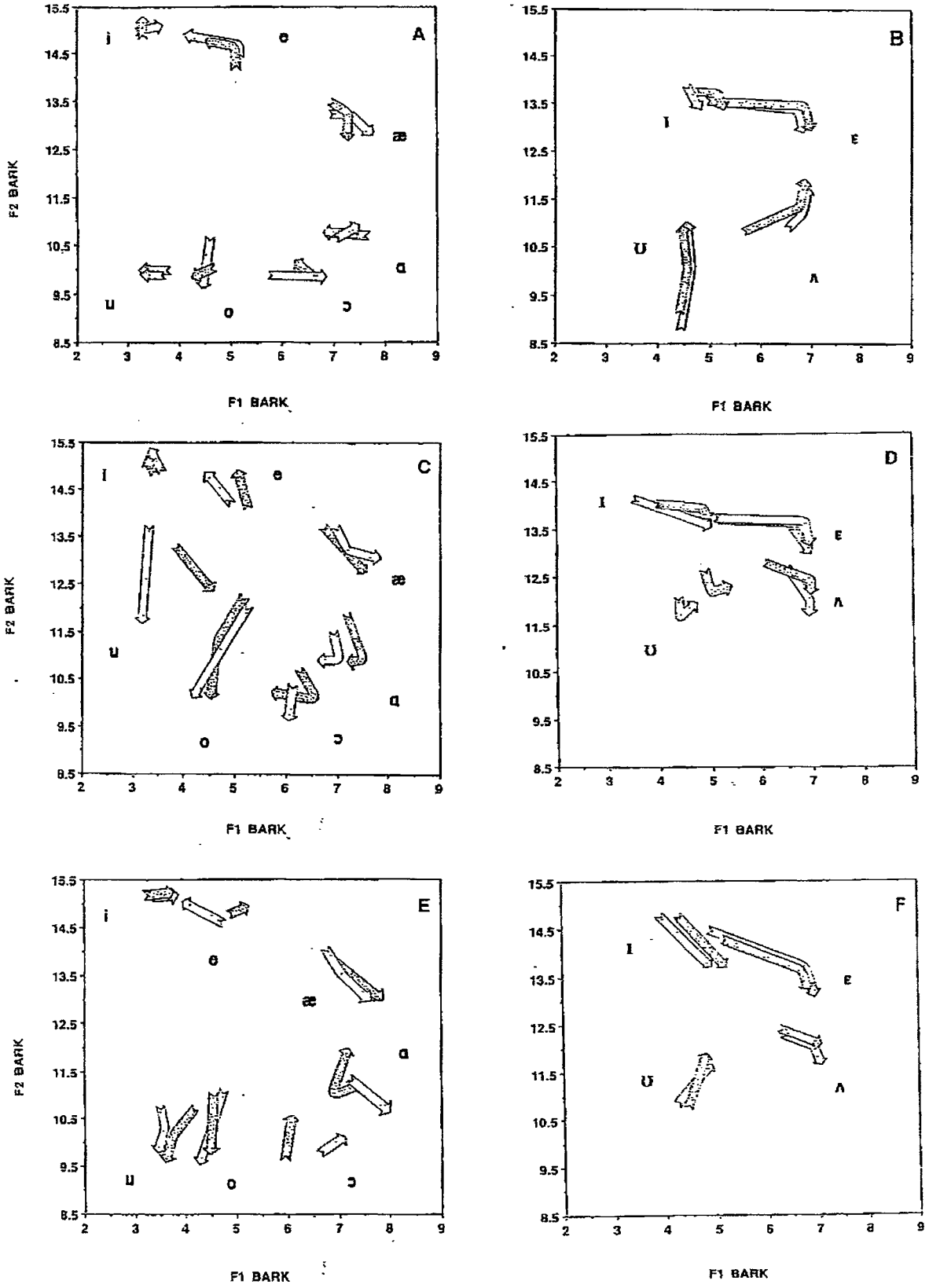


FIG. 4. Formant trajectories over the middle third of the Context corpus syllables, averaged over repetitions and final consonants. Long and mid vowels—left; Short vowels—right. Initial consonant: Labial—top; Alveolar—middle; Velar—bottom. Normal rate syllables—open arrows; Rapid rate syllables—stippled arrows. Arrows point to the values at the 2/3 location in the syllables.

"lax" in most phonological descriptions. We can conclude that temporal patterns of F1 provided distinctive information about tense and lax vowels with similar formant targets in this corpus. However, F1 Max did not provide consistent information about vowel height.

The above analyses indicate that the dynamic spectral parameters of VISC and F1 Max varied systematically as a function of the identity of the vowels, although not necessarily in the ways suggested in the previous literature. The change in speaking rate produced little variation in either

direction or extent of formant movement in the middle third of the syllables. However, the extent of movement (proportional to total duration) for several of the vowels differed markedly as a function of consonantal context. F1 Max for lax vowels was invariant across changes in speaker's rate, and the overall distributions for tense versus lax vowels yielded a highly significant difference when tested by Chi Square ($\chi^2=56.14, p<0.001$).

VI. DISCUSSION

A. Specific results

This study examined the perception of a large set of American English vowels under conditions in which speaking rate, consonantal context, and presence or absence of the sentence frame were varied. Speaking rate could be correctly classified as normal or "fast" by phonetically naïve listeners when full sentences were presented, but listeners made many errors in classifying speaking rate when only the excised syllables were presented. In spite of their inability to specify speaking rate explicitly, however, listeners were able to identify the vowels both in sentences and in the excised syllables at high levels of accuracy. These findings raise the question as to whether there is any need for special processing to cope with rate of speech. They direct attention to the information in the syllables themselves and raise questions concerning the intrasyllabic information for vowel identity.

Acoustic analyses showed that formant targets of coarticulated CVC syllables differed appreciably from the traditional canonical targets derived from vowels spoken in citation form. Remarkably, however, there was little difference between the vowel spaces for syllables produced in sentences at normal and rapid rates. The 31% increase (on average) in sentential speaking rate (which produced an average of a 20% decrease in durations of the target CVC syllables) led to very little *further* undershoot of canonical targets. The formant values of the vowels in Citation form syllables appear to be reliably different from the formant values achieved in sentential speech at whatever rate.

Midsyllable formant values for four spectrally-adjacent vowel pairs were ambiguous when the vowels were produced in sentence-length utterances. Yet the vowels were well identified. Consequently, the study examined other sources of information that have been shown to be related to vowel identification in past studies.

Absolute durations of CVC syllables varied systematically with vowel identity, speaking rate, and the voicing of the final consonant. Within a particular (final) consonantal context, the temporal compression of syllables was proportionally equal (i.e., a constant ratio), such that the relative durations of intrinsically short and long vowels in the same context were invariant over changes in rate. Thus, we can assume that relative duration information was available to disambiguate long and short coarticulated vowels with ambiguous target values.

Other dynamic spectrotemporal information also differentiated spectrally similar long and short vowels, and tended to be invariant over both speaking rate and consonantal context. VISC was shown to be invariant in direction, though not in extent, over changes in speaking rate and consonantal context for most of the vowels. While there was little effect of speaking rate on the extent of VISC for this speaker, significant variability was found with phonetic context in the extent of movement of F2 in the back rounded vowels /o, ʊ, u/ and the direction of movement of F2 for /ʌ/ (and to a lesser degree /ʊ/). Nevertheless, three of the four vowel pairs with overlapping formant targets could be differentiated by context-independent direction of VISC.

F1 Temporal Maximum (F1 Max) also differed systematically with vowel identity. The late F1 Max pattern characteristic of all five lax vowels /ɪ, ɛ, æ, ʌ, ʊ/ was consistent over changes in speaking rate and consonantal context, and contrasted with early F1 Max patterns associated with the tense vowels /e, o, u/ F1 Max patterns for the remaining tense vowels showed inconsistent patterns over rate and context, due in large part to the fact that F1 values changed very little over the middle third of the syllables containing these vowels.

From the above summary of the invariant patterns of VISC and F1 Max, one can see that the two descriptions yield the same results with respect to differentiation of three of the four vowel pairs with overlapping formant targets. For /ɛ-æ/ neither VISC nor F1 Max were distinctive; these vowels were differentiated only by their (context-dependent) intrinsic duration differences. It remains for future research to determine to what extent these patterns of VISC and F1 Max remain invariant over different final consonant contexts, different manners of articulation of initial and final consonants, and different speakers. Recent research by Hillenbrand and Nearey (1999) and Neel (2004), is beginning to explore how much detailed information on formant movement is necessary for accurate vowel identification, but we cannot yet say whether the directionally invariant VISC patterns shown in this corpus are sufficient to support differentiation of vowels with similar spectral targets for other speakers.

B. General findings

On the basis of the acoustical analysis of the three corpora, some general conclusions can be drawn about the effects of speaking rate and consonantal context on the acoustic specification of vowels for this speaker. First, in agreement with van Son and Pols (1992), the change in speaking rate had relatively minor effects on either formant targets or dynamic spectrotemporal parameters that differentiate coarticulated vowels. Second, changes in syllable duration as a function of speaking rate were highly consistent, supporting a (context-dependent) constant ratio hypothesis. Third, as is well known, consonantal context had complex effects on intrinsic durations, formant targets, and spectrotemporal trajectories. Compression ratios varied as a function of the voicing of the final consonant. However, the rapid/normal ratio for syllables ending in /d/ and /t/ differed by a nearly constant amount (in log units) over the range of syllable durations within each rate. There were large effects of the place of articulation of the initial consonants on target formant values, reflecting coarticulatory influences of alveolar (and velar) consonants on F2 for back vowels similar to those reported elsewhere (Hillenbrand *et al.*, 2001; Stevens and House, 1963; Strange 1989).

Overall, the most striking difference in vowel target information was revealed in the comparison between disyllables produced in sentence context and those produced as citation-form utterances. In other words, vowels in citation-form utterances were much more differentiated than vowels coarticulated in continuous speech contexts. This difference in "speaking style" produced much more perturbation of spectral information than either consonantal context or speaking rate.

1. Limitations

It must be recognized that these data, like those of van Son and Pols (1992), refer to the speech of a one speaker. The speaker was not selected as exceptional, however, and the accuracy data for the speaker are very much like those for other speakers we have used in past studies. The generality of the findings must, however, be limited pending further replication and extension.

ACKNOWLEDGMENTS

This research was funded by NIDCD Grant No. DC-00323 to one of the authors (W.S.). The authors wish to thank research assistants Manuela Aguilar, Salvatore Miranda, and Alba Evans Rodriguez for their help with collecting and analyzing data for this study.

[1]Confusion between /ɑ/ and /ɔ/ was not counted as errors because this distinction was not made in the dialect of many of our listeners.

[2]The small window, positioned such that the "ringing" portion of each pitch period was weighted most heavily, was used on the advice of Kenneth Stevens, in order that rapid formant movements would not be distorted by averaging over more than one pitch period. The disadvantage of this analysis may be an increase in measurement error for the lowest formant for some vowels.

[3]Values for final /d/ and /t/ syllables were averaged because inspection of the data showed only small and unsystematic differences between their target values. The average differences for F1 for the Citation corpus were 32 Hz (SD=40) and 65 Hz for F2 (SD=59). The average differences for the Context corpus were 34 Hz for normal rate F1 (SD=37), 48 Hz for normal rate F2 (SD=47), 33 Hz for F1 in rapid rate (SD=31), and 51 Hz for F2 in rapid rate (SD=47).

Andruski, J. E., and Nearey, T. M. (**1992**). "On the sufficiency of compound target specification of isolated vowels and vowels in /bVb/ syllables," J. Acoust. Soc. Am. **91**, 390–410.

Assmann, P. F., and Katz, W. F. (**2000**). "Time-varying spectral change in the vowels of children and adults," J. Acoust. Soc. Am. **108**, 1856–1866.

Caisse, M. (**1982**). "Context-induced vowel duration change and intrinsic vowel duration." J. Acoust. Soc. Am. **72**, (S1) S65 (A).

Crystal, T. H., and House, A. S. (**1988**). "The duration of American-English vowels: An overview," J. Phonetics **16**, 263–284.

Crystal, T. H., and House, A. S. (**1990**). "Articulation rate and the duration of syllables and stress groups in connected speech," J. Acoust. Soc. Am. **88**, 101–112.

Cummins, F. (**1999**). "Some lengthening factors in English speech combine additively at most rates," J. Acoust. Soc. Am. **105**, 478–480.

DiBenedetto, M. G. D. (**1989a**). "Vowel representation: Some observations on temporal and spectral properties of the first formant frequency," J. Acoust. Soc. Am. **86**, 55–66.

DiBenedetto, M. G. D. (**1989b**). "Frequency and time variations of the first formant: Properties relevant to the perception of vowel height," J. Acoust. Soc. Am. **86**, 66–77.

Gay, T. (**1978**). "Effect of speaking rate on vowel formant movements," J. Acoust. Soc. Am. **63**, 223–230.

Gay, T. (**1981**). "Mechanisms in the control of speech rate," Phonetica **38**, 148–158.

Gopal, H. S. (**1990**). "Effects of speaking rate on the behavior of tense and lax vowel durations," J. Phonetics **18**, 497–518.

Gottfried, T. L., Miller, J. L., and Payton, P. E. (**1990**). "Effect of speaking rate on the perception of vowels," Phonetica **47**, 155–172.

Hillenbrand, J. M., and Nearey, T. M. (**1999**). "Identification of resynthesized /hVd/ utterances: Effects of formant contour," J. Acoust. Soc. Am. **105**, 3509–3523.

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.

Hillenbrand, J. M., Clark, M. J., and Nearey, T. M. (**2001**). "Effects of consonantal environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748–763.

Huang, C. (**1985**). "Perceptual correlates of the tense/lax distinction in General American English," Master's thesis, MIT Cambridge, MA.

Johnson, T. L., and Strange, W. (**1982**). "Perceptual constancy of vowels in rapid speech," J. Acoust. Soc. Am. **72**, 1761–1770.

Kuehn, D. P., and Moll, K. L. (**1976**). "A cineradiographic study of VC and CV articulatory velocities," J. Phonetics **4**, 303–320.

Lindblom, B. E. F. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.

Lindblom, B. E. F., and Studdert-Kennedy, M. (**1967**). "On the role of formant transitions on vowel recognition," J. Acoust. Soc. Am. **42**, 830–843.

Miller, J. L. (**1981**). "Effects of speaking rate on segmental distinctions," in *Perspectives in the Study of Speech*, edited by P. D. Eimas and J. L. Miller (Erlbaum, Hillsdale, NJ), pp. 39–74.

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.

Nearey, T. M. (**1989**). "Static, dynamic, and relational factors in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Nearey, T. M., and Assmann, P. F. (**1986**). "Modeling the role of inherent spectral change in vowel identification," J. Acoust. Soc. Am. **80**, 1297–1307.

Neel, A. T. (**2004**). "Formant detail needed for vowel identification," ARLO **5**(4), 125–131.

Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.

Piterman, M. (**2000**). "Effect of speaking rate and contrastive stress on formant dynamics and vowel perception," J. Acoust. Soc. Am. **107**, 3425–3437.

Port, R. F. (**1976**). *Influence of Speaking Tempo on Duration of Stressed Vowel and Medial Stop in English Trochee Words* (Indiana University Linguistics Club, Bloomington, IN).

Smith, B. L. (**2001**). "Effects of speaking rate on temporal patterns of English," J. Acoust. Soc. Am. **110**, 2736.

Stack, J. W. (**1992**). "Effects of speaking rate and stress on vowel durations and formant structures," Master's thesis, University of South Florida (unpublished).

Stevens, K. N., and House, A. S. (**1963**). "Perturbation of vowel articulations by consonantal context: An acoustical study," J. Speech Hear. Res. **6**, 111–128.

Strange, W. (**1989**). "Dynamic specification of coarticulated vowels spoken in sentence context," J. Acoust. Soc. Am. **85**, 2135–2153.

Strange, W., and Bohn, O.-S. (**1998**). "Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies," J. Acoust. Soc. Am. **104**, 488–504.

van Son, R. J. J. H., and Pols, L. C. W. (**1990**). "Formant frequencies of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **88**, 1683–1693.

van Son, R. J. J. H., and Pols, L. C. W. (**1992**). "Formant movements of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **92**, 121–127.

Verbrugge, R. R., and Shankweiler, D. P. (**1977**). "Prosodic information for vowel identity," J. Acoust. Soc. Am. **61**, (Suppl. 1), S39.

Watson, C. I., and Harrington, J. (**1999**). "Acoustic evidence for dynamic formant trajectories in Australian English vowels," J. Acoust. Soc. Am. **106**, 458–468.

Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification

Mitchell S. Sommers[a)]
Department of Psychology, Washington University, Saint Louis, Missouri 63146

Joe Barcroft[b)]
Department of Romance Languages, Washington University, Saint Louis, Missouri 63146

(Received 26 May 2005; revised 11 January 2006; accepted 11 January 2006)

Three experiments were conducted to examine the effects of trial-to-trial variations in speaking style, fundamental frequency, and speaking rate on identification of spoken words. In addition, the experiments investigated whether any effects of stimulus variability would be modulated by phonetic confusability (i.e., lexical difficulty). In Experiment 1, trial-to-trial variations in speaking style reduced the overall identification performance compared with conditions containing no speaking-style variability. In addition, the effects of variability were greater for phonetically confusable words than for phonetically distinct words. In Experiment 2, variations in fundamental frequency were found to have no significant effects on spoken word identification and did not interact with lexical difficulty. In Experiment 3, two different methods for varying speaking rate were found to have equivalent negative effects on spoken word recognition and similar interactions with lexical difficulty. Overall, the findings are consistent with a phonetic-relevance hypothesis, in which accommodating sources of acoustic-phonetic variability that affect phonetically relevant properties of speech signals can impair spoken word identification. In contrast, variability in parameters of the speech signal that do not affect phonetically relevant properties are not expected to affect overall identification performance. Implications of these findings for the nature and development of lexical representations are discussed.
[DOI: 10.1121/1.2171836]

PACS number(s): 43.71.Bp [ARB] Pages: 2406–2416

I. INTRODUCTION

Considerable evidence is now available to suggest that when listeners are presented with speech signals varying in acoustic characteristics the resultant acoustic-phonetic variability can significantly impair both the identification of, and memory for, spoken words (Mullennix *et al.*, 1989; Sommers *et al.*, 1994; Nygaard *et al.*, 1995; Bradlow *et al.*, 1999). Mullennix *et al.* (1989), for example, reported that trial-to-trial changes in talkers resulted in poorer spoken word recognition relative to conditions in which the items were all spoken by the same talker. Moreover, reduced speech intelligibility from acoustic-phonetic variability is not limited to changes associated with different talker characteristics. Sommers *et al.* (1994), for instance, found that intraspeaker variability, in the form of trial-to-trial changes in speaking rate, produced approximately a 15% decline in identification performance compared with conditions in which words were all spoken at the same articulation rate. Similarly, Uchanski and Braida (1998) reported that another form of intraspeaker variability, token variability, could reduce vowel identification even when the talker and speaking rate remained constant. In their study, participants heard 1, 4, or 16 different utterances of a given vowel and identification performance was assessed using a two-interval forced-choice procedure. Uchanski and Braida observed significant declines in identification performance as a result of increasing from 1 to 4 tokens, but little change when the number of tokens was further increased from 4 to 16 exemplars.

Although the preceding findings suggest that a number of different sources of acoustic-phonetic variability can affect spoken word identification, it is important to note that not all types of variability impair speech perception. Sommers *et al.* (1994), for example, reported that variations in overall amplitude had little effect on spoken word identification. In explaining these findings, Sommers *et al.* suggested that acoustic-phonetic variability will impair spoken word identification only when that variability alters acoustic properties that are used for phonetic perception. According to this proposal, henceforth referred to as the phonetic-relevance hypothesis, talker, rate, and token variability all affect phonetic perception because each produces changes in acoustic features, including formant frequencies and transitions that are important cues for phonetic identification. In contrast, variations in overall amplitude result primarily from differences in the distance between talker and listener and therefore do not affect acoustic properties that serve as the primary cues for identifying spoken words (however, see Dorman and Dougherty, 1981).

One purpose of the present study was to further investigate the phonetic-relevance hypothesis by extending the re-

[a)]Electronic mail: msommers@wustl.edu
[b)]Electronic mail: barcroft@wustl.edu

 0001-4966/2006/119(4)/2406/11/$22.50

sults of previous investigations of stimulus variability to two other sources of intraspeaker variability, variations in speaking style[1] (normal, denasalized,[2] whispered, excited, elongated, and pitch shifted) and fundamental frequency (F0). Speaking style and F0 were selected as sources of variability for two reasons. First, as neither source of variability has been examined previously in studies of stimulus variability and perceptual identification, the findings provide new information about how different sources of acoustic-phonetic variability affect speech perception. Most importantly, however, the experiments provide a strong test of the phonetic-relevance hypothesis because the speaking style changes used in Experiment 1 introduced considerable variability in formant frequencies and other phonetically relevant spectrotemporal acoustic properties. Consequently, trial-to-trial variations in speaking style were expected to produce poorer word identification than conditions in which all items were spoken with a single speaking style. In contrast, the variations in F0 investigated in Experiment 2, were generated to affect only voice pitch, with minimal or no changes in formant frequencies or other acoustic properties used for phonetic identification. This type of F0 manipulation is not contrastive for the types of stimuli used in Experiment 2 [monosyllabic consonant-vowel-consonants (CVCs) for native English speakers] and, therefore, the F0 variations introduced in Experiment 2 were not expected to reduce spoken word recognition compared to single-F0 conditions. Evidence consistent with these proposals would suggest that listeners engage a perceptual mechanism to accommodate sources of variability that affect phonetically relevant features of the speech signal and that the additional demands associated with accommodating variability can impair spoken word identification. However, sources of acoustic variability, such as the specific type of F0 manipulation used in the present study, that do not affect phonetically important speech features are processed more automatically and therefore have minimal effects on speech perception.

A second goal of the study was to determine whether variations in speaking style and F0 interact with phonological similarity, as assessed by neighborhood density and neighborhood frequency. Neighborhood density has been operationalized (Luce and Pisoni, 1998) as the number of words that can be created from a target word by adding, deleting, or substituting a single phoneme (e.g., "COT," "HAT," and "CAB" are all neighbors of the word "CAT"). Neighborhood frequency refers to the average frequency of occurrence for items in a phonological neighborhood. A number of studies have demonstrated that "hard" words—items from high-density high-frequency neighborhoods (i.e., word that are phonologically similar to many other high-frequency words)—are identified less accurately and with longer latencies than "easy" words—items from low-density low-frequency neighborhoods (Pisoni *et al.*, 1985; Luce and Pisoni, 1998; Sommers and Danielson, 1999). Sommers (1997) reported a significant interaction between talker variability and lexical difficulty, such that talker variability had significantly greater negative effects on identification of hard words than on easy words. Sommers attributed this interaction to increased processing demands when listeners had both to identify lexically difficult words and simultaneously accommodate acoustic-phonetic variability. Furthermore, consistent with the proposal that variations in overall amplitude are processed more automatically, Sommers (1997) failed to find an interaction between amplitude variability and lexical difficulty (i.e., the effects of amplitude variability were comparable for lexically easy and hard words).

The present study was designed to determine whether parallel findings would be obtained for speaking-style and F0 variability. That is, variability in speaking style was expected to interact with lexical difficulty in a manner similar to that observed for talker variability, whereas no interaction was predicted between F0 variability and lexical difficulty. This pattern of results would be consistent with the proposal that the processes used to accommodate sources of variability that affect phonetically relevant acoustic features (e.g., talker and speaking style) are similarly affected by increased perceptual demands. Analogously, evidence that sources of variability which do not alter phonetically relevant speech features (e.g., F0 and amplitude variability) fail to interact with lexical difficulty would support more automatic processing in accommodating these sources of variability.

II. EXPERIMENT 1

A. Methods

1. Participants

Sixty undergraduate students (37 men and 23 women) served as participants and received course credit for their participation. All of the participants were native speakers of English and none reported a history of speech or hearing disorders.

2. Stimuli

The stimuli consisted of 144 highly familiar monosyllabic words. The words were a subset of the 150 words used by Sommers (1998).[3] All items had a familiarity rating (Nusbaum *et al.*, 1984) of 6.2 or higher on a 7-point scale. The mean familiarity rating for all 144 items was 6.8 [standard deviation (SD)=0.3]. The 144 words consisted of 72 items from low-density low-frequency neighborhoods (easy words) and 72 from high-density high-frequency neighborhoods (hard words). Neighborhood density values were obtained from the Hoosier Mental Lexicon (Nusbaum *et al.*, 1984). The mean number of neighbors was 10.8 and 26.3 for the easy and hard words, respectively. Mean neighborhood frequency (i.e., the mean frequency of all items in the neighborhood, excluding the target word) was 243.7 for hard words and 55.8 for easy words. Sommers (1996) provided a distribution of stops, fricatives, affricates, nasals, and liquids in both word-initial and word-final position for the complete 150-word set. The only significant differences that were observed was that fricatives were significantly more common (both word-initial and word-final) in easy than in hard words (this was also true for the subset of 144 words used in the current study). Neighbors were defined operationally as any word that could be created from the target stimulus by adding, deleting, or substituting a single phoneme. Mean word frequency (Kucera and Francis, 1967) was equated as closely

as possible for easy [mean (M)=43, SD=8.9] and hard (M =52, SD=11.4) words in order to examine the effects of neighborhood characteristics (density and frequency) independent of word frequency.

A male talker produced all 144 words in six different speaking styles (normal or conversational), excited, nasal, elongated (items were produced at a slow speaking rate), childlike (talker was asked to produce items "as if speaking like a young child"), and whispered. Each of the recorded words was produced in a carrier phrase "please say the word_____," where the blank corresponded to the intended item. The talker recorded all words while seated in a double-walled sound-attenuating booth. The productions were transduced using a Universal Serial Bus (USB) microphone, a 16-bit analog/digital (A/D) converter and a 44.1 kHz sampling rate. Target words were edited from the carrier phrase using a commercial wave form editor and the root-mean-square amplitude of all words was digitally equated. Thus, unlike the amplitude manipulation in Sommers *et al.* (1994), all of the stimuli in the current experiment were presented at approximately the same overall amplitude. The range of formant frequency values for the first five formants averaged across all 144 words for each of the six speaking styles was: F1: 449–996; F2: 1520–2123; F3: 2402–3106; F4: 3543–3827; and F5: 4105–4558. As expected, speaking style variability resulted in considerable differences in formant frequency values (as well as other acoustic parameters), especially for the first and second formants. Recall that the working hypothesis for the current experiment was that accommodating for this variability should impose additional processing demands and therefore produce poorer performance for the multiple, compared with the single speaking style condition.

Two pilot studies were conducted to determine whether the six speaking styles were perceptually distinct and to assess the quality of each production as an example of a given speaking style. In the first study, all 864 tokens (144 words $\times 6$ speaking styles) were presented in random order to four speech-language pathologists from the Central Institute for the Deaf at Washington University School of Medicine. The four participants had a minimum of 11 years experience as licensed speech-language pathologists. Participants were first asked to assign the production to one of the six speaking style categories. After this decision, they were asked to rate on a 7-point scale "how good an example the token was of their selected speaking style." Correct categorization (categorizing consistent with the speaking style designated at production) exceeded 98% for all four participants. Mean category goodness ratings (7 indicating an ideal example and 1 indicating a poor example) were as follows: nasal=6.7, whispered=6.9, excited=6.5, neutral=6.7, elongated=6.9, and childlike=6.9.

In the second pilot study, we examined whether the categorization results with the speech-language pathologists would generalize to a group of participants similar to those who would participate in the proposed experiments. Ten young adults from the student population at Washington University were asked to perform the speaking style categorization task.[4] Mean correct categorization for the group of students exceeded 96%, with three participants obtaining perfect scores. Taken together, the results of these two pilot studies indicate that the six speaking styles were easily distinguished and that the productions represented good examples of each speaking style.

3. Design and procedure

The experiment used a completely within-subjects design, with variability (single speaking style versus multiple speaking styles) and lexical difficulty (easy versus hard words) as the independent variables. The 144 words were divided into two sets with an equal number of easy and hard words in each set (i.e., each set contained 36 easy and 36 hard words). Assignment of easy and hard words to each set was done pseudo-randomly without replacement. A given participant received one set of words in the single speaking style condition and one set of words in the multiple speaking style condition. One-half of the participants received the single speaking style condition first, and one-half received the multiple speaking style condition first. Assignment of the two word sets to single versus multiple speaking style conditions was counterbalanced such that equal numbers of participants received each word set in the single and multiple speaking style conditions. Within the single speaking style condition, equivalent numbers of participants heard one of the six speaking styles (i.e., ten participants heard words produced in an excited speaking style for the single speaking style condition, ten heard words produced in a whispered speaking style for the single speaking style condition, etc.). Within the multiple speaking style condition, words were rotated across speaking styles such that equal numbers of participants heard a given word in one of the six speaking styles. Thus, the principal difference between the single and multiple speaking style conditions was that all of the words in the single speaking style condition were presented in the same speaking style, whereas in the multiple speaking style condition words were produced using all six speaking styles.

At the start of each trial, participants were instructed to ″press the space bar when ready for next trial." After a delay of 750 ms, a word was presented in a background of six-talker babble at a signal-to-babble ratio of 0 dB. The word file was digitally mixed with a file of the same duration containing the babble, converted to an analog signal using a 16-bit D/A converter and a 44.1 kHz sampling rate, and presented at approximately 70 dB sound pressure level (SPL) over headphones (Sennheiser HD250) to both ears. Presentation levels refer to the SPL measured at the output of the transducer in a NBS-9A 6-cm^3 coupler with a Bruel and Kjaer sound level meter (Model No. 2204) set on the linear weighting scale. Participants were instructed to listen to the word and to make their best guess as to the word that was presented. Participants then wrote their responses and pressed the space bar when they were ready for the next word. Scoring was based on exact phonetic matches such that words differing from the target item by even a single phoneme were counted as incorrect. Thus, homophones were counted as correct responses but misspellings that included a phoneme error were counted as incorrect as there was no way to distinguish misspellings from misperceptions.

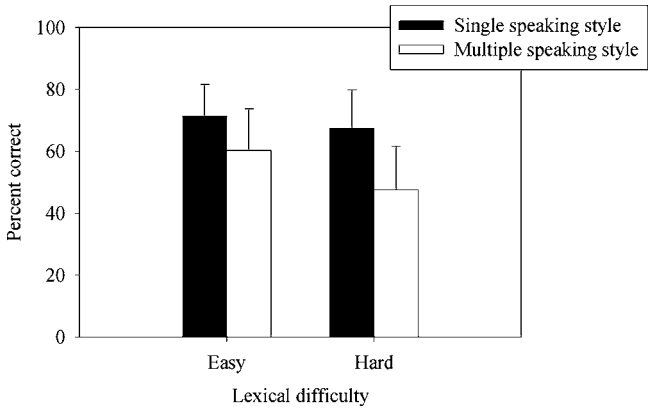


FIG. 1. Mean percent correct for words presented in single (dark bars) and multiple (clear bars) speaking-style conditions as a function of lexical difficulty. Error bars depict standard deviations of the mean.

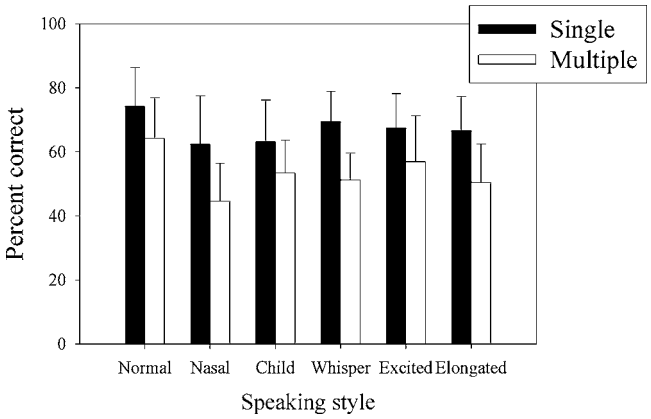


FIG. 2. Mean percent correct for the six different speaking styles used in Experiment 1 for the single (dark bars) and multiple (clear bars) speaking style conditions. Error bars depict standard deviations of the mean.

B. Results and discussion

1. Effects of speaking style variability and lexical difficulty

Figure 1 displays means (and standard deviations) for percent correct identification in the single and multiple speaking style conditions as a function of lexical difficulty. A 2 (lexical difficulty: Easy versus hard) $\times 2$ (variability: Single versus multiple speaking style) repeated measures analysis of variance (ANOVA) was conducted to examine the main effects and interactions between the variables. Consistent with a number of previous studies (Sommers, 1996; Luce and Pisoni, 1998), lexically easy words were identified better (M=65.8) than lexically hard words (M=57.8) $F(1,59)=35.2$, $p<0.001$. One of the novel findings from the current study is that, similar to variability in talker characteristics and speaking rate, speaking style variability also reduced identification performance relative to the single speaking style condition (mean difference=15.4) $F(1,59)=146.1$, $p<0.001$. Finally, the interaction between lexical difficulty and stimulus variability was also significant $F(1,59)=9.3$, $p<0.001$, indicating a greater reduction in identification performance for hard than for easy words as a consequence of changing from single to multiple speaking styles.

2. Intelligibility of words for each speaking style in single and multiple speaking style conditions

One of the advantages of varying speaking style within participants is the ability to compare performance for the different speaking styles across the single and multiple speaking style conditions. That is, the design allowed us to compare the performance for the ten participants who received, as an example, the whispered speaking style in the single speaking style condition with the performance for whispered items presented to the same participants in the multiple speaking style condition. Note that for each speaking style, data for the conditions with and without variability are based on different numbers of presentations. In the single speaking style conditions, 10 participants heard the 72 stimuli in the normal speaking style, 10 participants heard the stimuli in a whispered speaking style, etc. Therefore, data for the single speaking style conditions are based on 72 presentations for each participant (720 total presentations). In the multiple speaking style condition, however, each participant heard only 12 stimuli in each speaking style (120 total presentations). Despite this limitation, the data provide useful information regarding the effects of context (single versus multiple speaking styles) on identification performance for words spoken in each of the speaking styles.

Figure 2 displays percent correct identification performance for words produced in each of the speaking styles in the single and multiple speaking style contexts. Pair-wise comparisons for each speaking style indicated significant differences between the single and multiple speaking style conditions for all six speaking styles: Normal $t(9)=3.1$; nasal $t(9)=4.5$; whispered $t(9)=6.1$; excited $t(9)=3.7$; childlike $t(9)=4.2$; and elongated $t(9)=6.4$ ($p<0.001$ for all comparisons except for normal speaking style, $p<0.01$).

Taken together, the results of Experiment 1 contribute to the growing body of evidence that accommodating for sources of variability that affect phonetically relevant acoustic features of speech signals produces significant decrements in spoken word identification. In the current study, not only was speech perception poorer overall when items were produced by multiple speaking styles than when they were spoken with a single speaking style, but this pattern was obtained for all six speaking styles. In Experiment 2, we wanted to extend these results by examining whether a source of variability that did not affect phonetically relevant properties of the speech signal would also produce decrements in word identification. To date, only a single source of variability—variations in overall amplitude—has failed to produce performance decrements in spoken word identification. If similar null results are obtained in Experiment 2 with variations in F0, this finding would provide additional support for the proposal that only sources of variability that affect phonetically important acoustic cues will impair spoken word identification.

III. EXPERIMENT 2

A. Method

1. Participants

Sixty undergraduate students (33 men and 27 women) served as participants and received course credit for their

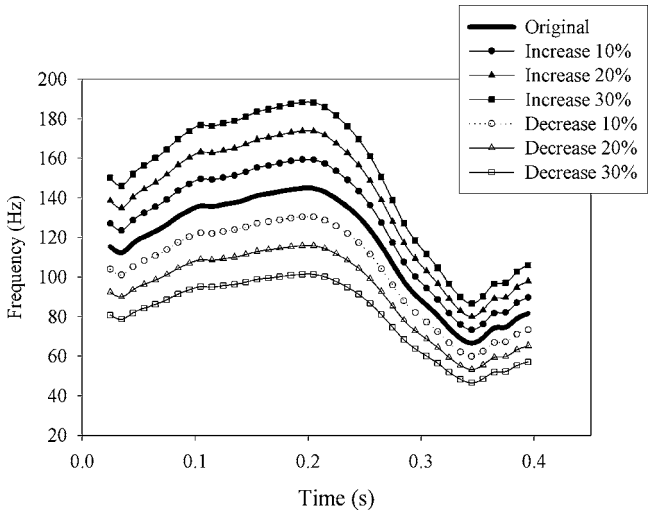


FIG. 3. Average pitch tracks for the original (solid black line) stimuli and six modified versions of the original. Pitch tracks depicted with solid symbols are for modified version with increments in the fundamental frequency and those depicted with open symbols show pitch tracks for modified versions with decrements in fundamental frequency.

participation. All of the participants were native speakers of English, and none reported a history of speech or hearing disorders. None of the participants in Experiment 2 participated in Experiment 1.

2. Stimuli

The same 144 monosyllabic consonant-vowel-consonants (CVCs) (72 easy and 72 hard) used in Experiment 1 were used in Experiment 2. However, rather than varying the speaking style, six new instances of each word were created by changing the F0 of the stimuli produced in the neutral or conversational voice. The six new instances of each word were created by incrementing and decrementing the F0 contour of the neutral-voice stimuli by 10%, 20%, and 30% using the Praat software package (Boersma, 2001). Modification of the F0 contours was accomplished by first extracting the pitch track for each word spoken in the neutral speaking style, changing the values for each time window to the desired value, and then resynthesizing the stimuli. The pitch track for each new stimulus was then extracted to ensure that it differed by the appropriate amount from the original. Thus, three sets of the new stimuli had F0 contours higher than the original neutral-voice stimuli, and three sets had F0s lower than the original. Pitch tracks were obtained for both the original and F0-altered stimuli, and the tracks for the 144 words produced at each F0 were averaged. Figure 3 displays these averaged pitch tracks for the original stimuli and the six F0-altered variants. To ensure that the F0 changes did not significantly alter formant frequencies, formant tracks were generated for the original stimuli and all six variants of each word using the Praat software. The individual formant tracks were used to compute mean formant frequencies for the first four formants of the original version and the six F0-altered variants. In all cases, the differences between the original and F0-altered formants were less than 1%.

To determine if the differences in F0 were perceptually salient to listeners, we conducted a pilot study with 20 participants. In each trial, participants were presented with two identical words (e.g., the word "bag" was presented twice) and had to decide whether the two words "had the same overall voice pitch." In one-half of the trials, the voice pitch was the same, with 18 presentations for each of the 6 different voice pitches (e.g., participants heard the 10% increase in F0 paired with the 10% increase in F0 18 times). In the remainder of the trials, participants again heard the same word presented twice but differing by 10% in overall F0.[5] Using the current stimuli, a total of four possible pairings provided a 10% difference in F0 (10% increase versus 20% increase, 20% increase versus 30% increase, 10% decrease versus 20% decrease, and 20% decrease versus 30% decrease). Participants received 27 trials for each of these pairs with the bigger variation presented first on approximately one-half of the trials, and the smaller variation presented first on approximately one-half of the trials. The specific words that were used in each pairing were selected randomly for both same-F0 trials and different-F0 trials. Thus, participants in the pilot study received a total of 216 trials, with one-half having the same F0 and one-half having a 10% change in F0. Participants were instructed to respond as quickly and accurately as possible whether the two stimuli presented had the same or different pitch. Mean percent correct for same-F0 and different-F0 trials were 98.1 and 99.2, respectively. Mean response latencies for the two types of trials did not differ significantly (618 and 637 ms, respectively, for the same and different trials). The results of the pilot study suggest that the 10% change in F0 was readily detectable by participants and that listeners were able to judge whether two stimuli had the same or different F0.

3. Design and procedure

The equipment, design, and procedures were identical to Experiment 1, except that rather than comparing identification performance for single versus multiple speaking styles, Experiment 2 compared conditions with single versus multiple F0s. As in Experiment 1, ten participants heard each of the six F0 variants in the no variability condition.

B. Results

Figure 4 displays percent correct identification as a function of lexical difficulty. A two-way repeated-measures ANOVA analogous to the one conducted in Experiment 1 revealed that identification scores were poorer for hard words than for easy words $F(1,59)=36.1$, $P<0.001$. In contrast to the findings with multiple speaking styles, however, identification scores did not differ for conditions with single versus multiple F0s $F(1,59)<1$, $p=0.41$. Furthermore, no significant interaction was observed between lexical difficulty and stimulus variability $F(1,59)=2.2$, $p=0.14$.

Figure 5 displays percent correct identification for stimuli presented in the single and multiple F0 conditions as a function of the extent and direction of change in F0. As was the case in Experiment 1, scores in the single F0 condition are based on a total of 720 presentations (72 presentations $\times$ 10 participants for each single F0 condition),

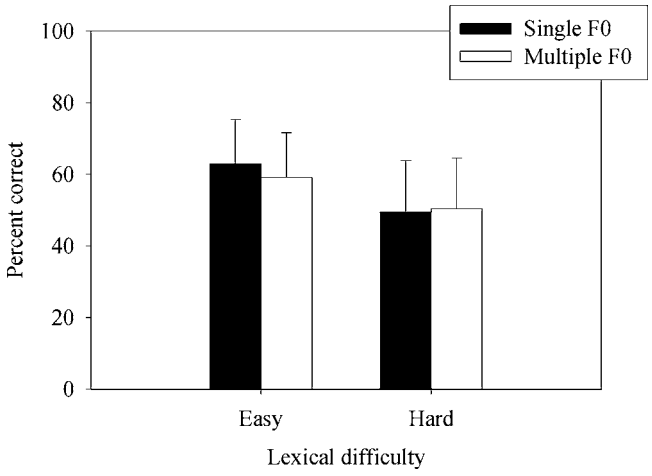


FIG. 4. Same as Fig. 1 except data are from Experiment 2 comparing conditions with single versus multiple F0s as a function of lexical difficulty.

whereas scores for the multiple F0 condition are based on only 120 presentations (12 presentations × 10 participants). A within-participants ANOVA with three factors (single versus multiple F0, increase versus decrease in F0, and amount of change in F0: 10%, 20%, or 30%) did not reveal any significant main effects or interactions. This finding suggests that identification performance was not affected by the amount or direction of F0 change and, furthermore, that equivalent performance was obtained for items presented in single and multiple F0 conditions. Separate pair-wise comparisons for each level of direction and extent of F0 change as a function of single versus multiple F0 context also failed to reveal any significant effects. Thus, the absence of significant differences between single and multiple F0s observed in the overall analysis was also observed when comparing individual levels of direction and extent of F0 change.

IV. EXPERIMENT 3

The purpose of Experiment 3 was to provide an initial test of an alternative explanation for the findings from Experiments 1 and 2 and previous results on acoustic-phonetic

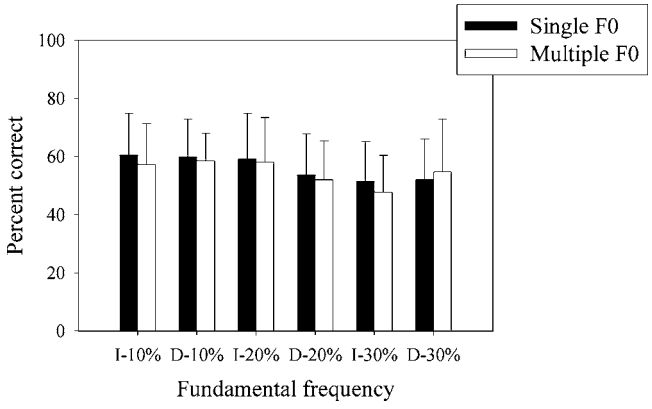


FIG. 5. Same as Fig. 2 except data are from Experiment 2 comparing identification performance for the six different fundamental frequencies presented in single F0 or multiple F0 conditions. Abscissa labels refer to increments (I) or decrements (D) in F0 relative to the original unaltered versions. Percentages after each I or D refer to the amount of increment or decrement in F0 relative to the original.

variability. Specifically, the findings from the current experiments on variability in speaking style and F0, combined with previous results on talker, rate, and amplitude variability (Mullennix *et al.*, 1989; Sommers *et al.*, 1994), are all consistent with the phonetic-relevance hypothesis in that sources of variability that do not affect phonetically relevant properties of the speech signal (amplitude and F0) do not impair spoken word recognition. Conversely, sources of variability that do affect phonetically relevant parameters of speech signals (talker characteristics, speaking rate, and speaking style) reduce perceptual identification, compared to conditions without such variability. A potential alternative account of these findings, however, is that the differential effects on identification performance simply reflect differences in the extent of acoustic variation resulting from a particular source of variability.[6] That is, variability in F0 (Experiment 1) and overall amplitude (Sommers *et al.*, 1994), were introduced by varying a single acoustic parameter and neither source of variability produced decrements in spoken word recognition. In contrast, variability in speaking style, speaking rate, and talker characteristics all alter multiple properties of the speech signal and each has been found to impair perceptual identification.

Experiment 3 was designed to provide an initial test of this alternative explanation by comparing the effects of naturally and digitally produced variations in speaking rate. Speaking rate was selected as the source of variability for two reasons. First, previous results (Sommers *et al.*, 1994; Sommers, 1997) found that naturally produced speaking rate variability reduces perceptual identification performance, compared to conditions without rate variability. Second, stimuli with naturally and digitally produced rate variations can be matched for overall duration but still vary with respect to the extent that each affects acoustic properties of the speech signal. Naturally produced variations in speaking rate (i.e., asking talkers to produce stimuli at different speaking rates) produce changes in multiple acoustic features, including both temporal (e.g., transition durations, voice-onset times) and spectral (e.g., vowel-space expansion) parameters that are important for phonetic identification. Experiment 3 compared the effects of this naturally produced rate variability with digitally produced rate variations that were introduced by removing (rate compression) or duplicating (rate expansion) steady-state portions of the speech signal (see below for additional details). According to the phonetic-relevance hypothesis, listeners should be predisposed to adapt to variations in speaking rate, independent of how those variations are produced, because in natural situations rate variations can affect phonetically relevant speech parameters. If, however, the absence of effects with F0 variability in Experiment 2 and with overall amplitude variability in Sommers *et al.* (1994) result because variability in these two parameters entails less extensive acoustic changes, then naturally produced rate variability should have greater negative effects than digitally produced variations on spoken word recognition.

A. Methods

1. Participants

One hundred and twenty undergraduate students (57 men and 63 women) were recruited from the same participant pool as in Experiments 1 and 2. One-half (60) of the participants (28 males and 32 females) heard the naturally produced rate alterations, and one-half of the participants (26 males and 34 females) heard the digitally produced rate alterations. All participants were native speakers of English. None of the participants reported a history of speech or hearing disorders and none had taken part in either of the two previous studies.

2. Stimuli

The same 144 monosyllabic words used in Experiments 1 and 2 were used in Experiment 3. The first step in stimulus generation was to get productions with naturally produced rate variations. A male talker (a different male than used in Experiment 1) recorded three variants of each word in the same carrier phrase as in Experiment 1. Prior to recording, the talker was told that he would be asked to record the stimuli at three distinct speaking rates (normal or conversational, slow, and fast). The talker was instructed to produce three distinct speaking rates, but no other instructions about how to produce the different speaking rates were provided. Each speaking rate was recorded in a separate block with the normal-rate items recorded first, the fast rate second, and the slow rate last. Equipment and procedures for recording were the same as those in Experiment 1. Average durations for the words were 780 ms (range: 675–815 ms), 460 ms (range: 427–501 ms), and 323 ms (range: 285–351 ms) for the slow, medium, and fast items, respectively.

The digitally rate altered stimuli were produced by either deleting (rate compression) or copying (rate expansion) steady-state portions of the medium-rate naturally produced tokens. The duration of each digitally altered stimulus was adjusted to match the corresponding values for the naturally produced fast and slow rate stimuli. Thus, for example, to produce the digitally altered versions of the word CAT, steady-state portions from the medium-rate recording of this item (duration of 482 ms) were either removed to match the duration of the fast-rate natural recording (319 ms) or duplicated to match the duration of the slow-rate version (710 ms duration). Steady-state portions of the stimuli were identified by visual inspection of the wave form and spectrogram. The steady-state portions were removed or duplicated in segments ranging from 10 to 25 ms and these segments always started and ended at zero crossings in the wave form. We specifically avoided altering transition or burst information.

Several pilot studies were conducted to assess the naturalness and saliency of the rate changes for the stimuli used in Experiment 3. The first pilot experiment was conducted to determine if listeners could detect differences in the naturalness of the two methods used to alter speaking rate. Fifteen young adults were presented with pairs of the 288 rate-altered words (recall that the medium- or normal-rate words were not altered). One member of each pair was the natural version of the stimulus and the other member was the digitally produced version of the same stimulus. In one-half of the trials, the natural stimulus was presented first, and in the other half the digital version was presented first. Participants were told that one member of each pair was a naturally produced stimulus and the other was a digitally altered version and were instructed to identify the digitally produced version. Overall percent correct was approximately 54.2, which was very close to chance level performance (50%). This finding suggests that the digitally altered stimuli were not perceived as less natural than the naturally produced rate alterations.

The second pilot experiment was conducted to determine if the natural and digital rate alterations had differential effects on intelligibility. Ten young adults were asked to identify one complete set of the 144 rate-altered items (i.e., 10 participants heard the 144 naturally produced fast items, a different set of 10 participants heard the 144 digitally produced fast items, etc.) in quiet at approximately 70 dB SPL. Mean percent correct for the digital and natural fast-rate stimuli was 80.1 (SD=4.4) and 81.8 (SD=7.9), respectively. Corresponding values for the slow-rate stimuli were 93.5 (SD=2.4) and 92.5 (SD=3.0). These results suggest that the two methods for varying speaking rate had similar effects on overall intelligibility.

The final pilot study was conducted to ensure that listeners perceived three distinct speaking rates for both the natural and digital rate alterations. An additional 20 subjects were recruited to judge the speaking rate of the 432 items (144 words $\times$ 3 speaking rates) produced using each method (10 participants judged the natural stimuli and 10 judged the digital stimuli). Although the medium-rate items were identical for both groups, they were included to increase the difficulty of the categorization task. Stimuli were again presented at approximately 70 dB SPL. Correct rate categorizations for the natural stimuli were 94%, 87%, and 89% for the slow-, medium-, and fast-rate words, respectively. Corresponding values for the digital alterations were 96%, 84%, and 93%. Considering that the listeners had no prior experience with the talker's voice and heard only isolated words, these relatively high accuracy levels indicated that the three speaking rates used in the present study were quite distinct and that this was true for both the natural and digital rate alterations.

3. Design and procedure

Experiment 3 used a mixed design with type of alteration (digital versus natural) as a between-participants variable, and condition (single versus multiple) and lexical difficulty (easy versus hard words) as repeated measures variables. In the single-rate conditions, participants heard all of the stimuli at the same speaking rate, with 20 participants hearing each of the three speaking rates for both the natural and digital rate alterations (i.e., 20 participants in the natural alteration and 20 participants in the digital rate alteration heard the fast-rate speech for the single-rate condition). Therefore, as noted earlier, only the fast- and slow-rate stimuli differed across the digital and natural alteration conditions. In the multiple-rate condition, participants heard one-third of the stimuli at each rate, with the rate for a given

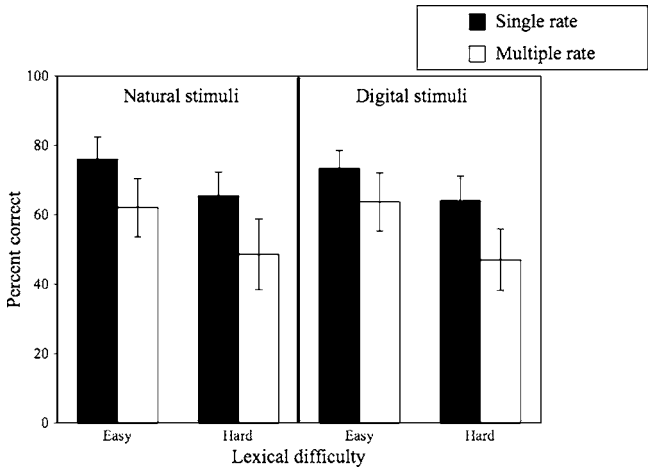


FIG. 6. Mean percent correct identification for the digitally and naturally produced rate alterations in Experiment 3 as a function of stimulus variability and lexical difficulty. Error bars indicate standard deviation of the mean.

trial selected pseudo-randomly. Within each type of rate alteration, one-half of the participants received the multiple-rate condition first, and the other half heard the single-rate condition first. Lexical difficulty was also distributed equally across conditions.

4. Results and discussion

Figure 6 displays percent correct identification for the two types of stimuli as a function of both stimulus variability and lexical difficulty. A three-way mixed ANOVA with stimulus variability (single versus mixed rate) and lexical difficulty (easy versus hard words) as repeated measures variables and stimulus type (natural versus digital) as an independent-measures variable was conducted on the data from Experiment 3. Consistent with previous results on rate variability (Sommers *et al.*, 1994), performance was significantly better in the single-rate (M=68.7) than in the multiple rate-condition (M=56.5) $F(1,118)=423.4$, $p<0.001$. As expected, easy words (M=59.7) were identified significantly better than hard words (M=55.5) $F(1,118)=305.5$, $p<0.001$. Of particular importance to the current study, neither the main effect of stimulus type $F(1,118)=1.6$, $p>0.2$ nor the interaction between stimulus type and stimulus variability $F(1,118)=1.9$, $p>0.1$ was significant. These findings indicate that performance did not differ for the naturally (M=63.1) and digitally (M=62.2) produced rate alterations and, most importantly, that the effects of stimulus variability were similar for the two types of rate variability.

V. GENERAL DISCUSSION

The results of the current study provide important new information about sources of variability and their effects on spoken word recognition. Specifically, variation in speaking style was found to impair recognition of spoken words and had greater negative effects on identification of hard words, compared with easy words. Moreover, the negative effects of speaking-style variability were observed in the within-participants analyses even for the neutral or conversational speaking style. That is, when comparing performance of the same individuals for items spoken in a conversational speaking style, performance was significantly poorer when the words were presented within the multiple speaking style condition than within the single speaking style condition. In contrast, the type of F0 variability used in Experiment 2 (global changes in pitch tracks) did not significantly affect spoken word recognition nor did it interact with lexical difficulty. Finally, the results of Experiment 3 demonstrating comparable effects of naturally and digitally produced variations in articulation rate on spoken word recognition are consistent with the proposal that phonetic relevance rather than extent of acoustic variation is the primary factor determining what types of variability do or do not affect speech perception.

Considered with previous findings on the effects of variability on speech perception (Mullennix *et al.*, 1989; Sommers *et al.*, 1994; Nygaard *et al.*, 1995), the current results add to the emerging picture that certain sources of variability, including changes in talker characteristics, speaking rate, and speaking style, can significantly impair identification of spoken words, whereas trial-to-trial variations in other features, including overall amplitude and F0, do not significantly alter spoken word recognition. This pattern of findings is consistent with the phonetic-relevance hypothesis in which sources of variability that affect phonetically relevant properties of the speech signal, such as formant frequencies and transitions, will degrade spoken word recognition. On the other hand, sources of variability, such as changes in overall amplitude and the F0 manipulations used in the current study, which do not alter parameters of the speech signal important for phonetic identification, will not impair identification performance for spoken words.

Before discussing theoretical implications of the current findings, it is important to consider two issues related to phonetic or phonemic relevance and how they limit potential conclusions from the present study. First, all three current experiments as well as most previous studies of stimulus variability (Mullennix *et al.*, 1989; Sommers *et al.*, 1994; Nygaard *et al.*, 1995) have used isolated monosyllabic words. Consequently, the notion of phonetic relevance and its possible role in explaining differential effects across sources of acoustic-phonetic variability applies only to the CVC monosyllables that have been the focus of most research. For example, both the F0 and amplitude variations that have been found to have null effects on monosyllabic word identification might have significant effects on more extended speech samples where such variations can be phonetically or phonemically relevant. In the case of F0, for instance, variability in this dimension may have significant effects at the sentence level where it is a critical feature for establishing prosodic structure (Cutler and Darwin, 1981). Similarly, changes in overall amplitude may not have direct phonetic consequences for phoneme identification in monosyllabic words but may be phonetically relevant as a correlate of word-level stress (Norris *et al.*, 1995). Although these considerations regarding phonetic relevance in different contexts limit the generalizability of the present findings, they also offer a potentially strong test of the phonetic-relevance hypothesis. Specifically, if phonetic relevance is a principal determinant of whether a given source of variability has

negative consequences for language processing, then the phonetic-relevance hypothesis would predict opposing results for the same source of variability in contexts where it does not have direct phonetic or phonemic consequences (e.g., overall amplitude in monosyllabic words) and those where it does (e.g., overall amplitude for multisyllabic words).

A second issue related to phonetic relevance that is important for the current discussion is a consideration of the results and limitations of Experiment 3. Specifically, Experiment 3 provided an initial investigation of whether the pattern of results with different sources of acoustic-phonetic variability could be accounted for based on the extent of acoustic variation each source introduces into the speech signal as opposed to the phonetic relevance of those changes. This proposal is a compelling potential alternative to the phonetic-relevance hypothesis because the sources of variability that have been shown to have null effects on spoken word recognition (overall amplitude, F0) introduce significantly fewer acoustic changes than sources that have negative effects on speech perception (speaking rate, speaking style, talker characteristics). Moreover, results from a study examining the perceptual consequences of intraspeaker variability (Newman *et al.*, 2001) suggest that identification latencies increase for talkers with greater acoustic variability in producing a phonetic category. The results of Experiment 3 indicate that, at least for the case of speaking-rate variability, reducing the extent of acoustic variations (by changing from natural to digitally produced changes in articulation rate) does not modulate the negative effects of rate variability.

For several reasons, however, these results should be considered as only a first step in evaluating the relative importance of extent of variation and phonetic relevance as factors contributing to whether a given source of variability will have negative effects on spoken word recognition. First, the findings of Experiment 3 were restricted to speaking rate as a source of variability and, therefore, do not rule out the importance of extent of acoustic variation as a factor determining how other sources of variability will affect spoken word recognition. Second, the extent of variation and phonetic relevance are not mutually exclusive, and both may function to determine whether a given source of variability has negative effects on speech perception. These considerations suggest that future research will be required to distinguish between the phonetic-relevance hypothesis and alternative explanations such as the extent of variation proposal.

Keeping these concerns in mind, the results of the current experiments and findings from previous research on acoustic-phonetic variability (Mullennix *et al.*, 1989; Mullennix and Pisoni, 1990; Sommers *et al.*, 1994; Nygaard *et al.*, 1995) have important implications for determining the nature and development of lexical representations. In particular, considerable debate exists regarding the extent to which indexical properties of the speech signal, such as talker characteristics and speaking rate, are retained in long-term lexical representations (Pisoni, 1993; Goldinger, 1996; Sheffert, 1998). According to one recent proposal (Luce *et al.*, 2003), form-based lexical representations contain both abstract (without indexical information) and episodic (with indexical information) codes that are differentially activated based on the time course of lexical access. Specifically, Luce *et al.* suggested that when there is a straightforward mapping between acoustic-phonetic information and lexical representations, such that lexical access can occur relatively quickly, then abstract representations will dominate spoken word recognition. In contrast, when the time course of lexical processing is slower, Luce *et al.* argued that episodic information will have an opportunity to exert its effects and will serve as the primary basis for responding during spoken word recognition.

In support of this time-course hypothesis, Luce and colleagues have conducted a number of studies examining the effects of allophonic and indexical variability on long-term priming. In a typical experiment, listeners are presented with lists of words during an encoding phase that contain either indexical variability (e.g., words are spoken by different talkers) or allophonic variability (e.g., stop consonants are either articulated carefully or as a flap). Participants are subsequently asked to perform a perceptual task such as shadowing, recognition, or lexical decision. The key manipulation in these studies was that one-half of the stimuli in the perceptual task were presented exactly as they were in the encoding phase (e.g., if Talker A produced a given item in the encoding phase then that same stimulus was repeated by the same talker in the perceptual task), and one-half were presented in a different form (e.g., if participants heard a flapped stop consonant in the encoding phase then they were presented with a carefully articulated stop in the perceptual task). Specificity of priming was indicated if there was an advantage for identical repetitions (e.g., same voice or same type of articulation) compared with changed repetitions (e.g., different voices or different types of articulation) in the perceptual task. Luce *et al.* (2003) suggested that because listeners frequently encounter allophonic variation, they should process this information more rapidly than indexical variation, which is encountered less frequently. Furthermore, Luce *et al.* (2003) proposed that this difference in the time course of processing should encourage the use of abstract representations for allophonic variability and episodic representations for indexical variability. Behaviorally, this distinction would be manifested in little or no specificity of long-term priming for allophonic variability (e.g., no difference between same- and different-articulation modes), but significant specificity for indexical variability (e.g., poorer performance for different-voice repetitions compared with same-voice repetitions). In a series of elegant studies (see Luce *et al.*, 2003), not only was the hypothesis about specificity of priming for allophonic versus indexical variability confirmed, but both task- and stimulus-related factors that slowed lexical processing (e.g., degrading stimuli with noise) were found to promote specificity in long-term priming, whereas factors that increased the speed of lexical processing (e.g., using a shadowing task rather than a lexical decision task) generally produced little or no specificity in long-term priming. Taken together, these findings suggest that abstract representations may dominate early in spoken word recognition, with episodic representations dominating later in the time course of speech perception.

This dual-code hypothesis proposed by Luce *et al.* (2003) provides a potentially useful framework for understanding both the current and previous results on the effects of stimulus variability. Specifically, variability in overall amplitude and F0 are the two sources of variability that have little or no effect on spoken word identification (Sommers *et al.*, 1994) and, similar to allophonic variation, these are dimensions that listeners are likely to encounter frequently during spoken word recognition. In contrast, variations in speaking style, speaking rate, and talker characteristics may be encountered less frequently, and these are the sources of variability that have been shown to significantly affect spoken word recognition (Sommers *et al.*, 1994; Nygaard *et al.*, 1995). Thus, differences in the effects of variability may reflect differences in the speed of lexical access; variations in overall amplitude and F0 may be processed more quickly and therefore may be more dependent on abstract representations, whereas variations in talker characteristics, speaking rate, and speaking style, which are less frequently encountered, may be processed more slowly and therefore may allow time for episodic representations to dominate when these features are introduced as sources of variability. As a direct test of this time-course hypothesis, experiments are currently underway in our laboratory to examine the relationship between latency of lexical access and the effects of acoustic-phonetic variability.

One additional finding from the current study that supports differential processing demands for variations in F0 and speaking style is the significant interaction with lexical difficulty that was observed for the latter but not for the former. That is, variations in speaking style had significantly greater effects for hard than for easy words, but this interaction was not significant for variations in F0. Similarly, Sommers (1997) reported significant interactions between talker variability and lexical difficulty, with the effects of variability significantly greater for more phonetically confusable words. Although additional research is clearly needed before making any strong statements about the basis for the differential effects obtained with different types of acoustic-phonetic variability, both the current findings and results from previous studies provide strong boundary conditions for any comprehensive account of stimulus variability and its effects on spoken word recognition.

ACKNOWLEDGMENTS

This work was supported by a Faculty Development Grant from Washington University. The authors would like to thank the Associate Editor and three reviewers for their comments which significantly improved the manuscript. We would also like to acknowledge Amy Katz for her help with data collection and analysis.

[1]The term "speaking style" used here refers to the same manipulation that has been referred to as "voice type" in previous work on acoustic variability and second language vocabulary learning (e.g., Barcroft and Sommers, 2005). We believe the term "speaking style" captures well the idea that both source and filter characteristics of the speech signal are affected by variations of this nature.

[2]When instructing speakers to produce a denasalized speaking style and in previous work on acoustic variability and second language vocabulary learning (e.g., Barcroft and Sommers, 2005), we have used the term "nasal" in its common, less specialized, usage to refer to speech produced with the nasal cavity occluded (e.g., participants were asked to produce items as when speaking with a cold). Although based on articulatory considerations this type of speech is actually denasalized, we previously have retained the term nasal for two reasons. First, in generating stimuli speakers were instructed to "speak in a nasal voice." Second, in evaluating the distinctiveness of different speaking styles (see pilot studies), participants were asked to rate "how nasal the voice sounded." We elected to use the term nasal in both of these instances because this is how the desired speaking style is commonly referred to and was easily understood by the participants in both cases.

[3]The easy words "gang," "sauce," and "joke" and the hard words "gill," "soak," and "jeer" were not included in the present study.

[4]We did not include the quality ratings task for the undergraduates because initial testing indicated poor test-retest reliability of the ratings (test-retest reliability for the undergraduates was approximately 0.27, compared with 0.88 for the speech-language pathologists).

[5]The 10% difference was selected because it represented the smallest difference possible using the current stimuli.

[6]We would like to acknowledge and thank the Associate Editor and several of the Reviewers for suggesting this alternative explanation.

Barcroft, J., and Sommers, M. S. (**2005**). "Effects of acoustic variability on second language vocabulary learning," Stud. Second Language Acquisition **27**, 387–414.

Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **9/10**, 341–345.

Bradlow, A. R., Nygaard, L. C., and Pisoni, D. B. (**1999**). "Effects of talker, rate, and amplitude variation on recognition memory for spoken words," Percept. Psychophys. **61**, 206–219.

Cutler, A., and Darwin, C. J. (**1981**). "Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency," Percept. Psychophys. **29**, 217–224.

Dorman, M. F., and Dougherty, K. (**1981**). "Shifts in phonetic identification with changes in signal presentation level," J. Acoust. Soc. Am. **69**, 1439–1440.

Goldinger, S. D. (**1996**). "Words and voices: Episodic traces in spoken word identification and recognition memory," J. Exp. Psychol. Learn. Mem. Cogn. **22**, 1166–1183.

Kucera, F., and Francis, W. (**1967**). *Computational Analysis of Present Day American English* (Brown University Press, Providence, RI).

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.

Luce, P. A., McLennan, C., and Charles-Luce, J. (**2003**). "Abstractness and specificity in spoken word recognition: Indexical and allophonic variability in long-term repetition priming," in *Rethinking Implicit Memory*, edited by J. S. M. Bowers and J. Chad (Oxford University Press, London), pp. 197–214.

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (**1989**). "Some effects of talker variability on spoken word recognition," J. Acoust. Soc. Am. **85**, 365–378.

Mullennix, J. W., and Pisoni, D. B. (**1990**). "Stimulus variability and processing dependencies in speech perception," Percept. Psychophys. **47**, 379–390.

Newman, R. S., Clouse, S. A., and Burnham, J. L. (**2001**). "The perceptual consequences of within-talker variability in fricative production," J. Acoust. Soc. Am. **109**, 1181–1196.

Norris, D., McQueen, J. M., and Cutler, A. (**1995**). "Competition and segmentation in spoken-word recognition," J. Exp. Psychol. Learn. Mem. Cogn. **21**, 1209–1228.

Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," *Res. Speech Percpt., Progress Rep. No. 10 (Indiana University, Bloomington, IN).*

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1995**). "Effects of stimulus variability on perception and representation of spoken words in memory," Percept. Psychophys. **57**, 989–1001.

Pisoni, D. B. (**1993**). "Long-term memory in speech perception: Some new findings on talker variability, speaking rate, and perceptual learning," Speech Commun. **13**, 109–125.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and Slowiaczek, L. M. (**1985**). "Speech perception, word recognition and the structure of the lexicon," Speech Commun. **4**, 75–95.

Sheffert, S. M. (**1998**). "Voice-specificity effects on auditory word priming," Mem. Cognit. **26**, 591–598.
Sommers, M. S. (**1996**). "The structural organization of the mental lexicon and its contribution to age-related changes in spoken word recognition," Psychol. Aging **11**, 333–341.
Sommers, M. S. (**1997**). "Stimulus variability and spoken word recognition. II. The effects of age and hearing impairment," J. Acoust. Soc. Am. **101**, 2278–2288.
Sommers, M. S. (**1998**). "Spoken word recognition in individuals with dementia of the Alzheimer's type: Changes in talker normalization and lexical discrimination," Psychol. Aging **13**, 631–646.
Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1994**). "Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude," J. Acoust. Soc. Am. **96**, 1314–1324.
Sommers, M. S., and Danielson, S. M. (**1999**). "Inhibitory processes and spoken word recognition in young and older adults: The interaction of lexical competition and semantic context," Psychol. Aging **14**, 458–472.
Uchanski, R. M., and Braida, L. D. (**1998**). "Effects of token variability on our ability to distinguish between vowels," Percept. Psychophys. **60**, 533–543.

Effects of introducing unprocessed low-frequency information on the reception of envelope-vocoder processed speech

Michael K. Qin[a] and Andrew J. Oxenham[b]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Harvard-MIT Division of Health Sciences and Technology, Speech and Hearing Bioscience and Technology Program, Cambridge, Massachusetts 02139

(Received 9 April 2005; revised 26 January 2006; accepted 29 January 2006)

This study investigated the benefits of adding unprocessed low-frequency information to acoustic simulations of cochlear-implant processing in normal-hearing listeners. Implant processing was simulated using an eight-channel noise-excited envelope vocoder, and low-frequency information was added by replacing the lower frequency channels of the processor with a low-pass-filtered version of the original stimulus. Experiment 1 measured sentence-level speech reception as a function of target-to-masker ratio, with either steady-state speech-shaped noise or single-talker maskers. Experiment 2 measured listeners' ability to identify two vowels presented simultaneously, as a function of the *F*0 difference between the two vowels. In both experiments low-frequency information was added below either 300 or 600 Hz. The introduction of the additional low-frequency information led to substantial and significant improvements in performance in both experiments, with a greater improvement observed for the higher (600 Hz) than for the lower (300 Hz) cutoff frequency. However, performance never equaled performance in the unprocessed conditions. The results confirm other recent demonstrations that added low-frequency information can provide significant benefits in intelligibility, which may at least in part be attributed to improvements in *F*0 representation. The findings provide further support for efforts to make use of residual acoustic hearing in cochlear-implant users.
[DOI: 10.1121/1.2178719]

PACS number(s): 43.71.Bp, 43.71.Es, 43.71.Ky [PFA] Pages: 2417–2426

I. INTRODUCTION

Despite many significant advances made in the development of cochlear implants (e.g., Dorman, 2000; Zeng, 2004), even the most successful cochlear-implant users do not hear as well as normal-hearing listeners, particularly in noise. Fu *et al.* (1998) and Friesen *et al.* (2001) reported that implant users require higher target-to-masker ratios in broadband noise to achieve levels of speech reception performance comparable to normal-hearing listeners. More recent findings suggest that the differences in performance between normal-hearing listeners and implant users (real and simulated) are especially pronounced for speech in the presence of more complex, fluctuating backgrounds, such as modulated noise or single-talker interference (Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004).

Part of the difficulty experienced by cochlear-implant users in fluctuating backgrounds may reflect an impaired ability to perceptually segregate the fluctuations of the target from those of the masker. Voice pitch, or the fundamental frequency (*F*0) of voicing, has long been thought to be an important cue in the perceptual segregation of speech sources (e.g., Bregman, 1990; Darwin and Carlyon, 1995). While, in principle, voice pitch information is available to implant users via envelope modulations (McKay *et al.*, 1994; Wilson, 1997; Geurts and Wouters, 2001; Green *et al.*, 2002; Moore, 2003), the pitch salience associated with envelope periodicity is known to be less robust than that associated with resolved lower-order harmonics in normal-hearing listeners (e.g., Burns and Viemeister, 1976; 1981; Shackleton and Carlyon, 1994; Kaernbach and Bering, 2001; Bernstein and Oxenham, 2003). Qin and Oxenham (2003) suggested that the difference in speech reception performance between (real and simulated) implant users and normal-hearing listeners in fluctuating maskers can be attributed in part to the loss of pitch as a segregation cue on the part of the implant users. They tested this idea more directly in a later study (Qin and Oxenham, 2005) and found that, under envelope-vocoder processing, large *F*0 differences, which were readily discriminated in a sequential task, could not be used to improve the recognition of two simultaneously presented vowels. A possible explanation for this deficit may be that the auditory system cannot use envelope-periodicity cues to extract the *F*0's of two concurrent sounds, and instead (in normal hearing) relies on low-order resolved harmonics as the primary segregation cue (Carlyon, 1996; Deeks and Carlyon, 2004). Listeners' lack of ability to use envelope-based pitch cues to segregate sources may therefore account for at least part of the difficulties encountered by cochlear-implant users in complex (fluctuating) backgrounds.

A relatively recent trend is to consider patients with some low-frequency residual hearing as candidates for cochlear implantation. Some of these individuals are able to

[a]Electronic mail: qin@nsmrl.navy.mil
[b]Current address: Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. Electronic mail: oxenham@umn.edu

use an implant in one ear while using an acoustic hearing aid in the other ear to form a hybrid electric-plus-acoustic system (EAS). While the residual acoustic hearing present in these implant users is unlikely to contribute directly to speech intelligibility, the additional low-frequency temporal fine-structure cue in acoustic hearing may provide sufficient information to aid in source segregation. In fact, a recent study by Kong *et al.* (2005) has shown that speech reception in the presence of interference improved with combined electric and acoustic hearing compared to either alone.

In another variant of an EAS application, short-insertion cochlear implants have been developed, which stimulate only the basal end of the cochlea, leaving the apical (low-frequency) end sufficiently intact to still preserve low-frequency residual hearing (von Ilberg *et al.*, 1999; Gantz and Turner, 2003). A study by Turner *et al.* (2004) showed that three patients with short-insertion cochlear implants and amplified residual hearing performed better than any of their patients with traditional cochlear implants for speech recognition in a background of competing speech, but not in steady noise. The studies of Turner *et al.* (2004) and Kong *et al.* (2005) suggest that residual acoustic hearing can lead to substantial improvements in performance over cochlear-implant stimulation alone.

In addition to testing implant patients, Turner *et al.* (2004) also tested normal-hearing listeners using a noise-excited envelope vocoder (EV) (e.g., Shannon *et al.*, 1995), which is a type of channel vocoder (Dudley, 1939) designed to simulate aspects of cochlear-implant processing. They investigated the effect of introducing unprocessed low-frequency information on listeners' ability to recognize two-syllable words (spondees) in a background of either steady-state noise or two-talker babble. Using a 16-channel EV and steady-state noise interference, they found no effect of processing: EV processing did not degrade performance, relative to the unprocessed condition, and adding back unprocessed acoustic information below 500 Hz did not improve performance. In contrast, performance in two-talker babble was degraded by 16-channel EV processing, and was partially restored by reintroducing the unprocessed signal below 500 Hz.

These results reinforce the idea that low-frequency fine-structure information might be particularly useful in complex, fluctuating backgrounds, where perceptual segregation is necessary (e.g., Qin and Oxenham, 2003). However, some questions remain. First, how much low-frequency information is required to observe an improvement in speech reception? Turner *et al.* (2004) used a cutoff frequency of 500 Hz, which provided a good approximation to the residual hearing of their implant patients. Information below 500 Hz provides *F*0 information, but also provides information about the first formant of many vowels. Is a similar benefit observed if the additional information is limited to even lower frequencies? Second, does the additional low-frequency information aid performance in situations where dynamic fluctuations and dynamic grouping cues are not present?

To address the two questions above, the current study used two low-pass cutoff frequencies (300 and 600 Hz) to simulate varying extents of residual hearing, and to probe the limits of improvement due to additional low-frequency information. Experiment 1 measured speech-reception accuracy for sentences as a function of target-to-masker ratio in the presence of speech-shaped steady-state noise or a competing single talker. Experiment 2 measured vowel identification for synthetic vowels presented either alone or simultaneously in pairs, as a function of the *F*0 difference between the two vowels. In both experiments the processing conditions included traditional EV processing and two conditions in which the lowest two or three EV frequency channels were replaced by a low-pass-filtered version of the original stimulus.

II. EXPERIMENT 1: SPEECH RECEPTION IN THE PRESENCE OF INTERFERENCE

A. Methods

1. Participants

Eighteen normal-hearing (audiometric thresholds between 125 and 8000 Hz were $<$20 dB HL) native speakers of American English participated in this study. Their ages ranged from 18 to 28.

2. Stimuli

All stimuli in this study were composed of a target sentence presented in the presence of a masker. The stimulus tokens were processed prior to each experiment. The targets were H.I.N.T. sentences (Nilsson *et al.*, 1994) spoken by a male talker. The maskers were either a different male competing talker or speech-shaped noise. The targets and maskers were combined at various target-to-masker ratios (TMRs) prior to any processing. TMRs were computed based on the token-length root-mean-square amplitudes of the signals. The maskers began and ended 500 ms before and after the target, respectively. They were gated on and off with 250-ms raised-cosine ramps.

The H.I.N.T. sentence corpus consists of 260 phonetically balanced high-context sentences of low-to-moderate difficulty. Each sentence contains four to seven keywords. The competing-talker maskers were excerpts from the audio book "Timeline" (novel by M. Crichton) read by Stephen Lang (as used in Qin and Oxenham, 2003). The competing-talker masker had a mean *F*0 (mean=111.4 Hz, s.d. =27.06 Hz) similar to that of the target talker (mean =110.8 Hz, s.d.=24.15 Hz), as estimated by the YIN program (de Cheveigné and Kawahara, 2002). To avoid long silent intervals in the masking speech (e.g., sentence-level pauses), the competing-talker maskers were automatically preprocessed to remove silent intervals longer than 100 ms. The competing-talker maskers and speech-shaped noise maskers were spectrally shaped to match the long-term power spectrum of the H.I.N.T. sentences. The maskers were then subdivided into nonoverlapping segments to be presented in each trial.

3. Stimulus processing

The experimental stimuli for each listener were digitally generated, processed, and stored on disk prior to each experi-

TABLE I. Filter cutoffs for the noise-excited vocoders (3 dB down points).

	EV_{1-8}		$LF_{300}+EV_{3-8}$		$LF_{600}+EV_{4-8}$	
Channel number	Low (kHz)	High (kHz)	Low (kHz)	High (kHz)	Low (kHz)	High (kHz)
1	0.080	0.221	Unprocessed (0.080–0.300)		Unprocessed (0.080–0.600)	
2	0.221	0.426				
3	0.426	0.724	0.426	0.724		
4	0.724	1.158	0.724	1.158	0.724	1.158
5	1.158	1.790	1.158	1.790	1.158	1.790
6	1.790	2.710	1.790	2.710	1.790	2.710
7	2.710	4.050	2.710	4.050	2.710	4.050
8	4.050	6.000	4.050	6.000	4.050	6.000

ment. Stimulus processing was performed using MATLAB (Mathworks, Natick, MA) in the following manner.

The experimental stimuli were presented in four processing conditions. In all conditions, the target level was fixed at 65 dB SPL and the masker levels were varied to meet the desired TMR. In the first processing condition (unprocessed), the stimuli were filtered between 80 Hz and 6 kHz, but were otherwise left unchanged. The second processing condition (EV_{1-8}), designed to simulate the effects of envelope-vocoder implant processing, used an 8-channel noise-excited vocoder (Qin and Oxenham, 2005). The input stimulus was first bandpass filtered (sixth-order Butterworth filters) into eight contiguous frequency channels between 80 and 6000 Hz (see Table I). The entire frequency range was divided equally in terms of the Cam scale (Glasberg and Moore, 1990; Hartmann, 1997). The envelopes of the signals were extracted by half-wave rectification and low-pass filtering (using a second-order Butterworth filter) at 300 Hz, or half the bandpass filter bandwidth, whichever was lower. The 300-Hz cutoff frequency was chosen to preserve as far as possible the *F*0 cues in the envelope. The envelopes were then used to amplitude modulate independent white-noise carriers. The same bandpass filters that were used to filter the original stimuli were then used to filter the amplitude-modulated noises. Finally, the modulated narrow-band noises were summed and scaled to have the same level as the original stimuli.

The last two processing conditions were designed to simulate "electric plus acoustic" systems (EAS). The unprocessed stimuli were low-pass filtered using a sixth-order Butterworth filter with 3-dB down points at 300 and 600 Hz (LF_{300} and LF_{600} conditions). To simulate the effects of EAS with residual hearing below 300 Hz ($LF_{300}+EV_{3-8}$), LF_{300} was paired together with EV_{3-8}, consisting of the upper six channels of the eight-channel vocoder. The first formant frequency of most vowels exceeds 300 Hz (Hillenbrand *et al.*, 1995), so by limiting information to below that frequency, most direct speech information should be eliminated. To simulate the effects of an EAS with residual hearing below 600 Hz ($LF_{600}+EV_{4-8}$), LF_{600} was paired with EV_{4-8}, consisting of the upper five channels of the eight-channel vocoder simulation. The 600-Hz cutoff was selected to incorporate the first formant frequency of most vowels, and to include the most dominant harmonics in the formation of pitch percepts (Moore *et al.*, 1985; Dai, 2000). The low-pass filter cutoff frequencies were somewhat below the lower cutoff frequencies of the EV, leaving small (roughly 1 critical band) spectral gaps between the two (unprocessed and processed) regions (see Table I). This was done to limit the low-frequency information to the desired region, while maintaining the same channel configuration that was used in the full EV_{1-8} condition.

4. Procedure

The 18 participants were divided into two groups of nine. The speech reception of each group was measured under only one of the masker types (competing talker or speech-shaped noise). The experiment involved the participants listening to sentences in a background masker and entering as much as they could of the target sentence via a computer keyboard. No feedback was given.

Prior to the experiment session, listeners were given practice performing the experimental tasks as well as given exposure to the processed stimuli. The target sentences used in the training came from the IEEE corpus (IEEE, 1969), whereas the target sentences in the experiment sessions came from the H.I.N.T. corpus. While the maskers in the training and experiment session came from the same corpus, care was taken to ensure that the same masker token was never repeated. During the training session, the listener was exposed to a total of 35 stimulus tokens (five lists, with seven sentences per list), in each of the four processing conditions. In each processing condition, the target sentences were presented at a TMR in the midrange of the experimental TMRs (see Table II). The listeners were instructed to enter their responses via a computer keyboard, as in the experiment. No feedback was given.

In the experiment session, speech reception was measured for each listener under all four processing conditions (unprocessed, EV_{1-8}, $LF_{300}+EV_{3-8}$, $LF_{600}\_EV_{4-8}$), at 5 TMRs (see Table II), in the presence of one masker type. The TMRs for each processing condition and masker type were determined in an earlier pilot study, using two to three listeners. The TMRs were chosen to minimize floor and ceiling effects in the psychometric function. For each listener and condition, the target sentence lists were chosen at random, without replacement, from among the 25 lists of H.I.N.T.

TABLE II. List of conditions tested.

Processing condition	Masker type	Target-to-masker ratio (dB)
Unprocessed	Competing talker	[−20,−15,−10,−5,0]
	Steady-state noise	[−10,−7,−5,−3,0]
NEV_{1-8}	Competing talker	[−5,0,5,10,15]
	Steady-state noise	[−5,−1,2,6,10]
$LPF_{300}+NEV_{3-8}$	Competing talker	[−10,−5,0,5,10]
	Steady-state noise	[−5,−1,2,6,10]
$LPF_{600}+NEV_{4-8}$	Competing talker	[−10,−6,−3,1,5]
	Steady-state noise	[−10,−6,−3,1,5]

sentences. This was done to ensure that no target sentence was presented more than once to any given listener. Data were collected using one list (i.e., ten sentences) for each TMR in each condition. For each listener, the different conditions were run in random order.

The stimuli were converted to the analog domain using a soundcard (LynxStudio, LynxOne) at 16-bit resolution with a sampling rate of 22 050 Hz. They were then passed through a headphone buffer (TDT HB6) and presented diotically via Sennheiser HD580 headphones to the listener seated in a double-walled sound-attenuating booth.

Listener responses were scored offline by the experimenter. Obvious misspellings of the correct word were considered correct. Each listener's responses to the ten sentences at a given TMR, under a given masker condition, were grouped together to produce a percent-correct score that was based on the number of correct keywords.

B. Results and discussion

1. Fits to the psychometric functions

The percent-correct scores as a function of TMR under a given masker condition for each listener were fitted to a two-parameter sigmoid model (a cumulative Gaussian function),

$$\text{Percent correct} = \frac{100}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\text{TMR}} \exp\left(\frac{-(x-\text{SRT})^2}{2\sigma^2}\right)dx, \tag{4.1}$$

where x is the integration variable, SRT is the speech reception threshold in dB at which 50% of words were correctly identified, σ is related to the slope of the function, and TMR is the target-to-masker ratio (dB). The two-parameter model assumes that listeners' peak reception performance is 100%. The two-parameter model provided generally good fits to the data. The individual standard-errors-of-fit had a mean of 2.65% with a standard deviation of 2.52% (median of 1.87% and a worst case of 12.93%).

2. Speech-reception thresholds (SRT)

The mean SRT values and standard errors of means were derived from the SRT values of individual model fits. In general, performance across listeners was consistent, so only the mean SRT values as a function of masker condition and processing condition are plotted in Fig. 1. A higher SRT value implies a condition more detrimental to speech reception. Figure 1 shows that in the unprocessed conditions (right-most bars), the steady-state noise masker was a more effective masker than the competing-talker masker. However, with eight-channel envelope-vocoder processing (EV_{1-8}) the reverse was true (i.e., competing talker was the more effective masker), consistent with the findings of Qin and Oxenham (2003). When unprocessed low-frequency information (LF) was added to the envelope-vocoder processed speech, the SRT associated with both maskers decreased, indicating a speech reception benefit.

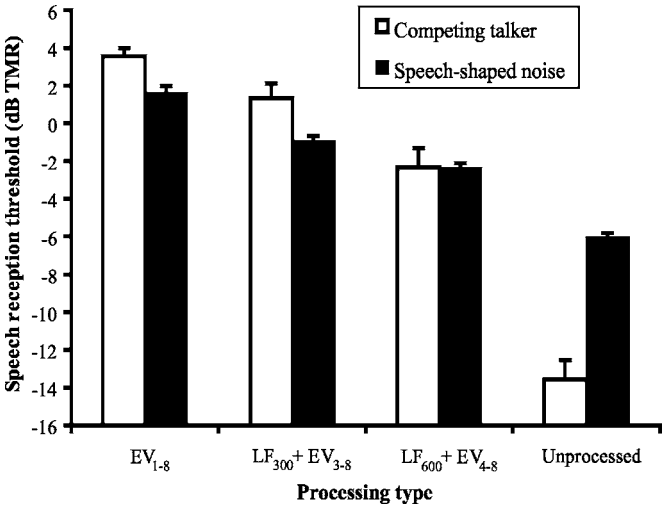


FIG. 1. Group mean speech reception threshold (SRT) values for the two types of background interference. The error bars denote ±1 standard error of the mean.

A two-way mixed-design analysis of variance (ANOVA) was performed on the data for the EV_{1-8}, $LF_{300}+EV_{3-8}$, and $LF_{600}+EV_{4-8}$ conditions and for both noise and single-talker maskers. The statistical significance of the findings was determined with SRT as the dependent variable, masker type as the between-subjects factor, and processing condition as the within-subjects factor. The ANOVA showed significant main effects of both processing condition ($F_{2,32}=57.16$, $p<0.05$) and masker type ($F_{1,16}=5.18$, $p<0.05$). There was also a significant interaction between masker type and processing condition ($F_{2,32}=3.49$, $p<0.05$). This may reflect the fact that adding low-frequency information seems to have a somewhat greater effect for the single-talker masker than for the steady-state noise masker. In particular, the SRT difference between the competing-talker and steady-state maskers, which was present in the EV_{1-8} condition, was not found in the $LF_{600}+EV_{4-8}$ condition. Note, however, that even with the 600-Hz low-pass information, listeners were still not able to benefit from the single-talker masker fluctuations as they were in the unprocessed condition.

Fisher's *post hoc* least-significance difference (LSD) test ($\alpha=0.05$) revealed some differences between individual conditions. In the presence of speech-shaped noise, SRTs were higher in the EV_{1-8} condition than in the $LF_{300}+EV_{3-8}$ condition, which in turn was higher than in the $LF_{600}+EV_{4-8}$ condition. Thus, increasing the amount of low-frequency information produced a systematic improvement in performance. Similar improvements were observed with the single-talker interferer, although the difference in SRT between the

EV_{1-8} condition and the $LF_{300}+EV_{3-8}$ condition failed to reach significance in the *post hoc* tests.

The larger effect of adding low-frequency information below 600 Hz for the competing talker than for the steady-state noise is consistent with the results of Turner *et al.* (2004). However, in contrast to their findings, we also observed a substantial improvement in the steady-state noise condition. This difference may be due to a number of factors. Most importantly, performance in their 16-channel EV condition was the same as in the unprocessed condition, leaving no room for improvement by reintroducing low-frequency information. In contrast, we used an eight-channel EV, which more accurately simulates the abilities of the best current implant users (Friesen *et al.*, 2001), and which also provided a substantial decrease in performance relative to the unprocessed condition. Also, their closed-set 12-alternative spondee identification task may have required fewer cues to perform correctly than our sentence task, leading to high performance in the EV condition. In any case, our results suggest that low-frequency residual hearing may be beneficial in a variety of backgrounds, including steady-state noise.

In summary, the addition of unprocessed low-frequency information improved performance in speech-shaped noise at both cutoff frequencies (300 and 600 Hz). The $LF_{600}+EV_{4-8}$ condition performance also produced a significant improvement compared to the EV-alone condition in the presence of competing talkers. Adding unprocessed information below 600 Hz to envelope-vocoded high-frequency information improved speech reception threshold by about 6 dB in the presence of a competing talker and by about 4 dB in the presence of speech-shaped noise compared to the EV-only condition. Although the addition of low-frequency information did not return speech reception performance to normal levels, it was nevertheless a significant improvement when compared with envelope-vocoder alone.

III. EXPERIMENT 2: CONCURRENT-VOWEL IDENTIFICATION

A. Rationale

Experiment 1 showed that low-frequency information can improve performance in a sentence recognition task under both steady-state noise and single-talker interference. The cues that are provided by the low-frequency information are primarily *F*0 and, particularly with the 600-Hz cutoff, first-formant information. Particularly in the case of the single-talker interference, amplitude fluctuations in the target and the interferer lead to many time instances where one or the other dominate. The addition of coherently fluctuating low-frequency *F*0 information may help listeners identify these instances and thus improve performance. Experiment 2 used a concurrent-vowel paradigm (Scheffers, 1983), which rules out these time-varying aspects and their accompanying perceptual segregation cues, such as onset asynchronies and different dynamic frequency and amplitude trajectories. The concurrent-vowel paradigm, despite its artificial nature, also eliminates possible semantic and grammatical context cues that might have influenced performance in our sentence recognition task. It has been a popular task to probe the role of *F*0, and other potential segregation cues (such as interaural time differences) in separating sounds in normal (Brokx and Nooteboom, 1982; Assmann and Summerfield, 1990; Summerfield and Assmann, 1991; Culling and Darwin, 1993; Assmann and Summerfield, 1994; Culling and Darwin, 1994; Darwin and Carlyon, 1995; de Cheveigné *et al.*, 1997; Bird and Darwin, 1998) and impaired (Arehart *et al.*, 1997) hearing.

TABLE III. Formant frequencies (Hz) for vowels. Values enclosed in parentheses represent formant bandwidths (Hz).

	Formant			
Vowel	*F*1 (60)	*F*2 (90)	*F*3 (150)	*F*4 (200)
/i/	342	2322	3000	3657
/A/	768	1333	2522	3687
/u/	378	997	2343	3357
/ɛ/	580	1799	2605	3677
/ɝ/	474	1379	1710	3334

By using concurrent vowels that differed only in *F*0, we were able to assess the extent to which reintroducing low-frequency acoustic information aided listeners in their ability to use *F*0 differences to segregate two simultaneous sounds. As in experiment 1, two levels of residual hearing (<300 Hz and <600 Hz) were tested.

B. Methods

1. Participants

Six native speakers of American English (audiometric thresholds between 125 and 8000 Hz were <20 dB HL) were paid for their participation in this experiment. Their ages ranged from 19 to 26.

2. Stimuli

Five American English vowels (/i/ as in *heed*, /A/ as in *hod*, /u/ as in *hood*, /ɛ/ as in *head*, /ɝ/ as in *herd*) were synthesized using an implementation of Klatt's cascade synthesizer (Klatt, 1980). They were generated at a sampling frequency of 20 kHz, with 16-bit quantization. The formant frequencies (see Table III) used to synthesize the vowels were based on the estimates of Hillenbrand *et al.* (1995) for an average male talker. The vowels were chosen because of their positions in the $F1-F2$ space; because their natural duration characteristics (House, 1960; 1961) are similar to the stimulus durations used in this experiment (i.e., 200 ms); and because they can all be reasonably approximated by time-invariant formant patterns. Each vowel was generated with seven different *F*0's ranging from 0 to 14 semitones above 100 Hz (100, 105.9, 112.2, 126.0, 158.7, 200.0, and 224.5 Hz).

The concurrent-vowel pairs were constructed by summing two single vowels with equal rms levels, with their onsets and offsets aligned, and with their pitch periods in phase at the onset of the stimulus. No vowel was paired with itself to generate the concurrent-vowel pairs. Each concurrent-vowel token was constructed using one vowel

with an $F0$ of 100 Hz and the other with an $F0$ of 100 Hz $+\Delta F0$, where the $\Delta F0$ ranged from 0 to 14 semitones. This yielded a total of 140 concurrent-vowel stimuli (20 vowel pairs $\times 7$ $\Delta F0$'s). Each stimulus had a total duration of 200 ms and was gated on and off with 25-ms raised-cosine ramps. The stimuli were presented at an overall rms level of 70 dB SPL.

3. Stimulus processing

All stimulus tokens were digitally generated, processed, and stored on computer disk prior to the experiments. The experimental stimuli were presented in seven conditions (unprocessed, LF_{300}, LF_{600}, EV_{3-8}, EV_{4-8}, $LF_{300}+EV_{3-8}$, and $LF_{600}+EV_{4-8}$). The additional low-frequency-only conditions (LF_{300} and LF_{600}) were added to test how much information could be gleaned from those frequencies in isolation. Given that the frequency of the first formant was always above 300 Hz but mostly below 600 Hz, it could be predicted that no identification would be possible in the LF_{300} condition, but that some degree of performance might be possible in the LF_{600} condition.

4. Procedure

Performance in single-vowel and concurrent-vowel identification tasks was measured using a forced-choice paradigm. Listeners were instructed to identify the vowels heard by selecting visual icons associated with the vowels. In the single-vowel identification task, listeners were instructed to identify the vowel heard by selecting from five different choices. In the concurrent-vowel identification task, listeners were instructed to identify both of the constituent vowels, and the response was only marked correct if both vowels were correctly identified. The responses were entered via a computer keyboard and mouse inside the booth. No feedback was provided.

Each listener took part in six 2-h sessions. Three sessions incorporated the unprocessed, LF_{300}, EV_{3-8}, and $LF_{300}+EV_{3-8}$ conditions, and the three other sessions incorporated the unprocessed, LF_{600}, EV_{4-8}, and $LF_{600}+EV_{4-8}$ conditions. The 300-Hz and 600-Hz sessions were interleaved, with the order randomized across subjects.

Each experiment session was subdivided into eight blocks, with each block involving a single condition (unprocessed, LF, EV, and LF+EV). The first four blocks measured single-vowel identification and the next four blocks measured concurrent-vowel identification. The order of the blocks was randomized from session to session.

Within a given block, the stimulus tokens were presented in random order. Within each single-vowel identification block, a total of 70 stimulus tokens was presented (7 $F0$'s $\times 5$ vowels $\times 2$ repetitions). For each listener, this translates to a total of 30 trials (5 vowels $\times 2$ repetitions $\times 3$ sessions) at each $F0$ under each processing condition. Within each concurrent-vowel identification block, a total of 140 stimulus tokens was presented (7 $\Delta F0$'s $\times 20$ vowel pairs). For each listener, this translates to a total of 60 trials (20 vowel pairs $\times 3$ sessions) at each $F0$ under each processing condition. In the unprocessed condition listeners were exposed to twice as many trials at each $F0$, because the unprocessed condition was presented in each of the six sessions. The six sessions took place over a span of 2–3 weeks, depending on the availability of the participants.

Prior to each experiment session, every listener was given practice performing the experimental tasks. As in the actual experiment, listeners were instructed to enter their responses via the computer keyboard, and no feedback was provided. On average, listeners were exposed to 40–80 stimulus tokens prior to data gathering.

C. Results and discussion

Figure 2 shows the identification accuracy, in percent correct, as a function of the $F0$ in the single-vowel identification task (dotted lines) and as a function of the $F0$ of the upper vowel in the concurrent-vowel identification task (symbols and solid lines), in units of semitones above 100 Hz. The unprocessed conditions are shown in Fig. 2(a), the LF conditions in Fig. 2(b), the EV conditions in Fig. 2(c), and the LF+EV conditions in Fig. 2(d). To investigate trends in the data, repeated-measures ANOVAs were conducted. All scores were arcsine transformed prior to statistical analysis.

In the unprocessed conditions [Fig. 2(a)] a single-factor repeated-measures ANOVA was conducted with $\Delta F0$ as the within-subject factor. The analysis revealed that the effect of $\Delta F0$ was statistically significant ($F_{6,30}=11.87$, $p<0.05$), as illustrated by the average improvement in performance from 78% to 95% as $\Delta F0$ increases from 0 to 2 semitones. Above this $\Delta F0$, performance plateaus until the $\Delta F0$ equals 12 semitones (one octave), consistent with previous studies (Brokx and Nooteboom, 1982; Culling and Darwin, 1993; Qin and Oxenham, 2005). When the $\Delta F0$ equals one octave, the harmonics of the two constituent vowels become inseparable, leading to a drop in identification performance. At $\Delta F0$ of 14 semitones, the identification performance seems to improve somewhat, although the performance difference between 12 and 14 semitones was not statistically significant (Fisher's LSD, $p>0.05$).

In the LF conditions [Fig. 2(b)], identification performance is greatly reduced, as predicted by the articulation index (AI) and the speech intelligibility index (SII) (ANSI, 1997). When a repeated-measures ANOVA with two within-subject factors ($\Delta F0$ and low-pass cutoff) was conducted, a statistically significant difference was found between LF_{300} and LF_{600} ($F_{1,5}=11.03$, $p<0.05$), but there was no statistically significant effect of $\Delta F0$ ($F_{6,30}=1.67$, $p>0.1$) and no interaction ($F_{6,30}<1$, n.s.). In the LF_{300} condition, performance for both single and double vowels was reduced to around chance (20% and 10%, respectively) and no benefits of $F0$ differences were seen. In the LF_{600} condition, although identification of both single and concurrent vowels improved, still no benefit of $F0$ differences was observed.

In the EV conditions [Fig. 2(c)], while single-vowel identification was generally high, concurrent-vowel identification was modest. When a repeated-measures ANOVA with two within-subject factors ($\Delta F0$ and low-pass cutoff) was conducted, no statistically significant difference was found between EV_{3-8} and EV_{4-8} ($F_{1,5}=1.67$, $p>0.1$). In addition,

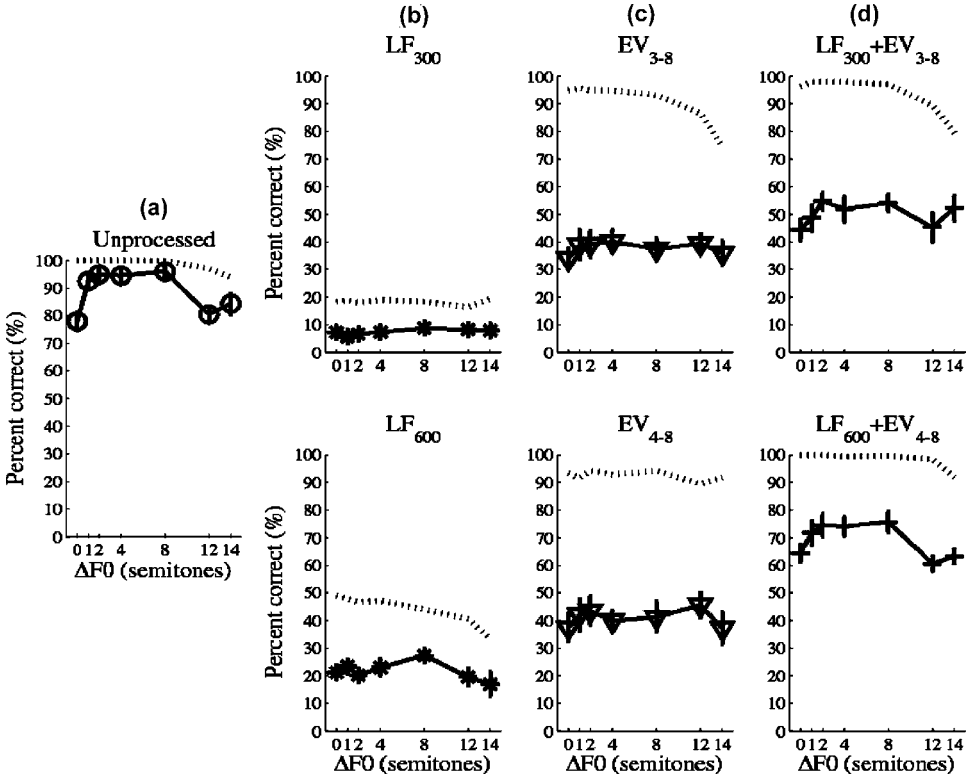


FIG. 2. Dotted lines show the percent of correct responses as a function of the $F0$ in the single-vowel identification task (dotted lines), where $F0$ is described in terms of the number of semitones from 100 Hz. Solid lines show the percent of correct responses as a function of the $\Delta F0$ (in semitones) between constituent vowels in the concurrent-vowel identification task, where the lower $F0$ was always 100 Hz. The error bars denote ±1 standard error of the mean. The unprocessed conditions are shown in the left panel (a), next the low-pass filtered (LF) conditions (b), then the envelope-vocoder (EV) conditions (c), finally the LF+EV conditions in the right panels (d). The top row of figures is associated with the 300-Hz low-pass sessions, and the bottom row of figures is associated with the 600-Hz low-pass sessions.

no effect of $\Delta F0$ ($F_{6,30}=1.67$, $p>0.1$), and no interaction ($F_{6,30}<1$, n.s.), was seen. This suggests that vowel identification performance does not improve as a function of $F0$ difference in either of the EV conditions, consistent with observations from our previous study using both 8- and 24-channel EV processing (Qin and Oxenham, 2005). A direct comparison of the EV_{4-8} condition in this experiment and the EV_{1-8} condition in Qin and Oxenham (2005) suggests only a small improvement due to the addition of the three lowest vocoder channels. In particular, the averaged percent correct for the eight subjects in that study, pooled across $\Delta F0$'s, was 45.3%, compared with 41.0% in the present study.

In the LF+EV conditions [Fig. 2(d)], identification performance improved compared to both LF-only and EV-only conditions, and overall performance in the $LF_{600}+EV_{4-8}$ condition was higher than that in the $LF_{300}+EV_{3-8}$ condition. A repeated-measures ANOVA with two within-subject factors ($\Delta F0$ and low-pass cutoff) showed significant main effects of cutoff ($F_{1,5}=23.71$, $p<0.05$) and $\Delta F0$ ($F_{6,30}=6.73$, $p<0.05$). These effects remained significant when the 12- and 14-semitone conditions were excluded, suggesting that the effect is truly due to the improvement in performance with increasing $\Delta F0$ at the lower $\Delta F0$'s. The interaction between cutoff and $\Delta F0$ was also found to be significant ($F_{6,30}=2.59$, $p<0.05$). However, when the data from the 12- and 14-semitone $\Delta F0$ conditions were excluded from the analysis, no significant interaction remained between cutoff and $\Delta F0$ ($F_{4,20}<1$, n.s.), indicating that the benefits of increasing $\Delta F0$ in the $LF_{300}+EV_{3-8}$ and $LF_{600}+EV_{4-8}$ conditions were not statistically different for moderate $\Delta F0$ between 0 and 8 semitones. Overall, the pattern of results in the $LF_{300}+EV_{3-8}$ and $LF_{600}+EV_{4-8}$ conditions is similar to that found in the unprocessed condition, although overall performance remains lower even in the $LF_{600}+EV_{4-8}$ condition, as was also found in experiment 1.

It has been suggested that part of the $\Delta F0$ benefit observed in concurrent-vowel identification at small values of $\Delta F0$ (1 semitone or less) may arise from "beating" between adjacent components with concurrent vowels (Assmann and Summerfield, 1994; Culling and Darwin, 1994; Darwin and Carlyon, 1995; Bird and Darwin, 1998). At least two types of interaction exist. The first involves two individual low-order harmonics which interact, producing a pattern of amplitude and frequency modulation; the second involves interactions between groups of unresolved harmonics that produce so-called pitch-period asynchronies (e.g., Assmann and Summerfield, 1994). Both types of interaction are present to some extent in our stimuli, with the interaction between individual harmonics present in the LF region, and only the pitch-pulse asynchronies present in the EV region (albeit somewhat masked by the inherent fluctuations of the noise carriers). It is interesting to note that the pitch-pulse asynchronies seem not to be sufficient on their own to improve vowel identification, given that the EV-alone conditions showed no benefit of differences in $F0$ (see also Qin and Oxenham, 2005). In any case, most of our conditions involved $\Delta F0$'s that were greater than those thought to be influenced by slowly varying beats.

In summary, reintroducing low-frequency information led to an overall improvement in the identification of concurrent vowels, and also led to a positive effect of $F0$ difference between the two vowels that was not present in either the LF-alone or the EV-alone conditions.

IV. DISCUSSION

The results from the two experiments can be summarized as follows. In experiment 1, speech reception for sentences in both steady-state noise and single-talker interference improved as (unprocessed) low-frequency information was added to EV simulations of cochlear-implant processing. In experiment 2, the addition of low-frequency information improved overall identification of two concurrently presented vowels, and led to a benefit of *F*0 differences between the two vowels, even when the cutoff frequency was as low as 300 Hz.

Our results are broadly consistent with those of Turner *et al.* (2004) in showing an improvement due to unprocessed low-frequency information. Our findings also extend the previous results in the following ways. First, sentence recognition can improve with additional low-frequency information even in steady-state noise. Second, improvements can be observed with a low-pass cutoff frequency as low as 300 Hz. Third, benefits of low-frequency information are found with steady-state vowel sounds, where dynamic grouping cues, and "glimpsing" of the target, are not available. This provides further support for approaches that attempt to combine acoustic with electric stimulation in cochlear-implant users with some residual hearing (Kong *et al.*, 2004; Turner *et al.*, 2004). However, it is important to note that our simulations involve "ideal" residual hearing, with no hearing loss and accompanying effects, such as broadened auditory filters. These conditions are unlikely to hold in real EAS users. On the other hand, real EAS users will have time to adjust to their processing scheme, and are perhaps more likely to perceptually fuse the two modes of presentation after prolonged exposure. In any case, care should be taken when interpreting the current findings in terms of potential benefits for implant patients. With that caveat in mind, some possible implications of our work can be examined.

A. Frequency extent of residual hearing necessary for tangible benefits in speech reception

Given the potential variability in the amount of residual hearing available across the population of cochlear implant candidates, one aim of the current study was to investigate the frequency extent of residual hearing necessary to show tangible speech reception benefits. Our findings suggest that even an extremely limited range (<300 Hz) of residual hearing may be beneficial to the reception of speech in the presence of interference. Both experiment 1 and 2 showed that when unprocessed information below 300 Hz was added to envelope-vocoder processing, significant improvements in speech identification could be observed. In experiment 1, the SRT decreased by 2.5 dB in steady-state noise. This 2.5-dB improvement in SRT translates to an improvement in intelligibility of about 20 percentage points for the sentence material used in our tests. In experiment 2, concurrent-vowel identification improved beyond that of the vocoder-only condition. In addition, listeners exhibited $\Delta F0$ benefits that were not observed in the vocoder-only conditions. While the addition of low-frequency information did not return speech reception performance to normal levels in either experiment, the improvement was nevertheless significant when compared with envelope-vocoder alone. The current findings, taken together with the positive results from real EAS users (Tyler *et al.*, 2002; Ching *et al.*, 2004; Turner *et al.*, 2004; Kong *et al.*, 2005), lead us to be cautiously optimistic about the ability of combined electric and acoustic stimulation to enhance the perceptual segregation of speech.

B. Role of *F*0 representation in improving performance

As stated in the Introduction, previous researchers (Turner *et al.*, 2004; Kong *et al.*, 2005) have suggested that speech reception benefits of residual hearing may be attributable to an improvement in *F*0 representation. However, the cochlear-implant subjects in their studies had residual hearing up to frequencies as high as 1 kHz. With this level of residual acoustic hearing, speech-reception benefits could be attributed at least in part to increased spectral resolution and more accurate formant frequency information, rather than to improvements in *F*0 representation alone. Even in the acoustic simulations of Turner *et al.* (2004) the cutoff frequency of 500 Hz was probably too high to rule out the improved representation of the first formant in the LF region. The current simulation experiment examined the effect of adding unprocessed information, low-pass filtered at 300 Hz. Because the lower cutoff frequency of the first noise band was at 426 Hz, and because all the stimuli were filtered with sixth-order Butterworth filters, the crossover frequency between the unprocessed low-frequency stimuli and the first noise band was at around 360 Hz, at which point the filters were roughly 10 dB below their peak value. As the lower-frequency noise band would likely mask any information beyond this crossover frequency, it is reasonable to assume that only reduced (and spectrally distorted) unprocessed low-frequency fine-structure information was available beyond 300 Hz, and none was available beyond 360 Hz. This frequency range is thought to contain very little speech information (ANSI, 1997) and, in the case of our vowels, would not have been sufficient to fully represent the spectral peak of any of the first formants in our experiment. This is supported by the fact that our listeners performed at chance in the single-vowel identification task when listening only to the unprocessed stimuli, low-pass filtered at 300 Hz. Thus, it is reasonable to conclude that an improvement in *F*0 representation aided performance for both the vowels and the sentences.

With the higher cutoff frequency of 600 Hz, part of the improvement in performance was likely due to improved spectral representation of the first formant. This is suggested in the results of experiment 2 by the fact that single- and double-vowel identification was no longer at chance in the LF_{600} condition. In contrast, the benefits of increasing $\Delta F0$ in concurrent-vowel identification results were similar for both the 300- and 600-Hz cutoff frequencies, supporting the notion that the addition of unprocessed low-frequency information improved *F*0 representation and thus improved speech segregation abilities.

For most conditions in experiment 2, the *F*0's were sufficiently far apart (2 semitones or more) to provide at least

one or two resolved harmonics for each vowel. In the 1-semitone condition that might not be the case, given that 1 semitone corresponds to 6%, which is smaller than the limit usually ascribed to peripheral resolvability (e.g., Plomp, 1964; Bernstein and Oxenham, 2003). However, depending on the vowel pairing, the amplitude of the fundamental components of concurrent vowels can be significantly different. The amplitude of the component at the *F*0 varies as the frequency of the first formant varies. For example, the amplitude of *F*0 for the vowel /i/ is higher than that of the vowel /A/, because the /i/'s *F*1 (i.e., 342 Hz) is closer to the *F*0 than /A/'s *F*1 (i.e., 768 Hz). The difference in amplitude between fundamental components of vowels may aid in the resolvability of the dominant *F*0 component by rendering the other *F*0 less audible, thereby increasing the ability to segregate the concurrent vowels.

It is currently unclear exactly how *F*0 information from the low-frequency region is used to help perceptual segregation and speech recognition of EV-processed speech sounds. It is possible that the additional *F*0 information aids the grouping of channels in the envelope-vocoder processed region. In experiment 1 with the competing talker, this could be on a moment-by-moment basis, with the *F*0 information cuing which voice is dominant at a given time; with the steady-state masker, it may simply improve pitch contour perception, which has been shown to aid lip-reading (Rosen *et al.*, 1981) and may similarly aid in interpreting EV-processed information. In experiment 2, when two vowels are presented simultaneously at similar intensities, there can still be spectral regions that are dominated by the energy from one of the vowels. An improved representation of *F*0 through low-frequency information may aid the listener in grouping together different spectral regions, perhaps in conjunction with top-down or "template" mechanisms that make use of prior knowledge of vowel structures as well as the *F*0 cues in the envelope that alone are too weak a cue for perceptual segregation.

As mentioned above, the benefit of adding low-frequency information, even when it is low-pass filtered at 300 Hz, is reminiscent of a technique that presented the *F*0 of voiced speech to profoundly hearing-impaired listeners as a potential aid to visual speech cues (Rosen *et al.*, 1981; Faulkner *et al.*, 1992). This technique was originally proposed as a possible alternative to cochlear implantation, but the present results, along with encouraging previous simulation results (Turner *et al.*, 2004) and results in actual implant patients (Turner *et al.*, 2004; Kong *et al.*, 2005), suggest that such a scheme (either via *F*0 tracking, or simply presenting unprocessed low-frequency acoustic information) may be a valuable supplement to cochlear implant patients with some residual hearing.

V. SUMMARY

(1) In a sentence recognition task, adding unprocessed information below 300 or 600 Hz led to significant improvements in speech reception for envelope-vocoder-processed speech in both steady-state speech-shaped noise and single-talker interference.
(2) In a concurrent-vowel identification task, significant benefits of *F*0 differences between the two vowels were observed when low-frequency information was introduced, even when it was over a very limited range (<300 Hz).
(3) The improvements, particularly in the case of the 300-Hz cutoff frequency, can probably be attributed to an improvement in *F*0 representation through the presence of low-order harmonics.
(4) The findings presented here extend those of Turner *et al.* (2004) and Kong *et al.* (2005), providing further support for schemes involving a combination of electric and acoustic stimulation (EAS), and suggest that even very limited residual hearing has the potential to provide substantial speech recognition benefits, when amplified and combined with a cochlear implant.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (NIDCD Grant R01 DC 05216). We thank Joshua Bernstein, Louis Braida, Christophe Micheyl, Barbara Shinn-Cunningham, Andrea Simonson, and the reviewers for helpful comments on earlier versions of this manuscript.

ANSI (**1997**). S3.5-1997 "Methods for calculation of the speech intelligibility index," (American National Standards Institute, New York).
Arehart, K. H., King, C. A., and McLean-Mudgett, K. S. (**1997**). "Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss," J. Speech Lang. Hear. Res. **40**, 1434–1444.
Assmann, P. F., and Summerfield, Q. (**1990**). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," J. Acoust. Soc. Am. **88**, 680–697.
Assmann, P. F., and Summerfield, Q. (**1994**). "The contribution of waveform interactions to the perception of concurrent vowels," J. Acoust. Soc. Am. **95**, 471–484.
Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?" J. Acoust. Soc. Am. **113**, 3323–3334.
Bird, J., and Darwin, C. J. (**1998**). "Effects of a difference in fundamental frequency in separating two sentences," in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 263–269.
Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Bradford Books, MIT Press, Cambridge, MA).
Brokx, J. P. L., and Nooteboom, S. G. (**1982**). "Intonation and the perceptual separation of simultaneous voices," J. Phonetics **10**, 23–36.
Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.
Burns, E. M., and Viemeister, N. F. (**1981**). "Played again SAM: Further observations on the pitch of amplitude-modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.
Carlyon, R. P. (**1996**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.
Ching, T. Y. C., Incerti, P., and Hill, M. (**2004**). "Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears," Ear Hear. **25**, 9–21.
Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by *F*0," J. Acoust. Soc. Am. **93**, 3454–3467.
Culling, J. F., and Darwin, C. J. (**1994**). "Perceptual and computational separation of simultaneous vowels: Cues arising from low-frequency beating," J. Acoust. Soc. Am. **95**, 1559–1569.
Dai, H. P. (**2000**). "On the relative influence of individual harmonics on pitch judgment," J. Acoust. Soc. Am. **107**, 953–959.
Darwin, C. J., and Carlyon, R. P. (**1995**). "Auditory Grouping," in *Hearing*, edited by B. C. J. Moore (Academic, San Diego).
de Cheveigné, A., and Kawahara, H. (**2002**). "YIN, a fundamental frequency

estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930.
de Cheveigné, A., Kawahara, H., Tsuzaki, M., and Aikawa, K. (**1997**). "Concurrent vowel identification. I. Effects of relative amplitude and *F*0 difference," J. Acoust. Soc. Am. **101**, 2839–2847.
Deeks, J. M., and Carlyon, R. P. (**2004**). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," J. Acoust. Soc. Am. **115**, 1736–1746.
Dorman, M. F. (**2000**). "Speech perception by adults," in *Cochlear Implants*, edited by S. Waltzman and N. Cohen (Thieme, New York), pp. 317–329.
Dudley, H. W. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 169–177.
Faulkner, A., Ball, V., Rosen, S., Moore, B. C. J., and Fourcin, A. J. (in press). "Speech pattern hearing aids for the profoundly hearing impaired: Speech perception and auditory abilities," J. Acoust. Soc. Am..
Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.
Fu, Q. J., Shannon, R. V., and Wang, X. S. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.
Gantz, B. J., and Turner, C. W. (**2003**). "Combining acoustic and electrical hearing," Laryngoscope **113**, 1726–1730.
Geurts, L., and Wouters, J. (**2001**). "Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants," J. Acoust. Soc. Am. **109**, 713–726.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Green, T., Faulkner, A., and Rosen, S. (**2002**). "Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants," J. Acoust. Soc. Am. **112**, 2155–2164.
Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (Springer, New York).
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.
House, A. S. (**1960**). "Formant band widths and vowel preference," J. Speech Hear. Res. **3**, 3–8.
House, A. S. (**1961**). "On vowel duration in English," J. Acoust. Soc. Am. **33**, 1174–1178.
IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **AU-17**(3), 225–246.
Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.
Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
Kong, Y. Y., Stickney, G. S., and Zeng, F. G. (**2005**). "Speech and melody recognition in binaurally combined acoustic and electric hearing," J. Acoust. Soc. Am. **117**, 1351–1361.
Kong, Y. Y., Cruz, R., Jones, J. A., and Zeng, F. G. (**2004**). "Music perception with temporal cues in acoustic and electric hearing," Ear Hear. **25**, 173–185.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). "Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees," J. Acoust. Soc. Am. **96**, 2664–2673.
Moore, B. C. (**2003**). "Coding of sounds in the auditory system and its relevance to signal processing and coding in cochlear implants," Otol. Neurotol. **24**, 243–254.
Moore, B. C. J., Glasberg, B. R., and Peters, R. W. (**1985**). "Relative dominance of individual partials in determining the pitch of complex tones," J. Acoust. Soc. Am. **77**, 1853–1860.
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.
Nilsson, M., Soli, S., and Sullivan, J. (**1994**). "Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise," J. Acoust. Soc. Am. **95**, 1085–1099.
Plomp, R. (**1964**). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.
Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on *F*0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Rosen, S. M., Fourcin, A. J., and Moore, B. C. J. (**1981**). "Voice pitch as an aid to lipreading," Nature (London) **291**, 150–152.
Scheffers, M. T. M. (**1983**). "Sifting vowels: Auditory pitch analysis and sound segregation," Groningen University, The Netherlands.
Shackleton, T. M., and Carlyon, R. P. (**1994**). "The role of resolved and unresolved harmonics in pitch perception and frequency-modulation discrimination," J. Acoust. Soc. Am. **95**, 3529–3540.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Stickney, G., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Acoust. Soc. Am. **116**, 1081–1091.
Summerfield, A. Q., and Assmann, P. F. (**1991**). "Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignment," J. Acoust. Soc. Am. **89**, 1364–1377.
Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.
Tyler, R. S., Parkinson, A. J., Wilson, B. S., Witt, S., Preece, J. P., and Noble, W. (**2002**). "Patients utilizing a hearing aid and a cochlear implant: Speech perception and localization," Ear Hear. **23**, 98–105.
von Ilberg, C., Kiefer, J., Tillein, J., Pfenningdorff, T., Hartmann, R., Sturzebecher, E., and Klinke, R. (**1999**). "Electric-acoustic stimulation of the auditory system—New technology for severe hearing loss," ORL **61**, 334–340.
Wilson, B. S. (**1997**). "The future of cochlear implants," Br. J. Audiol. **31**, 205–225.
Zeng, F.-G. (2004). "Trends in cochlear implants," Trends in Amplification **8**, 1–34.

The influence of perceived sexual orientation on fricative identification

Benjamin Munson,[a] Sarah V. Jefferson, and Elizabeth C. McDonald
Department of Speech-Language-Hearing Sciences, University of Minnesota, Minneapolis, Minnesota 55455

(Received 9 May 2005; revised 11 January 2006; accepted 17 January 2006)

Listeners are more likely to hear a synthetic fricative ambiguous between /s/ and /ʃ/ as /ʃ/ if it is appended to a woman's voice than a man's voice [Strand and Johnson, in *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference* (Mouton de Gruyter, Berlin, 1996), pp. 14–26]. This study expanded on this finding by replicating the result with a much larger group of male and female talkers than had been examined previously, by examining whether phonetic context mediates the influence of talker sex on fricative identification, and by examining whether talkers' perceived sexual orientation influences fricative identification. Stimuli were created by pairing a synthetic nine-step /s/-/ʃ/ continuum with tokens of /æk/ and /ɪp/ taken from productions of *shack* and *ship* by 44 talkers whose perceived sexual orientation had been reported previously [Munson *et al.*, J. Phonetics (in press)]. Listeners participated in a series of two-alternative sack-shack and sip-ship identification experiments. Listeners identified more /ʃ/ tokens for women's voices than for men's voices for both continua. Lesbian/bisexual-sounding women elicited more sack and sip responses than heterosexual-sounding women. No consistent influence of perceived sexual orientation on fricative identification was noted for men's voices. Results suggest that listeners are sensitive to the association between fricatives' center frequencies and perceived sexual orientation in women's voices, but not in men's voices. © *2006 Acoustical Society of America.*
[DOI: 10.1121/1.2173521]

PACS number(s): 43.71.Es, 43.71.Bp, 43.70.Fq, 43.70.Gr [ALF] Pages: 2427–2437

I. INTRODUCTION

A. Sex and sex typicality in speech perception and production

It is axiomatic that the acoustic characteristics of men and women's speech are different. Childers and Wu (1991) and Wu and Childers (1991) found that automatic classification algorithms could correctly identify the sex of 52 speakers based on a variety of global and local acoustic measures, such as vocal-tract resonant frequencies, fundamental frequency and spectral tilt of the voicing source, and vowel formant frequencies. Human identification of sex from audio-only speech signals is also very high, even when these signals are short in duration and have had acoustic information removed or filtered (Lass *et al.*, 1980, 1979).

The differences between the acoustic characteristics of men and women's voices are the consequence of both anatomical differences between the sexes and learned, socially conditioned differences. The influence of anatomical differences on speech acoustics arises in part because women, on average, have smaller-sized vocal tracts than men (Fitch and Giedd, 1999). Consequently, women produce vowels with higher frequency formants than men (Fant, 1966, 1975; Hillenbrand *et al.*, 1995, *inter alia*). In addition, men's longer, thicker vocal folds lead them, on average, to have lower fundamental frequencies than women (Hirano *et al.*, 1983; Titze, 1989). However, talkers may actively manipulate their articulation to change the same acoustic parameters, such as formant frequency, that are affected by vocal-tract size. Consequently, not all male-female differences in speech production result from anatomic and physiologic differences between the sexes. This can be seen in studies showing that there is a relatively poor correlation between measures of body size (presumably correlated with vocal-tract size) and both formant frequencies and perceived size (González, 2004; van Dommelen and Moxness, 1995). Learned, socially conditioned sex differences were found by van Bezooijen (1995), who demonstrated that male-female differences in vocal pitch were larger in Japanese speakers than in Dutch speakers. This presumably reflects a learned pattern associated with the Japanese cultural practice of ascribing positive affective ratings to women whose fundamental frequencies are high. Sex differences influence children's early speech acquisition. Perry *et al.* (2001) showed that children's sex can be reliably identified from speech in advance of the developmental anatomical changes that would lead to sex-related speech differences.

The relative influence of learned factors and anatomical variation on sex differences can be seen in studies of the acoustic characteristics of women and men's fricatives. Jongman *et al.* (2000) showed that women produce the sibilant fricatives /s/ and /ʃ/ with higher peak frequencies than men, plausibly because of their smaller-sized vocal tracts. These differences are perceptible to listeners, as shown by Schwartz's (1968) finding that talker sex could be identified at greater-than-chance levels from isolated voiceless fricatives. However, Stuart-Smith (2004) provides evidence that learned patterns can mediate the influence of sex on fricative production. Stuart-Smith examined the acoustic characteris-

[a]Electronic mail: munso005@umn.edu

tics of /s/ in groups of male and female talkers who varied in age and social class. She found that the difference between men and women was smallest for the younger working-class talkers. Stuart-Smith conjectured that the younger working-class girls were actively manipulating their production of fricatives in a way that minimized sex differences.

The focus of this investigation is the influence of sex and sex typicality (the extent to which an individual's speech conforms to the norms for their sex) on speech perception. Sex differences in speech production are salient to listeners in a variety of different listening tasks. As mentioned previously, listeners can identify a talker's sex from audio speech signals alone, even when these are quite short, have had some acoustic information removed by low-pass filtering, and are by talkers of an unfamiliar dialect (Lass *et al.*, 1979, 1980; Schwartz, 1968). *Tacit* knowledge of sex differences can be seen in studies showing that phonetic identification is sensitive to perceived talker sex. One example of this is given in Johnson *et al.* (1999). Johnson *et al.* examined perception of a synthetic /ʊ/-/ʌ/ continuum in an audiovisual speech perception experiment. Male and female faces were paired with four voices, which varied in talker sex and listener-perceived sex typicality. When the continuum was matched with a female face, listeners reported more /ʊ/ percepts than when it was paired with a male face, perhaps reflecting their belief that the speech was produced by someone with a smaller vocal tract. The sex and sex typicality of the voice also influenced identification boundaries, with the most /ʊ/ percepts for the stereotypical female voice and the least for the stereotypical male voice. In a separate experiment, Johnson *et al.* (1999) showed that these effects emerge when listeners are asked to merely imagine that a gender-neutral synthesized vowel had been produced by a man or a woman. The latter finding is consistent with Johnson's (1990) earlier finding that identification of a synthetic /ʊ/-/ʌ/ continuum is different when embedded in carrier phrases with different fundamental frequencies. The f0 of the carrier phrase gave the perceptual illusion that the target continuum had been produced by different talkers; the influence of carrier phrase f0 on /ʊ/-/ʌ/ perception presumably reflects listener accommodation to the perception of it having been produced by different talkers.

Other research has shown that fricative identification is affected by talker sex. When a synthetic /s/-/ʃ/ continuum is appended to a woman's voice, listeners report more /ʃ/ percepts than when the same stimuli are appended to a man's voice (Strand, 1999; Strand and Johnson, 1996). This suggests that listeners are sensitive to the fact that women produce both fricatives with higher-frequency energy than men. Strand and Johnson (1996) present some evidence suggesting that the influence of talker sex on phoneme identification is gradient: less-prototypical male and female voices elicited a smaller effect than more-prototypical ones.

One reasonable hypothesis is that the perceptual phenomena in Johnson (1991), Johnson *et al.* (1999), and Strand and Johnson (1996) represent listeners' perceptual adjustments to talkers' vocal-tract sizes. Women, in general, have smaller-sized vocal tracts than men (Fitch and Giedd, 1999). The influence of talker sex on speech perception may simply reflect listeners gauging talkers' vocal tract sizes from available acoustic cues, then making the appropriate perceptual adjustment. Indeed, this was proposed by May (1976), who showed that synthetic /s/-/ʃ/ perception was influenced by manipulating the center frequency and spacing of formants in a synthetic /æ/ to give the illusion of a larger or smaller talker. The direction of the effect was similar to that in Strand and Johnson (1996): larger-sounding talkers elicited more /s/ percepts than smaller-sounding ones. Johnson *et al.* (1999) argued against this hypothesis. A simple process in which listeners compute vocal-tract size from the available cues is incompatible with Johnson *et al.*'s audio-visual integration experiments and their imagined-gender experiment. Instead, they proposed that "… speaker normalization in speech perception is based on abstract, subjective talker representations, and that listeners perceive talker identify from the totality of listening situations, including direct acoustic cues for vocal tract length […], indirect cues […], visual cues, and even imagined talker characteristics" (p. 380). Johnson *et al.* proposed that the influence of gender on speech perception reflects listeners' detailed memories of the acoustic characteristics of male and female talkers whom they have encountered during language use.

B. The current investigation

1. Perceived sexual orientation and fricative identification

The purpose of the current investigation is to explore further the influence of sex and sex typicality on fricative identification. This study expands on previous studies of this phenomenon in three ways. The use of a larger cohort of talkers allows us to make a more definitive statement about the influence of sex and sex typicality on fricative identification than has been offered by earlier work. The way in which we examine the gradient influence of sex typicality on fricative perception is by examining the identification of a synthetic /s/-/ʃ/ continuum appended to naturally produced base tokens from talkers whose voices vary in their perceived sexual orientation. There are, according to popular stereotypes, distinctive speech styles associated with sexual orientation. These speech styles purportedly allow naive listeners to identify talkers' sexual orientation at greater-than-chance levels in the absence of an overt disclosure. This has been supported by a variety of recent studies (Gaudio, 1994; Linville, 1998; Pierrehumbert *et al.*, 2004; Smyth *et al.*, 2003). These studies have examined the relations among self-identified sexual orientation, acoustic measures of speech production, and listener judgments of sexual orientation.

Recently, Munson *et al.* (in press) examined acoustic and perceptual characteristics of 44 adults, including equal numbers of gay, lesbian, or bisexual (henceforth *GLB*) and heterosexual men and women. Munson *et al.* found that gay/bisexual (henceforth *G*/B) men produced a higher F1 frequency in /æ/ and /ɛ/ and a more negatively skewed /s/ spectrum than heterosexual men. Lesbian/bisexual (*L*/*B*) women produced a lower F1 frequency in /ɛ/ and a lower F2 frequency in /oʊ/ than heterosexual women. GLB and hetero-

sexual people did not differ in height, suggesting that the observed differences were not related to vocal-tract size. In a perception experiment, Munson *et al.* (in press) showed that naive listeners judged GLB people to be more GLB sounding than heterosexual people, though there was some overlap between the groups. Regression analyses showed that men who produced low vowels with a high F1, back vowels with a high F2, and /s/ with a highly negatively skewed spectrum were more likely to be rated as G/B sounding than men with the opposite characteristics. Women were likely to be rated as L/B sounding if they produced a low F1 in low vowels and a low F2 in back vowels.

The current investigation examines the influence of talker sex and perceived sexual orientation on fricative identification using the 44 talkers from Munson *et al.* (in press). Perceived sexual orientation in speech is used in this study as an index of sex typicality. Individuals who were reliably rated as GLB sounding in Munson *et al.* (in press) would presumably be comparable to the less-prototypical speakers from Johnson *et al.* (1999) and Strand and Johnson (1996). The perceived sexual orientation measures from Munson *et al.* are used in the current investigation in two ways. First, they are used to categorize the 44 speakers into groups of GLB- and heterosexual-sounding talkers for analyses of variance examining the influence of perceived sexual orientation on fricative identification. For these analyses, we hypothesize that GLB-sounding people will elicit different patterns of phoneme identification than heterosexual-sounding people, with G/B-sounding men eliciting more /ʃ/ judgments from listeners than heterosexual-sounding men, and L/B-sounding women eliciting more /s/ percepts than heterosexual-sounding women.

Second, perceived sexual orientation measures from Munson *et al.* are used in regression analyses predicting summary measures of fricative identification. These analyses allow us to make a stronger test of Strand and Johnson's hypothesis that there is a gradient relationship between sex typicality and fricative identification than their investigation offered. This investigation further tests this hypothesis through regression analyses using a second set of measures of sex typicality as predictors of fricative identification patterns. These were listener ratings of perceived masculinity/femininity, collected specifically for this investigation. Previous research on men's speech has suggested that measures of perceived sexual orientation and measures of perceived masculinity are correlated but not identical (Gaudio, 1994; Smyth *et al.*, 2003). However, no previous study has examined these relationships in women's speech, and some of the specific findings of Munson *et al.* (in press) suggested that perceived sexual orientation might not be as strongly correlated with masculinity and femininity in those 44 talkers as it had in earlier research. Thus, we collected perceived masculinity and femininity judgments for this study to examine whether that measure was more closely related to fricative identification than perceived sexual orientation. Together, these two analyses provide a rigorous test of the hypothesis that sex typicality influences fricative identification.

2. Acoustic variables mediating the influence of sex on fricative identification

The second way in which this study differs from previous research is that we examine which acoustic characteristics of the stimuli predict patterns of fricative identification. Both Johnson *et al.* (1999) and Strand and Johnson (1996) used too few talkers to identify which of the many acoustic parameters that differ between men and women were likely to have influenced listeners' phonetic identification. The use of 44 talkers in the current investigation allows us to conduct an exploratory analysis in which we use regression to predict summary measures of fricative identification from acoustic characteristics of the stimuli, including the f0, F1, F2, and duration of the vocalic portions of the stimuli.

3. Vowel context and fricative perception

The final way in which this study differs from previous research is by examining whether the influence of talker sex and perceived sexual orientation on fricative identification differs systematically as a function of the vowel with which the synthetic fricative stimuli are paired. Strand and Johnson's (1996) study utilized one vowel context, /ɑ/: listeners identified a *sod-shod* continuum. Previous studies of the influence of sex and sexual orientation on vowel production have found that differences are mediated by vowel height. Fant's classic (1966) and (1975) studies showed larger sex differences in low vowels than in non-low vowels. Munson *et al.* (in press) extended this finding by showing greater acoustic differences between G/B and heterosexual men's low vowels than non-low vowels. Given these findings, it is reasonable to hypothesize that the influence of sex and perceived sexual orientation on fricative identification would be proportional to the influence of those variables on vowel acoustics. A stronger influence of sex and perceived sexual orientation might be found for synthetic continua appended to low-vowel words than high-vowel words, given that these variables influence low-vowel acoustics more strongly than high-vowel acoustics. This is tested in the current study by examining identification of a synthetic /s/-/ʃ/ continuum appended to two types of words: those containing low vowels and those containing high vowels.

II. METHODS

A. Stimuli

Stimuli were 44 *sack-shack* and 44 *sip-ship* continua created by splicing synthetic fricatives with natural productions by the words shack and ship by male and female adults who varied in their perceived sexual orientation, as gauged by naive listeners. The use of one continuum containing a low front vowel, sack-shack, and one containing a high-front vowel, sip-ship, allowed us to assess whether the strength of the influence of sex and perceived sexual orientation on fricative identification is proportional to the influence of those variables on the acoustic characteristics of the vowels the fricatives were appended to. The words had an average familiarity value of 6.7 or higher on a 7-point scale, using the values described in Pisoni *et al.* (1985).

1. Talker characteristics

Natural speech tokens from 44 adults were used to create the fricative-identification stimuli. These talkers are described in detail in Munson *et al.* (in press). This group included equal numbers of self-identified GLB and heterosexual men and women (11 in each group) from the North Central American English dialect region (Labov *et al.*, 2005), which comprises Minnesota, Western Wisconsin, Eastern North and South Dakota, and Northern Iowa. All were native speakers of English. Talkers produced multiple repetitions of isolated CVC words containing a variety of vowels and consonants. In a second experiment reported in Munson *et al.* (in press), a group of 40 listeners was played a subset of the words that the 44 talkers produced, and asked to judge their perception of the talkers sexual orientation on a five-point equal-interval scale. Each listener rated each talker four times: once based on words containing low front vowels and sibilant fricatives, once based on words containing low front vowels and no sibilant fricatives, once based on words with back round vowels and sibilant fricatives, and once based on words with back round vowels and no sibilant fricatives. Average perceived sexual orientation was calculated for each talker, collapsed across the different conditions of the experiment. These values were used to divide the 44 talkers into groups of GLB- and heterosexual-sounding talkers for analyses of variance. The distribution of average perceived sexual orientation scores was relatively flat, making a simple arbitrary median split an inappropriate method to group the talkers. However, we were able to identify two gaps in the distribution (between 2.95 and 3.45 for men and between 2.17 and 2.70 for women) that were used as the cutoff points for defining GLB- and heterosexual-sounding talkers. Ten men were classified as G/B sounding (mean rating=4.04, SD=0.36) and 12 were classified as heterosexual sounding (mean rating=2.44, SD=0.35). Nine women were classified as L/B sounding (mean rating=3.40, SD =0.36) and 13 were classified as heterosexual sounding (mean rating=1.96, SD=0.16).

One objective of the current study was to examine whether measures of talkers' perceived masculinity or femininity predicted patterns of fricative identification. To examine this, a second perception experiment was conducted to measure perceived masculinity/femininity. The listeners in this experiment were ten adults who did not participate in the fricative-perception experiment in the current study or in the earlier perception studies reported by Munson *et al.* (in press). The design and procedure for the experiment closely paralleled the perception experiment in Munson *et al.* Stimuli were 14 words each from the 44 talkers: three words with front vowels and sibilant fricatives (*gas, said, same*), three words with front vowels and no sibilant fricatives (*bell, fade, path*), three words with back-round vowels and sibilant fricatives (*loose, soap, soon*), and three words with back-round vowels and no sibilant fricatives (*hoop, note, tooth*), as well as single tokens of the words ship and shack. These had a 22.05-kHz sampling rate with 16-bit quantization and had been processed through an 11.025-kHz antialiasing filter. The amplitudes of the 616 stimuli were peak normalized.

The experiment was designed and executed using the *E-prime* experiment-management software (Schneider *et al.*, 2002). On each trial, three words were played. These were either the triplets of words with similar phonemic composition, described in the previous paragraph, or three repetitions of the token of shack or ship. An orthographic display of the words was presented on a 17-in. computer monitor in 36-point courier font concurrent with their audio presentation. After each trial, listeners rated the talkers' sexual orientation on a 5-point equal-interval scale. The experiment was blocked by talker sex. For the men, 1 indicated *sounds very masculine*, 3 indicated *sounds somewhat masculine*, and 5 indicated *sounds not at all masculine*. For the women, 1 indicated *sounds very feminine*, 3 indicated *sounds somewhat feminine*, and 5 indicated *sounds not at all feminine*. Cards with the above wording were placed above buttons on a button-box. Participants responded by pressing buttons; their responses were logged automatically. Words were presented over a powered speaker (Roland DS-90A) located 0° azimuth from the listener's head at a level of approximately 65 dB HL, as calibrated prior to the experiment using the slow, dB A-scale of a sound-level meter positioned at the approximate location of the listener's head during the experiment. Experimental blocks were preceded by a short practice block containing two talkers and two words not used in the experiment. Experimental stimuli were presented in fully randomized order.

Three average perceived masculinity/femininity ratings were calculated for each talker. One of these was the average of the four ratings made over front- and back-vowel words with and without a sibilant fricative. The other two were ratings made over shack and ship, respectively. These values were used in regression analyses predicting different summary measures from the fricative-identification experiment. These measures correlated strongly with each other and with the measures of perceived sexual orientation. For the 22 women, all correlations were significant at the $\alpha < 0.001$ level. The correlation coefficients (Pearson's r) ranged from 0.76 to 0.93. For the 22 men, all correlations were significant at the $\alpha < 0.01$ level. The correlation coefficients were slightly lower than those for women, ranging from 0.52 (for the correlation between perceived sexual orientation and perceived masculinity based on ship tokens) to 0.87. The significant correlations between the judgments made over the larger group of stimuli and those made on shack and ship alone suggest that the specific tokens used to create the stimuli were well representative of the larger set of tokens in Munson *et al.* (in press).

2. Naturally produced bases

The naturally produced bases were the ship and shack tokens used in the perceived masculinity/femininity experiment described in the previous section. The /ʃ/ was edited off of these. The frication was removed at a point of zero crossing, so that the resulting stimuli would not include any transients between the frication noise and the natural bases. The duration of the vocalic portion of the natural base tokens was measured, as were the f0 and F1 and F2 frequencies at vowel midpoint. The f0 and formant-frequency measures were made using the psychophysically motivated ERB and Bark

TABLE I. Mean values of acoustic parameters of the vocalic portions of the naturally produced VC portions of the ship and shack stimuli. Standard deviations in parentheses.

Word	Sex	Perceived sexual orientation	f0 (ERB)	F1 (Bark)	F2 (Bark)	Duration (ms)	H2-H1 (dB)
Ship	Women	L/B-sounding	5.0 (0.3)	5.1 (0.5)	12.6 (0.5)	123 (34)	
		Heterosexual-sounding	5.4 (0.9)	5.4 (0.7)	13.1 (0.6)	120 (38)	
	Men	G/B-sounding	3.5 (0.4)	4.7 (0.4)	12.1 (0.4)	123 (19)	
		Heterosexual-sounding	3.3 (0.5)	4.6 (0.3)	12.1 (0.6)	123 (21)	
Shack	Women	L/B-sounding	4.4 (0.7)	7.2 (0.5)	12.9 (0.5)	225 (46)	5.5 (4.1)
		Heterosexual-sounding	5.1 (0.4)	7.9 (0.5)	13.1 (0.4)	205 (43)	5.9 (3.3)
	Men	G/B-sounding	3.0 (0.3)	6.8 (0.6)	11.9 (0.4)	223 (32)	2.1 (1.9)
		Heterosexual-sounding	3.0 (0.3)	6.4 (0.5)	12.1 (0.5)	231 (22)	2.0 (3.2)

scales, respectively (Hermes and van Gestel, 1991, Zwicker and Ternhardt, 1980). In addition, an acoustic index of vocal breathiness, H2-H1 amplitude (Klatt and Klatt, 1990), was measured for /æ/ in the /æk/ base. The specific method used to calculate H2-H1 is detailed in Munson *et al.* (in press). The H2-H1 amplitude was not measured for /i/, as the low first formant frequency of this vowel precludes the reliable measurement of H2 and H1 amplitudes. These measures were made to determine the extent to which the tokens used in this experiment reflected the findings of Munson *et al.*'s analysis of the larger speech corpus from this group, as well as for use in regression analyses predicting fricative identification patterns. These measures were made using the PRAAT signal-processing program (Boersma and Weenink, 2004). Formant measures were made automatically from a LPC formant track with eight coefficients. These values are shown in Table I.

To examine how well the stimuli chosen for this investigation represent the entire corpus of words analyzed by Munson *et al.* (in press), we examined correlations between the acoustic measures in Table I and the average values for the same parameters in a larger set of words examined by Munson *et al.* (in press). For this analysis, the values from Munson *et al.* that were chosen were the average duration, f0, F1, F2 and H1-H2 of /æ/ and /i/, excluding the tokens used in this investigation. The correlations between the average measures and the measures in Table I were very high. Pearson's r values ranged from 0.75 to 0.92 ($p < 0.001$ for all correlations). All individual correlations were significant ($\alpha < 0.05$) when men and women were examined separately. This analysis provides further evidence that the tokens selected for this study were well representative of the larger set from Munson *et al.* (in press).

3. Synthetic fricatives

The nine-step synthetic fricative continuum with which the /æk/ and /ɪp/ bases were paired was made with the parallel/cascade synthesizer described by Klatt and Klatt (1990). The fricatives varied in their third and fourth formant frequencies. The step differed in equal Bark-sized units. The values for the F3 and F4 center frequencies and bandwidths of the nine steps of the continuum are shown in Table II. Table II also shows the four spectral moments of the frica-

TABLE II. Synthesis parameters and spectral characteristic of the /s/-/ʃ/ continuum.

Step	Center of gravity (Hz)	Standard deviation (Hz)	Skewness	Kurtosis	F3[a]	F4[b]	B3[c]	B4[d]
1	2508	489	−1.92	7.48	2466	3108	247	311
2	2884	579	−1.94	7.34	3265	3573	327	357
3	3353	690	−1.88	6.84	2838	4107	284	411
4	3917	768	−2.09	8.13	3754	4717	376	472
5	4546	823	−2.15	9.72	4314	5417	431	542
6	5287	903	−2.08	10.41	4954	6219	496	622
7	6201	963	−2.06	11.35	5689	7137	569	714
8	7182	1005	−1.63	10.93	6530	8191	653	819
9	8572	1117	−1.46	8.30	7500	9405	750	941

[a]Third formant frequency (Hz).
[b]Fourth formant frequency (Hz).
[c]Third formant bandwidth (Hz).
[d]Fourth formant bandwidth (Hz).

tives (Forrest *et al.*, 1988). These measures have been shown to discriminate between naturally produced tokens of /s/ and /ʃ/ (Forrest *et al.*, 1988; Jongman *et al.*, 2000). They were 160 ms and had a natural amplitude contour.

The amplitudes of the naturally produced stimulus bases were peak normalized prior to concatenating them with the synthetic fricatives. The relative amplitude of the fricative and vowel portions were held constant across the nine stimuli, as this has been shown previously to influence fricative identification (Hedrick and Ohde, 1993). All 176 endpoint stimuli were reliably identified as sack, shack, sip, or ship, as determined by the first two authors.

B. Listeners and procedures

Ten adult listeners, ages 18–35, with no history of any speech, language, or hearing problems, participated in the fricative-identification experiment. Each adult was paid US$50.00 for participating. Given the large number of stimuli, the experiment was blocked by talker sex and continuum type. The choice to block by continuum type was made so that the responses would be equivalent across different trials. Though the sex of all of the 44 talkers was reliably identified by listeners in Munson *et al.* (in press), we chose to block by sex in case some of the listeners misperceived the sex of individual talkers. There were 22 continua in each block. Within blocks, each stimulus was presented five times. Stimulus order was fully randomized within blocks. The listeners heard each block of stimuli twice, for a total of 7920 responses per listener (8 blocks $\times$ 22 continua per block $\times$ 9 steps per continuum $\times$ 5 presentations per step). Block order was quasi-randomized, with the constraint that listeners heard two different blocks per session; approximately equal numbers of listeners participated in the different orders. The participants each completed four listening sessions, with two blocks per session. Each session took approximately an hour and a half, for a total of approximately six hours of listening per subject.[1]

Testing took place in a double-walled sound-treated booth. The stimuli were output from a computer through Sennheiser HD 280 Pro headphones at a level of approximately 65 dB HL. On each trial, the word *listen* was presented in 36-point courier font in the center of a 17-in. monitor for 1.5 s, followed by a stimulus, followed by a screen instructing the listeners to respond by pressing a button on a button-box. The buttons were labeled with the responses written (e.g., sack, sip, shack, or ship). Prior to each experimental block, practice words were given. These stimuli were created by appending the endpoint /s/ and /ʃ/ stimuli to base tokens recorded by two talkers not used in the experiment (the first and third authors). Throughout the experiment, nonspeech stimuli consisting of a 250-ms warble tone with a center frequency of 1 kHz were presented in random intervals, after which the participant was instructed to press a button. The purpose of this was to decrease the monotony of each session.

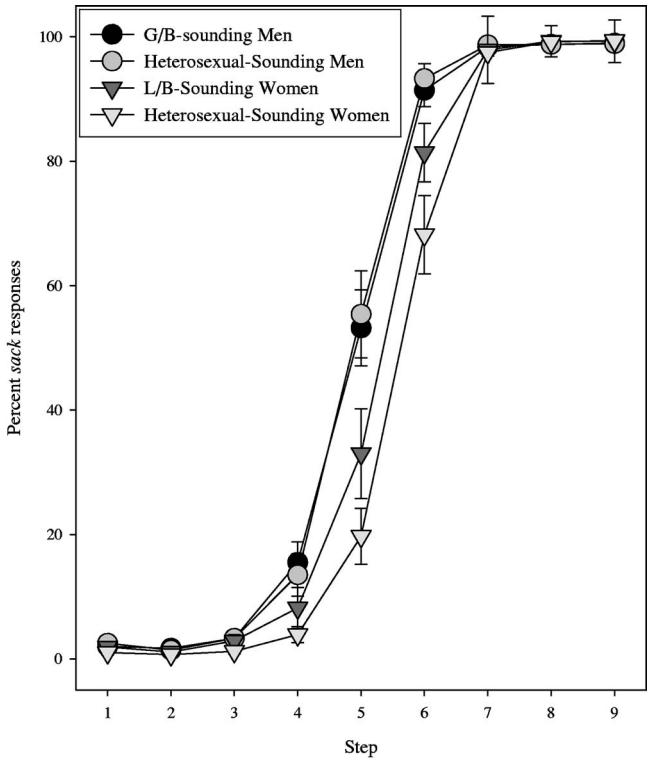


FIG. 1. Percent sack responses for GLB- and heterosexual-sounding men and women on the nine-step sack-shack continuum. Error bars represent one standard error of measurement.

C. Analysis

For each of the ten listeners, average percent sip and sack responses at each step were tallied separately for the 10 G/B- and the 12 heterosexual-sounding men, and for the 9 L/B- and 13 heterosexual-sounding women. These were used as the dependent measures in analyses of variance. For each of the 44 talkers, average percent sip and sack responses across the ten listeners were tallied separately for each step. These were used as the dependent measures in regression analyses.

III. RESULTS

A. Analyses of variance

Rationalized arcsine-transformed percentage sack and sip responses were submitted to separate three-factor (9 step $\times$ 2 sex $\times$ 2 perceived sexual orientation) within-subjects ANOVAs. This ANOVA design is identical to that used by Strand and Johnson (1996). Average sack-shack and sip-ship identification functions for the four groups are shown in Figs. 1 and 2.

For the sip-ship continuum, there were significant main effects of sex, $F[1,9]=48.9$, $p<0.001$, $\eta^2_{\text{partial}}=0.85$; perceived sexual orientation, $F[1,9]=12.7$, $p<0.01$, $\eta^2_{\text{partial}}=0.59$; and step, $F[8,72]=694.6$, $p<0.001$, $\eta^2_{\text{partial}}=0.99$. There was also a significant interaction between sex and step, $F[8,72]=32.7$, $p<0.001$, $\eta^2_{\text{partial}}=0.78$. Tests of significant main effects showed an effect of sex on performance at steps 4, 5, and 6 only: women elicited more /ʃ/ percepts than men at all three steps ($F[1,9]=\{40.5–35.0\}$, $p<0.001$ for all

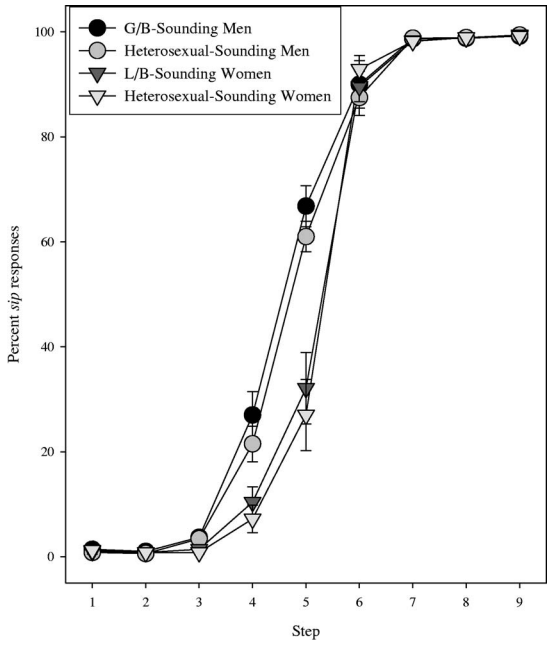


FIG. 2. Percent sip responses for GLB- and heterosexual-sounding men and women on the nine-step sip-ship continuum. Error bars represent one standard error of measurement.

tests). At steps 4 and 5, L/B-sounding women elicited fewer /ʃ/ percepts than heterosexual-sounding women ($F[1,9]=\{8.9,20.6\}$, $p<0.01$ for both tests). Contrary to predictions, G/B-sounding men also elicited significantly fewer /ʃ/ percepts than heterosexual-sounding men at steps 4, 5, and 6 ($F[1,9]=\{8.9–10.2\}$, $p<0.01$ for all tests).

For the sack-shack continuum, there were significant main effects of sex, $F[1,9]=96.9$, $p<0.001$, $\eta^2_{\text{partial}}=0.93$; perceived sexual orientation, $F[1,9]=10.9$, $p<0.05$, $\eta^2_{\text{partial}}=0.55$; and step, $F[8,72]=108.5$, $p<0.001$, $\eta^2_{\text{partial}}=0.98$. There were significant interactions between sex and perceived sexual orientation, $F[1,9]=2.4$, $p<0.05$, $\eta^2_{\text{partial}}=0.21$; sex and step, $F[8,72]=29.7$, $p<0.001$, $\eta^2_{\text{partial}}=0.77$; and among sex, step, and perceived sexual orientation, $F[8,72]=6.4$, $p<0.01$, $\eta^2_{\text{partial}}=0.42$. *Posthoc* simple effects ANOVAs showed an effect of sex on performance at steps 4, 5, and 6 only: women elicited more /ʃ/ percepts than men at all three steps ($F[1,9]=\{26.7–109.0\}$, $p<0.01$ for all tests). At steps 5 and 6, L/B-sounding women elicited fewer /ʃ/ percepts than heterosexual-sounding women ($F[1,9]=\{14.8–81.8\}$, $p<0.01$ for both tests). There were no statistically significant differences between G/B- and heterosexual-sounding men for any step along the continuum.

The next set of ANOVAs compared performance on the two continua. As Figs. 1 and 2 show, the largest effects of sex and perceived sexual orientation were at step 5. Thus, to assess whether performance on the two continua differed systematically, we conducted a three-factor (2 sex $\times$ 2 perceived sexual orientation $\times$ 2 continuum type) mixed-model ANOVA examining average /s/-initial word judgments (sack or sip) at step 5. There were significant main effects of sex ($F[1,9]=75.6$, $p<0.001$, $\eta^2_{\text{partial}}=0.89$) and perceived sexual orientation ($F[1,9]=19.9$, $p<0.01$, $\eta^2_{\text{partial}}=0.69$). These were qualified by significant interactions between sex and perceived sexual orientation ($F[1,9]=11.1$, $p<0.01$, $\eta^2_{\text{partial}}=0.55$) and among continuum, sex, and perceived sexual orientation ($F[1,9]=11.0$, $p<0.01$, $\eta^2_{\text{partial}}=0.55$).

TABLE III. Stepwise multiple regressions predicting the average identification of sack at step 5 from selected acoustic characteristics of the stimuli. $F[2,41]=54.0$, $p<0.001$, $R^2=0.73$ for the full model based on the average sack judgments for all talkers; $F[2,19]=8.5$, $p<0.01$, $R^2=0.26$ for the full model based on the average sack judgments produced by women. There were no significant predictors of sack responses for men's voices.

	Step	Variable	ΔR^{2a}	B[b]	SE B[b]	β^b
All talkers	1	f0	0.656[c]	−10.6[c]	1.78	−0.611[c]
	2	F1 frequency	0.069[c]	−7.6[c]	2.39	−0.329[c]
Women	1	F1 frequency	0.257[d]	−7.8[d]	2.97	−0.507[d]

[a]Increase in R^2 over the model containing all previous steps.
[b]Coefficients for the full model.
[c]$p<0.01$.
[d]$p<0.05$.

Posthoc simple-effects ANOVAs were used to examine the effects of continuum and perceived sexual orientation separately for men and women. For men, there were no significant main effects of continuum type or of perceived sexual orientation, but these did interact significantly ($F[1,9]=6.1$, $p<0.05$, $\eta^2_{\text{partial}}=0.40$). Paired comparisons showed a significant effect of continuum type for G/B-sounding men (G/B-sounding men elicited more sip responses than sack responses), but not for heterosexual-sounding men. In contrast, there was a significant main effect of perceived sexual orientation for women's voices ($F[1,9]=22.2$, $p<0.01$, $\eta^2_{\text{partial}}=0.71$) (L/B-sounding women elicited more sip and sack responses than heterosexual-sounding women) but no effect of continuum type.

B. Regression

1. Stimulus acoustics

The first set of regression analyses examined which acoustic characteristics of the stimuli predict patterns of phoneme identification. The dependent measures in these regressions were rationalized arcsine-transformed sack or sip percepts at step 5, averaged across the ten listeners. These were chosen as the dependent measures because this was the step at which the largest influences of sex and perceived sexual orientation were found.

The first two regressions examined predictors for the entire group of 44 talkers. For the sip regression, the vowel duration as well as midpoint F1 frequency, F2 frequency, and f0 of the sip stimuli were the independent measures. For the sack regression, the same acoustic measures of the sack stimuli served as independent measures, as did the H2-H1 amplitude at vowel midpoint. These were entered into the regression stepwise if they accounted for a significant proportion of variance ($\alpha<0.05$) beyond what was accounted for by the variable(s) entered on the prior step(s). The results of these regressions, as well as the other regressions presented in this section, are shown in Tables III and IV.

When the entire group of 44 talkers was examined, f0 accounted for a large proportion of variance (65.6%) in sack

TABLE IV. Stepwise multiple regressions predicting the average identification of sip at step 5 from selected acoustic characteristics of the stimuli. $F[2,41]=51.8$, $p<0.001$, $R^2=0.72$ for the full model based on the average sip judgments for all talkers; $F[2,19]=11.4$, $p<0.01$, $R^2=0.54$ for the full model based on the average sip judgments produced by women talkers; and $F[1,20]=4.5$, $p<0.05$, $R^2=0.18$ for the full model based on the average sip judgments produced by men talkers.

	Step	Variable	ΔR^{2a}	B[b]	SE B[b]	β[b]
All talkers	1	f0	0.624[c]	−11.2[c]	1.64	−0.634[c]
	2	F1 frequency	0.092[c]	−10.8[c]	2.95	−0.342[c]
Women	1	F1 frequency	0.455[c]	−7.2[c]	1.62	−0.686[c]
	2	f0	0.090[d]	−2.7[d]	1.41	−0.301[d]
Men	1	F2 frequency	0.183[d]	−7.1[d]	3.29	−0.428[d]

[a]Increase in R^2 over the model containing all previous steps.
[b]Coefficients for the full model.
[c]$p<0.01$.
[d]$p<0.05$.

responses. As expected, people with higher f0 elicited fewer sack responses than those with lower f0s. F1 frequency accounted for an additional 6.9% of variance in sack responses. Individuals with higher F1 elicited fewer sack responses than those with lower F1s. When only women were examined, F1 frequency was the only significant predictor of sack responses, accounting for 25.7% of the variance. Women with higher F1 frequencies elicited fewer sack responses than those with lower F1s. No significant predictors were found for male talkers. F2 frequency, vowel duration, and H2-H1 amplitude never accounted for a significant proportion of variance in sack responses in any of the regression analyses.

When the proportion of sip responses at step 5 was examined in the entire group of 44 talkers, f0 accounted for the largest proportion of variance, 62.4%. As with the sack responses, listeners perceived more tokens of sip from talkers with lower f0s than from those with higher f0s. F1 frequency accounted for an additional 9.2% of the variance. Again, talkers with higher F1s elicited fewer sip responses than those with lower F1s. When women were examined, F1 frequency accounted for 45.5% of the variance in sip responses, and f0 accounted for an additional 9.0%. The direction of the effects was the same as in the entire group of 44 subjects. Finally, when the 22 men's voices were examined, F2 frequency accounted for 18.3% of the variance in sip responses. The negative β coefficient shows that men with higher F2 frequencies elicited fewer sip judgments than men with lower F2s. Vowel duration never accounted for a significant proportion of variance in sip judgments in any regression.

2. Perceived sexual orientation, masculinity, and femininity

The second set of regressions examined whether measures of the 44 talkers' perceived sexual orientation [taken from Munson *et al.* (in press)] and perceived masculinity or femininity (collected for the current study) predicted patterns of phoneme identification. Recall that three perceived masculinity/femininity measures were calculated for each subject. The first of these was the average of the four judgment words varying in their phonemic composition, as described in Sec. II A 1. The other two measures were the averages of listeners' ratings based on either the shack or ship token used to create the stimuli in this study. The dependent measure in these regressions was the percentage of sack or sip responses at step 5. Women and men were examined separately.

The first set of regressions examined average sack judgments. The results of these are shown in Table V. For women's voices, all three perceptual measures accounted for a significant proportion of variance in the average percent of sack judgments at step 5. The variance accounted for ranged from 37.2% to 44.8%. In all of these regressions, the standardized β coefficients were positive, indicating that women who were rated as L/B sounding or less feminine elicited more sack percepts than those rated as more-heterosexual or more-feminine sounding. In contrast, only one of the regressions was significant for men's voices. This was the regression predicting average sack judgments from measures of perceived masculinity based on the shack tokens used to create the fricative-identification stimuli. The variance accounted for was 21.9%. The negative β coefficient indicated

TABLE V. Simple regressions predicting the average identification of sack at step 5 from average perceived sexual orientation, average perceived masculinity or femininity, and masculinity or femininity based on the shack base stimulus only (see text for details). Men and women examined separately.

Group	Independent measure	F test	R^2	B	SE B	β
Women	Perceived sexual orientation	$F[1,20]=14.2$, $p<0.001$	0.415[a]	7.73[a]	2.06	0.644[a]
	Perceived femininity	$F[1,20]=16.3$, $p<0.001$	0.448[a]	6.74[a]	1.67	0.670[a]
	Perceived femininity, sack only	$F[1,20]=11.9$, $p<0.001$	0.372[a]	5.27[a]	1.53	0.610[a]
Men	Perceived sexual orientation	$F[1,20]<1$, n.s.	N/A	N/A	N/A	N/A
	Perceived masculinity	$F[1,20]=2.6$, n.s.	N/A	N/A	N/A	N/A
	Perceived masculinity, sack only	$F[1,20]=5.6$, $p<0.05$	0.219[b]	−8.02[b]	3.39	−0.468[b]

[a]$p<0.01$.
[b]$p<0.05$.

TABLE VI. Simple regressions predicting the average identification of sip at step 5 from average perceived sexual orientation, average perceived masculinity or femininity, and masculinity or femininity based on the ship base stimulus only (see text for details). Men and women examined separately.

Group	Independent measure	F test	R^2	B	SE B	β
Women	Perceived sexual orientation	$F[1,20]=7.0$, $p<0.05$	0.258[a]	4.38[a]	1.66	0.508[a]
	Perceived femininity	$F[1,20]=5.2$, $p<0.05$	0.208[a]	3.29[a]	1.44	0.456[a]
	Perceived femininity, sip only	$F[1,20]=6.0$, $p<0.05$	0.232[a]	3.23[a]	1.31	0.482[a]
Men	Perceived sexual orientation	$F[1,20]=4.1$, $p<0.05$	0.149[a]	3.67[a]	1.97	0.386[a]
	Perceived masculinity	$F[1,20]<1$, n.s.	N/A	N/A	N/A	N/A
	Perceived masculinity, sip only	$F[1,20]<1$, n.s.	N/A	N/A	N/A	N/A

[a]$p<0.05$.

that men who were rated as less-masculine sounding elicited fewer /s/ percepts than those were rated as more-masculine sounding, as hypothesized.

The second set of regressions examined average sip judgments at step 5 (Table VI). When women were examined, all three regressions were significant. The variance accounted for ranged from 20.8% to 25.8%. As with the sack regressions, all of the β coefficients were positive, indicating that women who were rated as L/B or less-feminine sounding elicited more sip judgments than those rated to sound more feminine or heterosexual. When men's voices were examined, only one regression was significant. Men's perceived sexual orientation accounted for 14.9% of the variance in sip judgments. The β coefficient for this regression was positive. This indicates that men who were rated as more-G/B sounding elicited more sip judgments than men rated as more-heterosexual sounding, contrary to expectations.

The results in this section suggest that women who were rated in Munson *et al.* (in press) as L/B sounding, and in the current study as less-feminine sounding, elicited more /s/ percepts than women rated to sound heterosexual and feminine. However, there is an alternative explanation for these findings, which arises when considering the results of the regressions in Sec. III B 1. Specifically, listeners may have been making judgments about fricative identity based on their perception of the women talkers' overall stature, rather than on perceived sexual orientation or perceived femininity *per se*. Women elicited more /s/ percepts if the acoustic characteristics of their voices were consistent with having a larger stature—i.e., a lower f0, F1, and F2. Indeed, this finding may explain the regressions presented in this section, as well. Munson *et al.* (in press) found that there was a strong, statistically significant correlation between measures of perceived sexual orientation and measures of perceived height (measured on a five-point equally occurring interval scale) for women's voices. Perceived sexual orientation was not correlated with actual height for women or for men.

To examine this possibility, a final set of regressions was calculated for women's voices only. They assessed whether perceived sexual orientation and perceived femininity predicted a significant proportion of variance in the percentage of sack and sip responses at step 5 when actual and perceived height (as reported in Munson *et al.*, in press) were forced as the first variables in the regression. In all of these regressions, perceived sexual orientation or perceived femininity predicted a significant proportion of variance ($\alpha<0.05$) in sack or sip responses beyond what was accounted for by perceived and actual height. For sack judgments, perceived sexual orientation accounted for 24.1% of additional variance, perceived femininity based on the full set of stimuli accounted for an additional 26.7% of variance, and perceived femininity based on the token sack only accounted for 15.4% of variance. For sip judgments, perceived sexual orientation accounted for 26.1% of additional variance, perceived femininity based on the full set of stimuli accounted for an additional 18.7% of variance, and perceived femininity based on the token sack only accounted for 21.8% of variance. These findings suggest that perception of vocal-tract size alone cannot account for the apparent influence of perceived sexual orientation and perceived femininity on identification of fricatives appended to women's voices.

IV. SUMMARY AND DISCUSSION

A. Summary

The principle findings of the study are summarized below:

(1) As in earlier research (Strand and Johnson, 1996), listeners report more /ʃ/-initial word percepts when a synthetic fricative ambiguous between /s/ and /ʃ/ is appended to a woman's voice than when it is appended to a man's voice.
(2) The effect of talker sex on fricative perception was not mediated by vowel contexts manipulated in this study: it was equally strong for /i/ and /æ/ contexts.
(3) Some evidence was found to support a gradient relationship between talker sex typicality and fricative perception. Women who were rated as lesbian/bisexual sounding elicited more sack and sip percepts than women who were rated as heterosexual sounding. In contrast, there were no differences between gay/bisexual-sounding men and heterosexual-sounding men in the rate of listeners'

sack responses. A separate set of regression analyses examined the acoustic characteristics that might have influenced listeners' fricative-identification patterns. For women, F1 frequency and f0 predicated phoneme identification: women with low F1 frequency and low f0 elicited phoneme-identification patterns that were more malelike (i.e., associated with more /s/-initial word judgments) than women with high F1 frequency and high f0. This was true for both the sack-shack and the sip-ship continua. For men, only one acoustic measure predicted phoneme-identification patterns. On the sip-ship continuum, men with a high F2 frequency elicited fewer sip-percepts than men with lower F2 frequencies; no predictors were found for performance on the sack-shack continuum.

B. Discussion

The most striking finding in the current study is the asymmetry between the identification patterns associated with men and women's voices. An influence of perceived sexual orientation was found for women's voices in the manner predicted, but not for men's voices. One hypothesis is that the range of perceived sexual orientation scores for men was more restricted than that of women, and that the asymmetry reflects this. This conjecture is not supported. Indeed, men demonstrated a *wider* range of perceived sexual orientation and fricative-identification values than women. It is also unlikely that the findings reflect a restriction in range of the acoustic variables that might cue listeners' fricative identifications. The range in F1 values in shack was 2.66 Bark for men's voices, which was larger than the 2.28-Bark range seen in women's voices.

One possible explanation of this seemingly paradoxical asymmetry is provided by a *posthoc* analysis of the acoustic characteristics of fricatives from Munson *et al.* (in press). An analysis of the average first spectral moments of the productions of /s/ and /ʃ/ by GLB- and heterosexual-sounding people in that study shows that the L/B-sounding women produced both fricatives with lower centers of gravity than the heterosexual-sounding women ($F[1,20]=7.3$, $p<0.05$). G/B-sounding men differed from heterosexual-sounding men in spectral skewness of /s/, but not in the first spectral moment of /s/ or /ʃ/ ($F[1,20]<1$, $p>0.05$). That is, L/B-sounding women produced lower F1 values and fricatives with a lower center of gravity. G/B-sounding men produced higher F1 values than heterosexual-sounding men, but *did not* produce fricatives with a higher center of gravity. The listener identification patterns in the current study showed sensitivity to this pattern: when listeners were presented with women's voices in which F1 varied, their fricative identification showed an expectation that fricatives' center frequency would also vary. When they were presented with men's voices in which the F1 frequency varied, their identification of fricatives was not affected, showing tacit knowledge that F1 variation in G/B- and heterosexual-sounding male talkers is not accompanied by a difference in the center of gravity of fricatives. Recall that the synthetic fricatives used in this study varied in their first spectral moment, but did not vary in their skewness. As shown in Table II, while skewness did vary across the nine fricatives, it did not increase monotonically as center frequency increased. Presumably, identification of a synthetic fricative continuum in which skewness did vary would be affected by perceived sexual orientation in men, as this variable was shown by Munson *et al.* (in press) to differ between G/B- and heterosexual-sounding men.

One paradoxical finding in these analyses, as well as in the ANOVAs, was the fact that G/B-sounding men elicited more sip responses in the sip-ship continuum than heterosexual-sounding men. In regression analyses of men's voices, a higher F2 was associated with fewer sip judgments, as we would expect. However, the G/B-sounding men in this study did not produce a higher F2 in their ship base tokens; the average F2 values for the two groups were nearly identical. The specific reason for this finding remains unclear. However, this paradoxical finding provides further support for the conjecture that gender typicality in men's voices does not have the same effect as it does for women's voices, as the G/B-sounding men elicited a phoneme-identification pattern quite unlike that of women.

A growing body of research, including that presented in this study, suggests that a variety of talker-related speech production patterns influences speech perception. The asymmetry between perception of men and women's speech in this study suggests that this perception is influenced by fine-grained knowledge of the co-occurrence between F1 and the fricative center of gravity in men and women's speech. That talker-specific detail influences speech perception is clear. The next logical set of questions in this area concerns how the perception of social variables interacts with other perception process, and the implications of these findings for models of the cognitive architecture that underlies speech perception. Specifically, research should delimit the extent to which perception of socially conditioned speech variants interacts with other variables known to affect perception, such as lexical frequency, prosodic structure, and overall task difficulty, to determine whether existing models of perception need to be modified to accommodate findings like those in this study.

ACKNOWLEDGMENTS

Portions of this research were conducted as part of the second author's *magna cum laude* B.A. thesis from the University of Minnesota Department of Speech-Language-Hearing Sciences, completed in December 2004. The authors thank Mary R. T. Kennedy for extensive input into that project. This research was supported by a Grant-in-Aid of Research, Scholarship, and Artistry from the University of Minnesota Graduate School. The authors thank Nancy DeBoe and Aubrey White for assistance with parts of this research. The authors gratefully acknowledge Keith Johnson and Elizabeth Strand for providing the Klatt Synthesizer parameter files used to create the fricative continua. This research was presented at the Spring 2005 meeting of the Acoustical Society of America. The authors thank conference participants for their extremely useful feedback about this work.

[1]A reviewer pointed out that this experimental design provided the listeners with ample opportunities to implicitly learn the fine acoustic differences among talkers and among the individual tokens of the fricative continuum. Consequently, the results may in part reflect some learning that occurred during the experiment.

Boersma, P., and Weenink, D. (**2003**). *Praat v. 4.1.7 [Computer Software]* (Institute of Phonetic Sciences, Amsterdam).

Childers, D. G., and Wu, K. (**1991**). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. **90**, 1841–1856.

Fant, G. (**1966**). "A note on vocal tract size factors and non-uniform F-pattern scalings," Speech Technology Laboratory: Quarterly Progress and Status Report **4**, 22–30.

Fant, G. (**1975**). "Nonuniform vowel normalization," Speech Technology Laboratory: Quarterly Progress and Status Report **2–3**, 1–19.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**, 1511–1522.

Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data." J. Acoust. Soc. Am. **84**, 115–123.

Gaudio, R. (**1994**). "Sounding gay: Pitch properties in the speech of gay and straight men." Am. Speech **69**, 30–57.

González, J. (**2004**). "Formant frequencies and body size of speaker: a weak relationship in adult humans," J. Phonetics **32**, 277–287.

Hedrick, M., and Ohde, R. (**1993**). "Effect of relative amplitude of frication on perception of place of articulation," J. Acoust. Soc. Am. **94**, 2005–2026.

Hermes, D. J., and van Gestel, J. C. (**1991**). "The frequency scale of speech intonation," J. Acoust. Soc. Am. **90**, 97–102.

Hillenbrand, J., Getty, L., Clark, M., and Wheeler, K.(**1995**). "Acoustic characteristics of American English Vowels." J. Acoust. Soc. Am. **97**, 3099–3111.

Hirano, M., Kurita, J., and Nakahima, T. (**1983**). "Growth, development, and aging of the vocal folds," in *Vocal Fold Physiology: Contemporary Issues and Clinical Research*, edited by D. Bless and P. Abbs (College-Hill, San Diego, CA) pp. 22–43.

Johnson, K. (**1990**). "The role of perceived speaker identity in f0 normalization for vowels," J. Acoust. Soc. Am. **88**, 642–654.

Johnson, K. (**1991**). "Differential effects of speaker and vowel variability on fricative perception," Lang Speech **34**, 265–279.

Johnson, K., Strand, E., and D'Imperio, M. (**1999**). "Auditory-visual integration of talker gender in vowel perception," J. Phonetics **27**, 359–384.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.

Klatt, D. H., and Klatt, L. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Labov, W., Ash, S., and Boberg, C. (in press). *Atlas of North American English* (Mouton de Gruyter, Berlin).

Lass, N. J., Almerino, C. A., Jordan, L. F., and Walsh, J. M. (**1980**). "The effect of filtered speech on speaker race and sex identifications," J. Phonetics **8**, 101–112.

Lass, N. J., Tecca, J. E., Mancuso, R. A., and Black, W. I. (**1979**). "The effect of phonetic complexity on speaker race and sex identifications," J. Phonetics **7**, 105–118.

Linville, S. (**1998**). "Acoustic correlates of perceived versus actual sexual orientation in men's speech," Folia Phoniatr Logop **50**, 35–48.

May, J. (**1976**). "Vocal-tract normalization for /s/ and /ʃ/," Haskins Laboratory: Status Report on Speech Research, SR-48, pp. 67–73.

Munson, B., McDonald, E. C., DeBoe, N. L., and White, A. R. "The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech," J. Phonetics (in press).

Perry, T. L., Ohde, R., and Ashmead, D. (**2001**). "The acoustic bases for gender identification from children's voices," J. Acoust. Soc. Am. **109**, 2988–2998.

Pierrehumbert, J. B., Bent, T., Munson, B., Bradlow, A. R., and Bailey, J. M. (**2004**). "The influence of sexual orientation on vowel production," J. Acoust. Soc. Am. **116**, 1905–1908.

Pisoni, D., Nusbaum, H., Luce, P., and Slowiacek, L. (**1985**). "Speech perception, word recognition, and the structure of the lexicon," Speech Commun. **4**, 75–95.

Schneider, W., Eschman, A., and Zuccolotto, A. (**2002**). *E-Prime User's Guide* (Psychology Software Tools, Inc., Pittsburgh).

Schwartz, M. (**1968**). "Identification of speaker sex from isolated, voiceless fricatives," J. Acoust. Soc. Am. **43**, 1178–1179.

Smyth, R., Jacobs, G., and Rogers, H. (**2003**). "Male voices and perceived sexual orientation: An experimental and theoretical approach," Lang. Soci. **32**, 329–350.

Strand, E. (**1999**). "Uncovering the role of gender stereotypes in speech perception," J. Lang. Social Psychol. **18**, 86–99.

Strand, E., and Johnson, K. (1996). "Gradient and visual speaker normalization in the perception of fricatives," in *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference, Bielfelt*, October 1996, edited by D. Gibbon (Mouton de Gruyter, Berlin), pp. 14–26.

Stuart-Smith, J. (2004). "Empirical Evidence for Gendered Speech Production: /s/ in Glaswegian," presentation given at the 9th Conference on Laboratory Phonology, Urbana-Champaign, IL.

Titze, I. (**1989**). "Physiological and acoustic differences between male and female voices," J. Acoust. Soc. Am. **85**, 1699–1707.

van Bezooijen, R. (**1995**). "Sociocultural aspects of pitch differences between Japanese and Dutch women," Lang Speech **38**, 253–265.

van Dommelen, W. A., and Moxness, B. H. (**1995**). "Acoustic parameters in speaker height and weight identification: sex-specific behaviour." Lang Speech, **38**, 267–287.

Wu, K., and Childers, D. G. (**1991**). "Gender recognition from speech. Part I: Coarse analysis." J. Acoust. Soc. Am. **90**, 1828–1840.

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.

The ability of listeners to use recovered envelope cues from speech fine structure

Gaëtan Gilbert[a] and Christian Lorenzi
Laboratoire de Psychologie de la Perception, CNRS, Ecole Normale Supérieure, DEC, 29 rue d'Ulm, France, and Groupement de Recherche en Audiologie Expérimentale et Clinique, France

(Received 27 June 2005; revised 16 January 2006; accepted 17 January 2006)

Recent work has demonstrated that auditory filters recover temporal-envelope cues from speech fine structure when the former were removed by filtering or distortion. This study extended this work by assessing the contribution of recovered envelope cues to consonant perception as a function of the analysis bandwidth, when vowel-consonant-vowel (VCV) stimuli were processed in order to keep their fine structure only. The envelopes of these stimuli were extracted at the output of a bank of auditory filters and applied to pure tones whose frequency corresponded to the original filters' center frequencies. The resulting stimuli were found to be intelligible when the envelope was extracted from a single, wide analysis band. However, intelligibility decreases from one to eight bands with no further decrease beyond this value, indicating that the recovered envelope cues did not play a major role in consonant perception when the analysis bandwidth was narrower than four times the bandwidth of a normal auditory filter (i.e., number of analysis bands $\geqslant 8$ for frequencies spanning 80 to 8020 Hz). © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2173522]

PACS number(s): 43.71.Gv, 43.71.Es, 43.66.Mk [KWG] Pages: 2438–2444

I. INTRODUCTION

A number of speech perception studies have investigated the role of two temporal features of the speech signal in speech understanding: its slow varying component, referred to as the signal's envelope, and its fast varying component, referred to as the signal's fine structure. Several signal-processing techniques allow the extraction of the signal's fine structure while removing the temporal envelope; among them, infinite compression and the Hilbert transform (which leads to a decomposition of the signal into its envelope and fine structure) are the most commonly used. Based on these techniques, studies conducted by Licklider and Pollack (1948), Drullman *et al.* (1994a, b), and Smith *et al.* (2002) have shown that the removal of the temporal envelope while keeping the fine structure intact does not affect strongly speech intelligibility, when the processing is applied within a single, wide frequency band (also called the "analysis band") or a limited number of broad frequency bands (i.e., six, 1-oct bands). However, when the speech fine structure is extracted within 24 $\frac{1}{4}$-oct frequency bands, Drullman *et al.* (1994a, b) and Drullman (1995) showed that the intelligibility is degraded substantially.

Modelling work by Ghitza (2001) has provided an insight into this apparent discrepancy by demonstrating that the degraded speech-envelope cues may be recovered at the output of auditory filters because the signal's envelope and instantaneous frequency information are related. The notion that the degraded envelope cues may be recovered at the output of auditory filters and used by listeners was recently confirmed by Zeng *et al.* (2004). They conducted a speech identification task where the contribution of the recovered envelope to speech identification was assessed. First, they removed the whole envelope of the sentences using the Hilbert decomposition in a single, wide frequency band (80–8020 Hz). Then, they extracted the "recovered" envelopes of the processed sentences at the output of a bank of gammachirp auditory filters (Irino and Patterson, 1997). They finally used these envelopes to amplitude modulate noise bands having the same bandwidth as the original auditory filters. In agreement with Ghitza (2001)'s predictions, the resulting processed sentences were found to be intelligible (40% mean correct identification).

Zeng *et al.* (2004) pointed out that the envelope recovery process depends upon the ratio between the analysis bandwidth (i.e., the bandwidth of the filters used to analyze the speech stimuli) and the bandwidth of auditory filters. The following simulation illustrates this dependency for elementary stimuli. The test signal was obtained by adding two 100% sinusoidally amplitude-modulated (SAM) tones having equal peak amplitude (65 dB SPL), A and B. A was a 1-kHz tone modulated at 4 Hz and B was a 2-kHz tone modulated at 8 Hz. The starting modulation phase of each SAM tone was chosen at random. The dotted line in Fig. 1 shows the envelope of the test signal at the output of a 1-ERB-wide auditory filter centered at 1 kHz. As expected, this envelope fluctuates at a 4-Hz rate. The solid line shows the envelope recovered from the fine structure of the test signal at the output of this auditory filter centered at 1 kHz when the fine structure of the test signal was extracted[1] from the compound stimulus (i.e., A+B). The dashed line shows the envelope recovered from the fine structure of the test signal at the output of the same auditory filter when the fine structure of the test signal was extracted from each SAM tone (i.e., A and B) taken separately, the two fine structure signals being added thereafter. The second case (referred to

[a]Present address: G. Gilbert. MRC Institute of Hearing Research, Glasgow Royal Infirmary, Queen Elizabeth Building, 16 Alexandra Parade, Glasgow G31 2ER, UK. Electronic mail: gaetan@ihr.gla.ac.uk

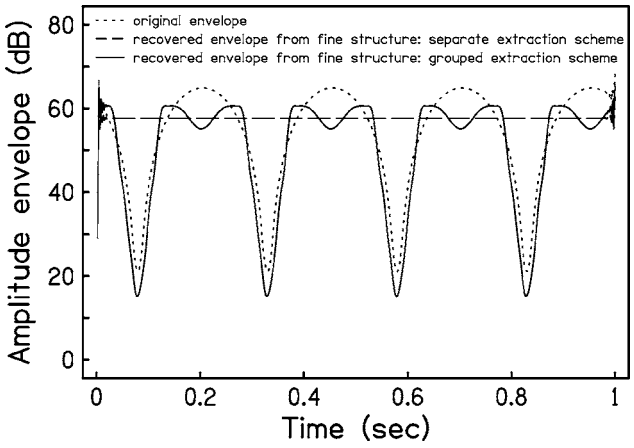


FIG. 1. Comparison between the envelopes obtained at the output of an auditory filter centered at 1 kHz (bandwidth of 1 ERB) for an elementary test signal. The test signal was obtained by adding two 100% sinusoidally amplitude-modulated (SAM) tones with equal peak amplitude (65 dB SPL): A and B. A was a 1-kHz tone modulated at 4 Hz and B was a 2-kHz tone modulated at 8 Hz. Dotted lines show the envelope of the test signal at the output of an auditory filter centered at 1 kHz. The dashed line shows the envelope recovered at the output of the same auditory filter from the fine structure of the test signal, when the fine structure of the test signal was extracted from each SAM tone taken separately, the two fine structure signals being added thereafter ("separated extraction" processing scheme). The solid line shows the envelope recovered at the output of the auditory filter centered at 1 kHz from the fine structure, when the fine structure was extracted from the compound stimulus ("grouped extraction" processing scheme).

as the "separate extraction scheme") corresponded to narrow analysis filtering while the first one (referred to as the "grouped extraction scheme") corresponded to broad analysis filtering. Figure 1 indicates clearly that envelope recovery occurs when the fine structure of the test signal was extracted from the compound stimulus. In other words, the broader the analysis bandwidth, the closer the recovered envelope is to the true envelope of the 1-kHz component. This suggests that when the bandwidth of analysis filters is substantially larger than the bandwidth of auditory filters, the envelope cues in those filters are mostly recovered.

In the current example, envelope recovery (solid line in Fig. 1) originated from an FM-to-AM conversion mechanism. As shown by Hartmann (1998), the beating of two (or more) frequency components creates a frequency modulation (FM) which depends on the relative amplitude of each component of the test signal (i.e., A and B): when the amplitude of tone A dominates over that of tone B, the instantaneous frequency of the test signal is closer to the frequency of tone A and vice versa. Thus, when the fine structure was extracted from the compound stimulus, that is when the analysis filter was broad, the variation over time of the amplitude ratio between the components A and B could be retraced from the FM. Such a FM was then converted into dynamic variations in the level of excitation (i.e., into an AM signal) at the output of auditory filters because the frequency excursion of the FM (within 1 and 2 kHz) was large compared to the bandwidth of auditory filters.

The goal of the present study was to extend the initial work by Zeng *et al.* (2004) by assessing the capacity of listeners to identify speech on the basis of the putative recovered envelope cues. This capacity was evaluated as a function of the analysis bandwidth, when envelope cues were removed by using Hilbert transform. A control identification task was conducted when the envelope was removed by infinite peak clipping.

II. EXPERIMENTS

A. Method

1. Speech material

One set of 48 unprocessed vowel-consonant-vowel (VCV) stimuli was recorded. These speech stimuli consisted of three exemplars of the 16 /aCa/ utterances (C $=/p,t,k,b,d,g,f,s,\int,m,n,r,l,v,z,ʒ/$) read by a French female speaker in quiet (mean VCV duration=648 ms; standard deviation=46 ms). Each signal was digitized via a 16-bit analog/digital converter at a 44.1-kHz sampling frequency.

2. Stimuli in the main identification task

The original speech signals were submitted to two different processing schemes. Stimuli processed using the first scheme (Hilbert fine structure conditions: HFS) contained speech information in their fine structure only, but envelope cues were potentially recoverable at the output of auditory filters. Stimuli were also generated using a second scheme (recovered envelope from Hilbert fine structure conditions: R-HFS) so as to force listeners to identify consonants primarily on the basis of the recovered envelope cues.

a. HFS conditions Each VCV signal was initially band-pass filtered using Butterworth filters (62 dB/oct rolloff) into 1, 2, 4, 8, or 16 complementary frequency bands (analysis bands) spanning the range 80–8020 Hz and following a logarithmic spacing inside this nominal bandwidth. Table I relates the number of frequency bands to their respective cutoff frequencies, and Table II relates the number of frequency bands to their bandwidths (in ERB units and in Hz). Forward and backward filtering were used to cancel phase delays. The Hilbert transform was then applied in each

TABLE I. Relationship between the number of frequency bands (i.e., analysis bands) and their respective cutoff frequencies.

No. of bands	Cutoff frequencies in Hz																
1	80																8020
2	80								801								8020
4	80				253				801				2535				8020
8	80		142		253		450		801		1425		2535		4509		8020
16	80	107	142	190	253	338	450	601	801	1068	1425	1900	2535	3380	4509	6013	8020

TABLE II. Bandwidth in ERB units as a function of the number of bands used in this study. The related bandwidth in Hertz is shown in italics below the value in ERB units.

No. of bands	Bandwidth of each band (in ERB units and *in Hz*) band number																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	Average
1	30.5																30.5
	7940																
2	11.2	19.3															15.3
	721	*7219*															
4	4.1	7.1	9.2	10.1													7.6
	171	*548*	*1734*	*5845*													
8	1.7	2.1	3.2	3.9	4.4	4.8	5.0	5.1									3.8
	62	*111*	*197*	*351*	*624*	*1110*	*1974*	*3511*									
16	0.8	0.9	1.1	1.3	1.5	1.7	1.9	2.0	2.1	2.3	2.3	2.5	2.4	2.5	2.5	2.6	1.9
	27	*35*	*48*	*63*	*85*	*112*	*151*	*200*	*267*	*357*	*475*	*635*	*845*	*1129*	*1504*	*2007*	

band in order to decompose the VCV signal into its envelope (module of the Hilbert analytic signal) and temporal fine structure (cosine of the argument of the Hilbert analytic signal). The envelope component was discarded. The fine structure was multiplied by the root mean square (rms) power of the band-pass filtered VCV, in order to compensate for the reduction in amplitude caused by envelope removal. The "power-weighted" fine structure signals were finally summed over all frequency bands and presented as such to the listeners.

b. R-HFS conditions The HFS signals were passed through a bank of 30 gammachirp auditory filters, each 1 ERB wide (Irino and Patterson, 1997) with center frequencies ranging from 123 to 7743 Hz, and spaced along an ERB scale. In each band, the temporal envelopes were extracted using the Hilbert transform and low-pass filtered (cutoff frequency=64 Hz, 62 dB/oct rolloff) using a Butterworth filter (again forward and backward filtering were used). These envelopes were then used to amplitude modulate sine waves having the same frequencies as the original center frequencies of the auditory filters, but with random starting phase.

Figure 2 shows the correlation between the original and the recovered envelopes. The filled symbols show the mean correlation coefficients computed across the 48 VCV utterances between the speech envelopes of the original VCV stimuli and the envelopes of the HFS stimuli at the output of six auditory filters, as a function of the number of analysis bands.[2] The values were averaged across the 48 VCV utterances. A high correlation coefficient means that there was a close resemblance between the original envelope and that recovered at the output of auditory filters.

As predicted by Ghitza (2001), envelope cues were recovered from fine structure information at the output of auditory filters when the processing used to extract the fine structure was applied within a single, wide frequency band (filled circles). Figure 2 also shows that, overall, increasing the number of bands from one to eight (i.e., decreasing the bandwidth of each band) had a detrimental effect on envelope recovery. It is noteworthy that the envelope was better recovered for the auditory filter centered at 1197 Hz and that the detrimental effect of increasing the number of bands was especially large for this auditory filter. The upper panel of Fig. 3 shows that the average excitation pattern of the VCVs used in this study peaked in the frequency region around 1000 Hz (the frequency region around 1000 Hz encompassed the energy of the first and second formant of the vowel /a/ used in this study). The lower panel of Fig. 3 shows the amount of envelope recovery for the stimuli used in the single-band HFS condition (as in Fig. 2, but with a greater frequency resolution). It is quite clear from Fig. 3 that envelope recovery peaked in auditory filters where the amount of excitation is highest. Additionally, Fig. 2 shows slight negative correlation coefficients in auditory filters adjacent to the auditory filters conveying most of stimulus' energy. In auditory filters whose output is weak, the FM-to-AM conversion mechanism (responsible for envelope recovery) recovered mostly the true amplitude envelope from the auditory filters whose output is strongest; however, this recovered envelope is in antiphase with the true envelope. Knowing that the amplitude-envelopes of speech stimuli are, to some extent, positively correlated across the whole spec-

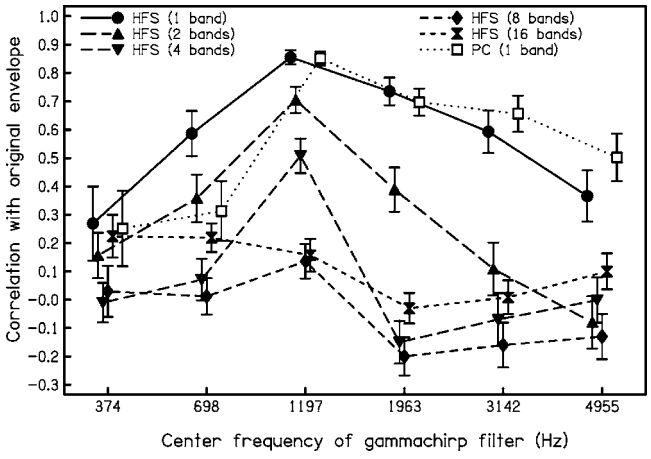


FIG. 2. Effect of the number of frequency bands (i.e., analysis bands) on the mean correlation coefficients computed (across the 48 VCV utterances) between the original speech envelopes and the envelopes of the stimuli in the Hilbert fine structure (HFS) conditions, at the output of six gammachirp auditory filters (filled symbols). Hilbert decomposition was used here to remove envelope cues while keeping fine structure intact. Open squares show the mean correlation coefficients computed between the original speech envelopes and the envelopes of stimuli in the peak clipping (PC) at the output of six gammachirp auditory filters, when infinite peak clipping in a single, wide frequency band is used to remove the temporal envelope. Error bars show ± one 95% confidence interval.

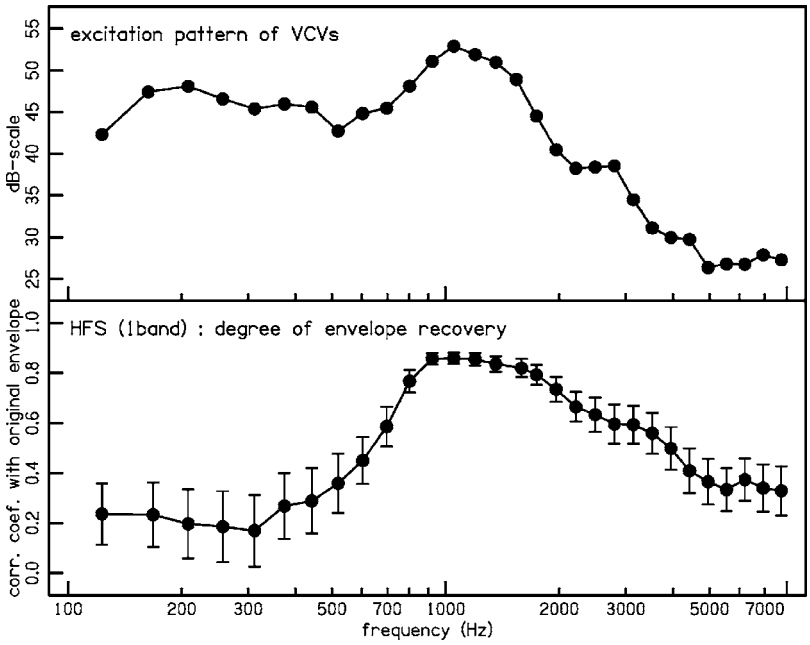


FIG. 3. The upper panel shows the excitation pattern averaged across the 48 VCV used in this study. Note the peak around 1 kHz, due to the first and second formant of the vowel /a/ used in this study. The lower panel shows the mean correlation coefficients computed between the original speech envelopes and the envelopes of the stimuli in the Hilbert fine structure (HFS) (1 band) condition at the output of 30 gammachirp auditory filters. The error bars in the lower panel represent ± one 95% confidence intervals. Note that the amount of envelope recovery is also peaking around 1 kHz.

trum (Crouzet and Ainsworth, 2001), this might explain the observation of those negative correlation coefficients.

3. Stimuli in the control identification task

An additional identification task was devised in order to assess to what extent the findings could be replicated when envelope cues were degraded using a different processing scheme. Indeed, in the case of *broadband* stimuli, the module and the argument of the analytic signal obtained by the Hilbert transform do not respectively correspond strictly to the envelope and instantaneous phase of the signal. Thus, fine structure information obtained by means of the Hilbert transform might be corrupted by the envelope information and vice versa (Ghitza, 2001). Said differently, the use of the Hilbert transform to demodulate broadband signals might be inadequate. In this additional identification task, infinite peak clipping in a single, wide frequency band was used to remove envelope cues from VCV signals (as in Licklider and Pollack, 1948). More precisely, each original signal was initially band-pass filtered (62 dB/oct rolloff, Butterworth filter, backward and forward filtering) between 80 and 8020 Hz. Infinite peak clipping was applied to each bandpass-filtered signal by replacing all positive amplitudes by +1 and all negative amplitudes by −1. Each clipped signal was then multiplied by the rms power of the VCV signal in that band (to compensate for the reduction in amplitude due to envelope removal). The resulting clipped stimuli were either presented as such (peak clipping condition: PC) or submitted to the second processing scheme (recovered envelope from peak clipping condition: R-PC), so as to force listeners to identify consonants primarily on the basis of the recovered envelope cues.

The open squares in Fig. 2 show the mean correlation coefficients computed across the 48 VCVs between the speech envelopes of the original VCVs and the envelopes of the infinite peak-clipped stimuli at the output of six auditory filters.[2] The correlation coefficients were similar to those obtained previously in the HFS single band condition, again indicating that envelope cues were recovered at the output of auditory filters. This reveals that the degree of envelope recovery reported in the broadband HFS conditions was not mainly due to an improper use of the Hilbert transform.

4. Procedure

All stimuli were generated using a 16-bit digital/analog converter operating at a sampling frequency of 44.1 kHz and delivered diotically via Sennheiser HD 580 earphones at an average level of 70 dB(A). The rms values of the stimuli were equalized. Listeners were tested individually in a sound-attenuating booth. In a typical experimental session, four complete and identical sets of the 48 VCV utterances corresponding to a given experimental condition (i.e., a given set of processed stimuli) were presented at random. Each listener was instructed to identify the presented consonant. The 16 possible choices were presented on the screen of the computer, and the listener entered his/her response by selecting a VCV on the screen with a computer mouse. No feedback was given to the listeners. The percentage of correct identification was calculated and a confusion matrix was built from the 192 VCV (4×48) utterances for a given set of stimuli.

All listeners were given at least 2 h of practice prior to data collection. In the main identification task, all sets of stimuli [2 experimental conditions (HFS vs. R-HFS) ×5 different number of bands (1 to 16 bands)] were presented in random order across listeners. In the control identification task, the two set of stimuli [two experimental conditions (PC vs. R-PC)] were presented in random order across listeners.

5. Listeners

A first group of five listeners participated in the main identification task. Their ages ranged from 20 to 36 years (mean age: 26 years; standard deviation: 6 years). A second group of five different listeners participated in the control

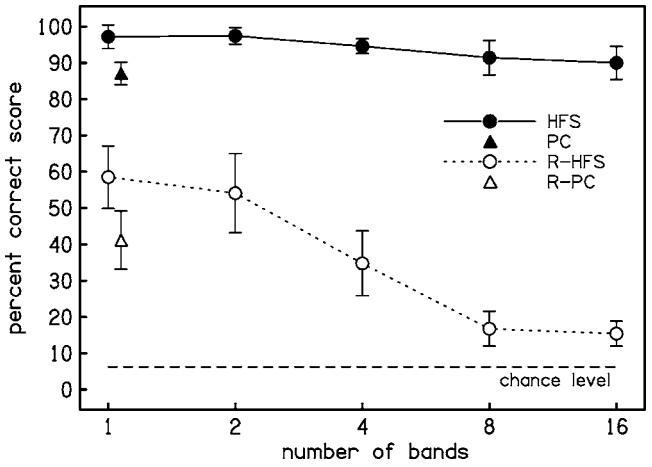


FIG. 4. Mean identification scores across listeners obtained in the Hilbert fine structure (HFS) (filled circles with continuous lines) and recovered envelopes from Hilbert fine structure (R-HFS) (open symbols with dotted lines) conditions. Hilbert decomposition was used here to remove envelope cues, while keeping fine structure intact. Data are plotted along with the mean data obtained in the peak clipping (PC) (filled triangle) and recovered envelopes from peak clipping (R-PC) (open triangle) conditions, where infinite peak clipping was used to remove envelope cues. Error bars show ± one 95% confidence interval. The dotted line shows chance level.

identification task. Their ages ranged from 20 to 22 years (mean age: 21 years; standard deviation: 1 year). All participants were native French speakers. They had audiometric pure tone thresholds less than 20 dB HL between 0.25 and 8 kHz and no history of hearing difficulty.

B. Results

1. Main identification task

Figure 4 shows the mean identification scores across listeners obtained in the HFS (filled circles with continuous lines) and R-HFS (open symbols with dotted lines) conditions. For all conditions, chance level corresponded to 6.25% correct.

The results showed that listeners identified perfectly or nearly perfectly consonants in the five HFS conditions (i.e., for 1 to 16 bands) with only a slight trend for performance to decrease (8%) as the number of bands increased from 2 to 16. The high scores observed in the present study contrast with the lower (70%–80%) scores obtained by Zeng *et al.* (2004) and Smith *et al.* (2002) using one-band, Hilbert fine structure sentences. This difference might be due to methodological differences between the three studies: in the present study, listeners received training before data collection and a closed-set format was used, whereas an open-set format was used in the two previous studies.

The identification scores measured in the R-HFS conditions were poorer than in the HFS conditions. They reached about 60% correct in the one- and two-band R-HFS conditions. This score was slightly higher than the 40% reported in a similar broadband condition by Zeng *et al.* (2004). This difference might be due to the same methodological differences listed above. This result indicates that the contribution of the recovered envelope cues to intelligibility in HFS conditions cannot be systematically neglected.

TABLE III. Phonetic features of the 16 French consonants used in this study (Martin, 1996).

Consonant	Voicing	Manner of articulation	Place of articulation
/p/	unoviced	occlusive	front
/t/	unvoiced	occlusive	middle
/k/	unvoiced	occlusive	back
/b/	voiced	occlusive	front
/d/	voiced	occlusive	middle
/g/	voiced	occlusive	back
/f/	unvoiced	constrictive	front
/s/	unvoiced	constrictive	middle
/ʃ/	unvoiced	constrictive	back
/v/	voiced	constrictive	front
/z/	voiced	constrictive	middle
/ʒ/	voiced	constrictive	back
/l/	voiced	constrictive	middle
/r/	voiced	constrictive	back
/m/	voiced	occlusive	middle
/n/	voiced	occlusive	front

However, Fig. 4 also shows that performance dropped significantly when the number of bands increased. A repeated measures analysis of variance (ANOVA) with factors number-of-bands and processing type confirmed this observation [main effect of factor number of bands: $F(4,16) = 32.12$, $p < 0.001$] [the percent correct identification scores were transformed into rationalized arcsine units (Studebaker, 1985) prior to the statistical analysis]. Nevertheless, this effect was much larger in the R-HFS condition than in the HFS condition [as shown by a significant interaction between factors processing condition and number of bands: $F(4,16) = 16.49$, $p < 0.001$]. Indeed, consonant identification in the R-HFS condition reached a minimum of about 15% in the 8–16 bands. Thus, a substantial contribution of recovered envelope cues to consonant identification was unlikely in the HFS condition when the stimuli were generated using 8 or 16 frequency bands.

Table II indicates that the bandwidth of these eight frequency bands varied from 1.7 ERB units for the lowest band to 5.1 ERB units for the highest band; the average bandwidth was 3.8 ERB units. This suggests that the ability to use recovered envelope cues for consonant identification was essentially abolished when the average analysis filters' bandwidth was less than approximately four times the bandwidth of a normal auditory filter.

The specific reception of the individual phonetic features of voicing (voiced versus unvoiced), manner (occlusive versus constrictive), and place (front versus middle versus back) was evaluated by an information-transmission analysis (Miller and Nicely, 1955) on the individual confusion matrices (see Table III for the assignment of the consonant features). The results of this analysis are presented in Fig. 5. The leftmost, middle, and rightmost panels show the results for the reception of voicing, manner, and place of articulation, respectively. In each panel, the filled circles with continuous lines and open circles with dotted lines correspond to the HFS and R-HFS conditions, respectively. The results show that the relatively good identification performance ob-

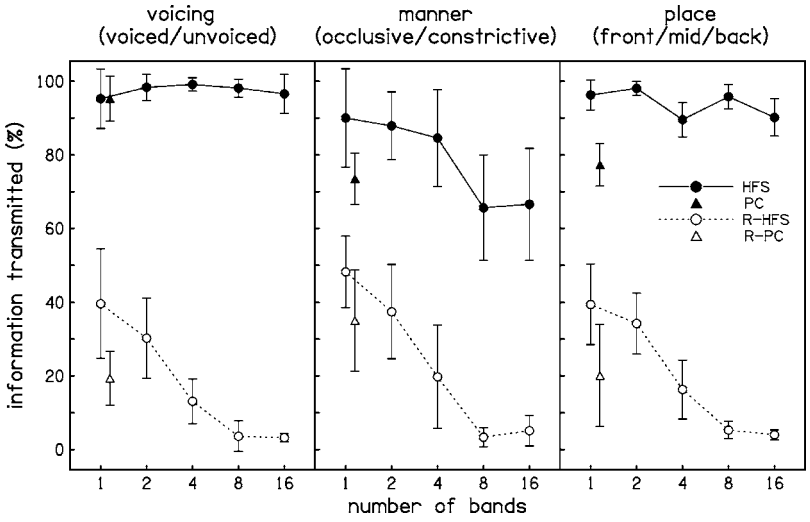


FIG. 5. Specific reception of three phonetic features as a function of the number of frequency bands. The rightmost, middle, and leftmost panels show the results obtained for the reception of voicing, manner, and place of articulation, respectively. In each panel, the filled circles with continuous lines and open circles with dotted lines correspond to the Hilbert fine structure (HFS) and recovered envelope from fine structure (R-HFS) conditions, respectively. Hilbert decomposition was used here to remove envelope cues, while keeping fine structure intact. Data are plotted along with the mean data obtained in the peak clipping (PC) (filled triangle) and recovered envelopes from peak clipping (R-PC) (open triangle) conditions where infinite peak clipping was used to remove envelope cues. Error bars show ± one 95% confidence interval.

served in the one- and two-band R-HFS conditions was due to a substantial reception of all three phonetic features (approximately 50% for the reception of manner, 40% for the reception of voicing and place in the single-band condition). Figure 5 shows that reception of voicing and place decreases when the number of bands increased from 1 to 16 in the R-HFS condition only. This is consistent with the observed significant interactions between the repeated factors number-of-bands and processing condition from two different repeated measures ANOVAs [in the case of voicing: $F(4,16)=28.91$; $p<0.001$; in the case of place: $F(4,16)=24.14$; $p<0.001$]. In contrast, reception of manner decreased when the number of bands increases from 1 to 16 in both experimental conditions (HFS and R-HFS), as also shown by the absence of interaction between the repeated factors number-of-bands and processing condition, from a repeated measures ANOVA performed on the transmitted information percentages in the case of manner [$F(4,16)=1.91$ $p=0.16$]. Thus, the slight decrease in identification performance with the frequency bandwith increasing observed in the HFS condition (cf. Fig. 4), was mainly caused by a decrease in the information transmitted by the manner of articulation. Moreover, the marked decrease in identification performance observed in the R-HFS condition (cf. Fig. 4) was caused by a decrease in the reception of all three phonetic features.

Taken together, the results presented in Figs. 4 and 5 demonstrate that the recovered envelope cues available at the output of auditory filters can contribute to consonant identification when the number of frequency bands used to extract the fine structure is four or less (i.e., when the average bandwidth of the analysis filters is greater than or equal to approximately 8 ERB units). However, when speech fine structure is extracted within eight bands or more (i.e., when the average bandwidth of the analysis filters is approximately smaller than or equal to 4 ERB units), recovered envelope cues are essentially abolished. These data also demonstrate that an *effective* removal of envelope and recovered envelope cues (i.e., when the number of bands is $\geqslant 8$ bands in HFS conditions) affects the transmission of manner, but not the transmission of voicing and place of articulation. This is consistent with previous work showing that nearly perfect reception of manner and poor reception of place are obtained when using mainly temporal envelope cues (e.g., Shannon *et al.*, 1995). However, this study also showed that nearly perfect reception of voicing can be obtained using mainly temporal envelope cues. This suggests that listeners can use either envelope or fine structure cues to reach nearly perfect reception of voicing; however, they must use essentially envelope cues to reach nearly perfect reception of manner and essentially fine structure cues to reach nearly perfect reception of place.

2. Control identification task

In the control identification task, infinite peak clipping was used to remove the amplitude envelope. This experiment was conducted in order to test whether or not the recovery of envelope cues in the broadband R-HFS conditions resulted from an inadequate use of the Hilbert transform to demodulate broadband signals (the Hilbert transform being, in theory, only applicable to demodulate signals that can be approximated as narrow band).

The mean identification scores across listeners obtained in the PC (filled triangle) and R-PC (open triangle) conditions are plotted in Fig. 4 along with the data of the first identification task. Overall, the data obtained with the single, wide band peak-clipped stimuli indicate that, as in Licklider and Pollack (1948), listeners identify nearly perfectly consonants (87% correct). Identification scores obtained in the R-PC condition (41% correct) were poorer than those measured in the PC condition. However, these scores were always substantially above chance level and comparable to those obtained in the main identification task using Hilbert decomposition. As shown in Fig. 5, the identification performance observed in the R-PC condition was, again, due to a substantial reception of all three phonetic features (approximately 35% for the reception of manner and 20% for the reception of voicing and place). This result confirms that the contribution of the recovered envelope cues to intelligibility, in conditions where speech items are processed within a single wide frequency band, cannot be neglected.

III. SUMMARY AND CONCLUSIONS

Taken together, the results indicate the following.

(1) Nearly perfect consonant identification can be obtained on the basis of speech fine structure cues extracted within 8 or 16 frequency bands. These data obtained with French material support the result of a recent study conducted by Xu and Pfingst (2003) with Mandarin material suggesting that speech identification in quiet does not rely entirely on temporal envelope cues.
(2) Consistent with Zeng *et al.*'s (2004) data, envelope cues recovered at the output of auditory filters can be used by listeners to identify consonants. When speech items are subjected to Hilbert decomposition or peak clipping within a single, wide frequency band so as to remove the amplitude envelope while keeping intact the speech fine structure, such a contribution is liable. This should be taken into account when interpreting the results obtained by Licklider and Pollack (1948), Drullman *et al.* (1994a, b), and Smith *et al.* (2002) in similar experimental conditions.
(3) The contribution of the recovered envelope cues to consonant identification is essentially abolished when the amplitude envelope of speech items is removed using Hilbert decomposition within analysis bands whose bandwidth is, on average, lower than or equal to four times the bandwidths of normal auditory filters, i.e., when the number of bands is equal to eight or more for frequencies ranging from 80 to 8020 Hz. This suggests that envelope recovery is unlikely to have significantly affected the results obtained by Drullman (1995), Drullman *et al.* (1994a, b), and Smith *et al.* (2002) when the fine structure of speech was extracted within eight frequency bands or more.

These results may help to process speech material in future studies attempting to compare speech identification based on envelope versus fine structure cues.

ACKNOWLEDGMENTS

This research was supported by a grant from ENTENDRE (GRAEC) to the first author, and a grant from the Institut Universitaire de France to C. Lorenzi. The authors wish to thank M. A. Akeroyd, K. W. Grant, F-G. Zeng, and one anonymous reviewer for valuable comments on preliminary versions of this manuscript. The authors are also thankful to I. Bergeras, D. Voillery, and C. Pinabiaux for their help in data collection.

[1]The fine structure of the tests signals corresponded to the cosines of the Hilbert analytic signals' arguments, multiplied by the original root mean square power of the test signals. Such a multiplication was required to correct for the reduction in amplitude due to envelope removal. The auditory filter used in the simulation was a gammachirp filter (Irino and Patterson, 1997).

[2]The envelopes were extracted using the Hilbert decomposition at the output of the gammachirp auditory filters. The resulting envelopes were passed forward and backward through a Butterworth low-pass filter (cutoff frequency=64 Hz, rolloff=62 dB/oct). A logarithmic transformation was applied before determining the correlation coefficient.

Crouzet, O., and Ainsworth, W. A. (**2001**). "On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation," Workshop on Consistent and Reliable Cues for Sound Analysis, Aalborg, Denmark, September.

Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust. Soc. Am. **110**, 1628–1640.

Hartmann, W. (**1998**). *Signals, Sound, and Sensation* (Springer-Verlag, New York), Chap. 17, pp. 393–398.

Irino, T., and Patterson, R. D. (**1997**). "A time-domain level dependent auditory filter: The gammachirp," J. Acoust. Soc. Am. **101**, 412–419.

Licklider, J. C. R., and Pollack, I. (**1948**). "Effects of differentiation, integration and infinite peak clipping upon the intelligibility of speech," J. Acoust. Soc. Am. **20**, 42–51.

Martin, P. (**1996**). *Éléments de phonétique, avec application au français* (Les Presses de l'Université Laval, Sainte-Foy).

Miller, G. A., and Nicely, P. E. (**1955**). "Analysis of perceptual confusions among some english consonants," J. Acoust. Soc. Am. **27**, 338–352.

Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.

Studebaker, G. A. (**1985**). "A rationalized arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Xu, L., and Pfingst, B. E. (**2003**). "Relative importance of temporal envelope and fine structure in lexical-tone perception (L)," J. Acoust. Soc. Am. **114**, 3024–3027.

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.-Y., and Chen, H. (**2004**). "On the dichotomy in auditory perception between temporal envelope and fine structure cues," J. Acoust. Soc. Am. **116**, 1351–1354.

Effect of masker type on native and non-native consonant perception in noise

M. L. Garcia Lecumberri[a)]
Department of English Philology, University of the Basque Country, Paseo de la Universidad 5, 01006, Vitoria, Spain

Martin Cooke[b)]
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield, S1 4DP, United Kingdom

(Received 24 August 2005; revised 4 February 2006; accepted 6 February 2006)

Spoken communication in a non-native language is especially difficult in the presence of noise. This study compared English and Spanish listeners' perceptions of English intervocalic consonants as a function of masker type. Three maskers (stationary noise, multitalker babble, and competing speech) provided varying amounts of energetic and informational masking. Competing English and Spanish speech maskers were used to examine the effect of masker language. Non-native performance fell short of that of native listeners in quiet, but a larger performance differential was found for all masking conditions. Both groups performed better in competing speech than in stationary noise, and both suffered most in babble. Since babble is a less effective energetic masker than stationary noise, these results suggest that non-native listeners are more adversely affected by both energetic and informational masking. A strong correlation was found between non-native performance in quiet and degree of deterioration in noise, suggesting that non-native phonetic category learning can be fragile. A small effect of language background was evident: English listeners performed better when the competing speech was Spanish.
 [DOI: 10.1121/1.2180210]

PACS number(s): 43.71.Hw, 43.71.Es, 43.66.Dc [ARB] Pages: 2445–2454

I. INTRODUCTION

Spoken communication in noise presents problems for all listeners, but is especially difficult in a foreign language (FL[1]). Several features distinguish native and non-native experience of a given language. Differences in the degree, type, quality, and time of exposure to the language inevitably result in less familiarity with linguistic patterning at all levels, from acoustic to pragmatic. Non-natives also have to deal with the possibility of interference from their first language (L1). In the case of phonological acquisition, L1 influences have been shown to be particularly strong and pervasive (Ioup, 1984; Leather and James, 1991; Polka, 1995). The relative contribution made by these and other factors to foreign language perception in noise is not well understood.

A number of studies have compared native and non-native speech perception performance in noise and reverberation (Florentine *et al.*, 1984; Takata and Nábělek, 1990; Mayo *et al.*, 1997; Hazan and Simpson, 2000; van Wijngaarden *et al.*, 2002; Bradlow and Bent, 2002; Cutler *et al.*, 2004; van Wijngaarden *et al.*, 2004). These studies differed in the speech and noise material employed, the languages tested and the range of language proficiencies of the participants, but three basic findings have emerged. First, native performance on speech in noise tasks exceeded that of non-natives, even for a bilingual-since-infancy group (Mayo *et al.*, 1997). Second, increasing FL experience correlated well with a reduced effectiveness of masking noise (Florentine *et al.*, 1984; Mayo *et al.*, 1997). Third, non-native listeners were less able to take advantage of linguistic context to decode speech presented in noise (Mayo *et al.*, 1997; van Wijngaarden *et al.*, 2004).

Many studies (Florentine *et al.*, 1984; Mayo *et al.*, 1997; Takata and Nábělek, 1990) have suggested that the effect of noise is greater for non-native than for native listeners. This additional native advantage has also been found in studies of easy versus hard word recognition (Bradlow and Pisoni, 1999; Imai *et al.*, 2005). However, two recent studies found that non-native listeners were not more adversely affected by noise (Bradlow and Bent, 2002; Cutler *et al.*, 2004). Bradlow and Bent (2002) compared native and non-native performance on a keyword identification task using sentences mixed with white noise at signal to noise ratios (SNRs) of −4 and −8 dB. The fall in performance between these SNRs was similar for natives and non-natives. However, Bradlow and Bent (2002) acknowledge that this might have been due to a floor effect for non-natives at the lower SNR. Cutler *et al.* (2004) compared the abilities of native and non-native listeners in identifying consonants and vowels in VC and CV syllables presented in babble noise at SNRs of 16, 8, and 0 dB. They found no interaction between language background and noise level: the native advantage in performance was approximately the same for each noise level.

[a)]Electronic mail: garcia.lecumberri@ehu.es; Tel: +34 945013967; Fax: +34 945013200.
[b)]Electronic mail: m.cooke@dcs.shef.ac.uk; Tel: +44 114 2221822; Fax: +44 114 2221810.

The primary purpose of the current study was to examine the effect of different kinds of masker on native and non-native speech perception. Speech perception in noise is governed by both energetic and informational masking. The former arises from the masking of stimulus components at the auditory periphery and produces uncertainty or complete loss of information about the level of the target signal in spectro-temporal regions where the masker is sufficiently intense. Sufficient redundancy exists in clean speech to allow robust identification in many masking conditions (Assmann and Summerfield, 2004) based on available "glimpses" of the target signal (Cooke, 2006). Informational masking (Carhart *et al.*, 1969) refers to the potentially distracting effect of the masker. Intelligibility may suffer if attentional resources are directed at processing the masker, or if the allocation of signal energy to the foreground or masker is unclear.

The role played by energetic and informational masking in non-native speech perception has received little attention. The reduction in acoustic information associated with energetic masking may expose deficiencies in non-native mental representations built from a more limited or less optimal exposure to speech signals. Attentional overload may have a disproportionate influence on non-native perceptual processing which is already taxed by the difficulties of listening to non-native sounds and processing higher level morphological, syntactic, semantic, and pragmatic structures in the FL (Bradlow and Bent, 2002; Cutler *et al.*, 2004). In addition, interference can arise from the non-native listeners' L1 sound system, which may be more activated when listening conditions render the task more difficult. Indeed, Cutler *et al.* (2004) suggest that L1 categories may exert greater influence when the FL stimuli are more difficult to perceive. Similarly, Mayo *et al.* (1997) proposed that for speech material with low predictability, competition between the L1 and FL sound systems may contribute to the added difficulty of non-native perception in noise.

Previous studies of non-native speech perception in noise have employed either steady-state or babble maskers. While N-talker babble can be an effective informational masker even for large N for tasks involving nonsense syllables (Simpson and Cooke, 2005), strong informational masking effects are typically observed when the masker consists of other speech material from one or more competing talkers (Brungart *et al.*, 2001). Non-native speech perception in the face of competing speech has not been investigated to date. The use of competing speech, babble, and stationary noise in the current study permits a comparison of native and non-native performance in differing degrees of energetic and informational masking. The presence of speech in the background in FL perception also raises the possibility of differential language-dependent informational masking: it may be easier for listeners to ignore competing foreign language speech.

Since noisy conditions present added difficulties for all listeners, they constitute a good test for the investigation of native/non-native confusion patterns. It has been suggested that differences between native and non-native phonological performance are a reflection of non-native listeners' less well-developed phonetic categories caused by such factors as L1 interference and differences in the quantity, quality, and onset of FL exposure. At various stages of acquisition, non-native listeners' FL categories will be based on the influence of categories and cues from the FL and L1 to differing degrees. Two models, the Perceptual Assimilation Model (PAM) (Best, 1995) and the Speech Learning Model (Flege, 1995) propose that in FL sound perception, the influence of L1 phonetic categories can largely account for non-native listeners' identifications. PAM takes into account the degree of perceived similarity (exemplar rating) of FL sounds to NL categories. Thus, FL sounds may be considered "uncategorizable" within the listeners' L1 space or even regarded as "non-speech sounds." However, if the FL sounds are assimilated to some L1 sound, they may be seen as a "good" exemplar of the L1 category, or as either an "acceptable" or "deviant" exemplar. The present study examined native versus non-native confusion patterns in different listening conditions to investigate whether masking produces the same confusion patterns independent of a listener's L1 background.

Native speaker competence in tasks such as consonant identification may be expected to be fairly uniform. However, the diversity of FL competences is a problem that permeates all FL research and especially that concerning sound acquisition. It is well known that pronunciation may develop quite separately from other language skills (the *Joseph Conrad* phenomenon; Scovel, 1969). Bradlow and Bent (2002) mentioned that their non-native listeners were highly proficient in written English but were self-reported to have problems with their aural and oral skills. As learners' phonological competence improves, the distance between their developing FL system and that of native speakers' decreases, so they may perform increasingly like native speakers. A recent study by Imai *et al.* (2005) in a task involving natively spoken English word recognition found that the performance of Spanish speaking learners of English with a high phonological proficiency was more similar to native listeners than to low proficiency non-native listeners. In the present study, we were interested in within-group differences in non-native listeners concerning their performance in quiet and their degradation in identification rates in the presence of noise maskers. If non-native listeners' emergent FL sound system is robust, more competent listeners in quiet might be expected to show a smaller intelligibility reduction in masking conditions than listeners at a lower stage of FL phonological acquisition.

This study used English and Spanish listener groups to compare native and non-native perception of intervocalic consonants in quiet and in four noise maskers. Section II describes the listener groups, speech and masker corpus and experimental procedure. It also summarizes the main differences between the English and Spanish phonetic systems. Section III presents the results of native and non-native consonant identification in the four masking conditions.

II. METHODS

A. Participants

Twenty-one native speakers of British English and 61 native speakers of (European) Spanish participated in the

study. The English group was composed of monolingual students at the University of Sheffield whose age ranged from 18 to 24 years (mean: 21.4 years). English students were paid for their participation. The Spanish group consisted of students at the University of the Basque Country studying English as a foreign language (age range:20–25, mean: 21.2 years). They were enrolled in a one semester course in English Phonetics in the second year of a four year B.A. degree in English Language and Literature. As is typical in such courses, competence in English within the non-native group was not uniform, but all had attained the level of the Cambridge Advanced Exam. Spanish students received course credit for participating in the listening tests.

English listeners were screened for hearing loss (better than 20 dB hearing level in the range 250–8000 Hz) while Spanish listeners were asked to report hearing loss. One Spanish participant was excluded after reporting some hearing loss. Another Spanish listener left the course before completing all the tests and was also excluded. Results are presented for the remaining 59 non-native listeners.

Many of the listeners were native or second language speakers of Basque. However, the Spanish and Basque sound systems are so similar that this was not considered to be a relevant variable: for the English consonants in intervocalic context employed in the present study, the only difference between Spanish and Basque is that the latter has a voiceless palato-alveolar phoneme /ʃ/. However, non-Basque speakers are familiar with this sound due to language contact, since it appears in some widely-known words such as "kaixo" /kaiʃo/ ("hello") and "xabi" /ʃabi/ (a common forename).

B. Differences between the English and Spanish phonetic systems

The English and Spanish consonant systems mainly differ in the following respects. Both languages have six plosives arranged in three voiced/voiceless pairs. However, Spanish intervocalic voiced plosives are lenited to approximants and have voicing lead, whereas English voiced plosives have voicing lag or are devoiced (Navarro Tomás, 1918; Harris, 1969). English voiceless plosives are often aspirated whereas in Spanish they are unaspirated in most accents. Unlike English alveolar plosives, Spanish /t d/ have a lamino-dental articulation.

The main difference for fricatives and affricates is the absence of voiced fricative phonemes in most varieties of Spanish. Whereas English has /f v θ ð s z ʃ ʒ tʃ dʒ h/, standard peninsular Spanish has /f θ s tʃ x/, although in many southern Spanish and Latin American accents /θ/ is absent and /s/ is used in its place. The sound /ʝ/ is variably classified as a weak fricative, an approximant or a realization of /i/ (Hualde, 2005). The other main difference between the two languages is that the [+back] fricative is velar in Spanish and glottal in English. The place of articulation of Spanish /s/ varies substantially across regional varieties. In our listeners' accent (Northern-Central Spain) /s/ is realized as an apico-alveolar whereas English /s/ can be characterized as lamino-alveolar. Some of the English voiced fricatives can be found as contextual variants or in other accents of Spanish. For instance, [v] is a variant of /b/ in some areas of Andalucia; [z] is a contextual allophone of /s/ before a voiced consonant; [ʒ] is the realization of /ʝ/ and /ʎ/ in Argentinian Spanish; [h] is found as a realization of /x/ in some areas such as the Canary Islands, Caribbean or as a realization of coda /s/ followed by consonant in central Spain; /dʒ/ is a variant of /ʝ/.

As for the nasals, Spanish has a palatal nasal phoneme /ɲ/ whereas the velar nasal is always a contextual allophone of /n/. Spanish has a larger inventory of liquids. As opposed to English postalveolar or retroflex approximant /ɹ/ Spanish contrasts an alveolar trill /r/ and a tap /ɾ/. The alveolar lateral which is often but differently velarized in American and British English is only velarized in Spanish as a result of anticipatory assimilation before velar plosives. The Spanish palatal lateral phoneme /ʎ/ is receding in most accents, converging with /ʝ/. However, because /ʎ/ is also a phoneme in Basque, it is still considerably present in the Spanish spoken in the Basque Country, though gradually giving way to /ʝ/ too. Finally, the English approximants /w j/ are present in Spanish either as allophones of the corresponding high vowels in rising diphthongs or, in the case of [j], as a variant of /ʝ/ (Navarro Tomás, 1918; Harris, 1969; Hualde, 2005).

A further issue is the differing relationship between phonemes and graphemes in the two languages. English is well known for the opacity of its spelling, whereas Spanish has often been called a "phonemic language." These statements are quite near the truth but need qualification. It is true that some English phonemes, particularly vowels and consonants such as /z θ ð/, are difficult to interpret from orthography; "g" may represent /g/ and /dʒ/, while "r" is only pronounced prevocalically in some varieties of English. However, other consonants such as /p t k b d v m n l/ have a much clearer spelling correspondence. On the other hand, Spanish sound-letter correspondence breaks down in a few cases. For instance, /b/ is represented by both "b" and "v" and "h" is not pronounced.

C. Speech and noise materials

Speech tokens were drawn from the vowel-consonant-vowel (VCV) corpus collected by Shannon *et al.* (1999). Although 23 consonants were available, a subset of 16 consonants /p b t d k g tʃ f v s z ʃ m n l r/ was employed. Members of the chosen subset have a clear consonantal character and are unambiguous with respect to their orthographic representations. This latter factor was important since the English listener group was not familiar with phonetic symbols. Consonants such as /θ ð ðʒ ʒ/ with a problematic orthographic-phoneme correspondence were not used, while /ŋ/ was excluded because it is not phonotactically possible in the vowel context chosen. The vowel context used was /ɑ_ɑ/ and remained constant to avoid coarticulatory differences between stimuli. This vowel context has been found to be better than high back and front vowels for consonant identification because of its well-defined F1 and F2 transitions which favor consonant place and voicing identification (Hazan and Simpson, 2000).

Two tokens of each VCV from each of five male talkers

made up a test set of 160 items. Data for the English listeners in the quiet condition was taken from a pretest condition of a larger study (which involved the same group of native listeners as the current study) and contained 80 items rather than 160. Tokens were normalised to have equal root-mean-square (rms) energy and sampled at 25 kHz. An additional two examples of each token were used to create a set of 32 practice items which were appended to the front of the test set but not scored. Listeners were not informed about the presence of the practice items.

Four masker types—8-talker babble, speech-shaped noise, and competing English and Spanish speech—were employed. All maskers apart from the Spanish speech were derived from male talkers within dialect regions d1, d2, and d3 of the TIMIT corpus (Garofolo *et al.*, 1992). Babble was produced by summing utterances whose rms energy had been equalized. Speech-shaped noise was generated by passing white noise through a filter whose magnitude response was the long-term average spectrum of speech material from the aforementioned dialect regions of TIMIT. Competing English sentences were chosen at random from TIMIT. Spanish speech material was provided by a native male speaker of the language at the University of the Basque Country.

VCVs were combined with each of the four masker noise types such that the masker signal started 1 s before the onset of the VCV and continued up to the VCV offset. Leading maskers were chosen to increase the likelihood of informational masking effects in the single competing talker conditions. Speech and noise signals were combined such that the SNR in the overlapped region was 0 dB for each token. The duration of the overlapped region varied from 0.57 to 1.21 s (mean=0.88, s.d.=0.14).

D. Procedure

The native group was tested at the University of Sheffield while the non-native group was tested at the University of the Basque Country. At both sites, stimulus presentation and response collection was under computer control. Participants used a mouse to select their response category from a 4×4 grid representing an orthographic version of the 16 consonants presented on a computer screen. English listeners were not familiar with phonetic symbols but their native knowledge and literacy made it possible to use orthographic symbols. Although Spanish listeners had been introduced to phonetic symbols in other English language courses, the degree to which they were familiar with English sounds was not known since they had just started a 15-week English Phonetics course. Both groups of listeners were informed that the test involved the identification of English consonants. Participants were instructed to listen out for the consonant that they would hear between vowels in nonsense words and click on the corresponding symbol box. After a short delay, the next stimulus was presented. In this way, participants governed the presentation rate of the stimuli.

Each of the five conditions required 8–10 min to complete. At the University of Sheffield, participants were tested individually in an IAC single-walled acoustically-isolated booth using Sennheiser HD250 headphones. At the University of the Basque Country, participants were tested in groups of 15–20 in a quiet laboratory using Plantronics Audio-90 headphones. Stimuli were presented diotically and listeners were able to adjust the level to a comfortable setting.

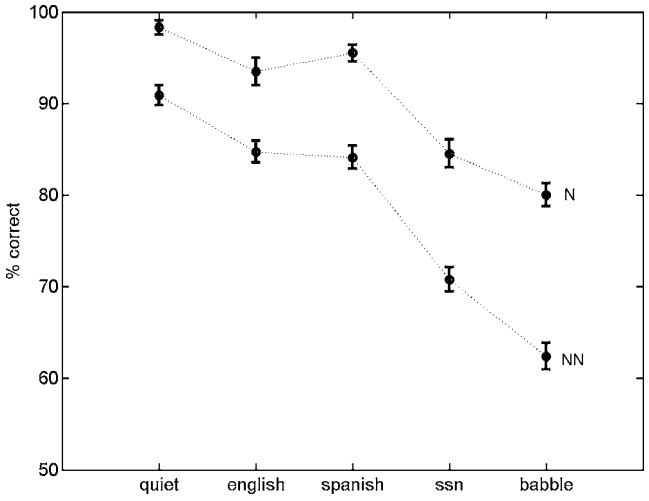


FIG. 1. Consonant identification scores of native (N) and non-native (NN) listeners in quiet and four masking conditions, each at a SNR of 0 dB: English=competing talker in English, Spanish=competing talker in Spanish, ssn=speech-shaped noise, babble=8-talker babble noise. Dotted lines indicate which points belong to the same listener group, and error bars define 95% confidence intervals.

To determine whether the difference in stimulus presentation equipment in the two countries could influence the results, a separate group of seven native English listeners were tested at the University of Sheffield in quiet and in the speech-shaped noise condition using a setup (quiet room, PC soundcard, Plantronics Audio-90 headphones) similar to that used in the University of the Basque Country. Identification results were compared with those obtained in quiet and speech-shaped noise by the main English experimental group. Means for both quiet (99.7 vs 98.3; $t(26)=1.44$, $p=0.162$) and speech-shaped noise (83.4 vs 84.6; $t(26)=-0.649$, $p=0.404$) were not significantly different.

Listener performance in quiet was measured first. Listeners were then presented with the four masking conditions in a random order. All five conditions (quiet plus the four maskers) were tested in a single session lasting 45–50 min.

III. RESULTS

A. Effect of masker type

Figure 1 compares native and non-native identification of consonants in quiet and in the four masking conditions. As expected, performance deteriorated for both groups in all masking conditions. A repeated measures ANOVA with one within-subjects factor (masker type) and one between-subjects factor (nativeness) confirmed the effect of masker ($F(4,75)=375$, $p<0.0005$, $\eta^2=0.952$) and group ($F(1,78)=212$, $p<0.0005$, $\eta^2=0.731$). Listeners in both groups performed best in the presence of competing speech and worst in a babble masker. Spanish listeners suffered more from the presence of noise. Post hoc comparisons (with Bonferroni adjustment for multiple comparisons) showed statistically significant ($p<0.0005$) differences between natives and non-

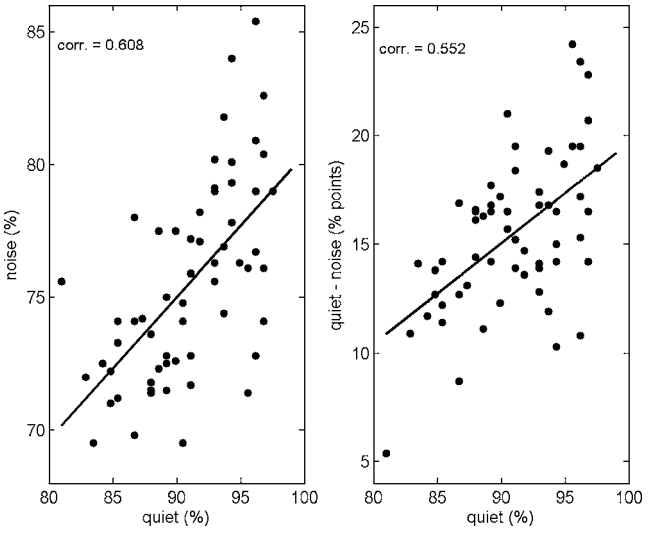


FIG. 2. Scatter plots of non-native performance in quiet versus noise (left panel) and quiet vs degradation in noise, measured as the percentage points difference between quiet and noise (right panel). Each point represents an individual listener. Noise scores are averaged over the four masking conditions. The best (least squares) linear fit and Pearson correlation coefficient are shown.

natives in each noise condition. Similarly, post-hoc comparisons of noise conditions demonstrated significant ($p < 0.0005$) differences in masking effectiveness apart from the two competing talker conditions ($p > 0.05$).

The interaction between masker type and nativeness was highly significant ($F(4,75) = 15.4$, $p < 0.0005$, $\eta^2 = 0.452$). The native advantage of 7 percentage points in quiet increased to an average of 10 for the two competing speech conditions, 14 for speech-shaped noise and 18 for babble. All pairwise comparisons of noise conditions (with the two competing speech conditions combined into a single mean measure) showed a significant interaction between masker type and nativeness. The largest interaction with nativeness was between quiet and babble ($F(1,78) = 60.24$, $p < 0.0005$, $\eta^2 = 0.436$) while the smallest was between the mean of the competing talker conditions and speech-shaped noise ($F(1,78) = 8.37$, $p < 0.005$, $\eta^2 = 0.097$).

B. Within-group differences for non-native listeners

As expected, the native group exhibited far less variability in quiet conditions than the non-native group (s.d. of 0.9 for natives vs 4.2 for non-natives). Non-native groups are far less homogenous than native listeners due to differences in the onset, amount and quality of exposure to the FL and the stage of language learning. Given the spread of identification rates in quiet for non-native listeners, it was possible to explore the extent to which performance in noise is predictable on the basis of performance in quiet conditions. Such an analysis is not meaningful for the native listeners since very few errors were made in quiet by this group. Figure 2 (left panel) plots performance in quiet against mean performance in the 4 masking conditions for each non-native listener. A clear relationship is visible: the better the ability to recognize VCVs in quiet, the better the performance in noise (Pearson correlation $= 0.608$, $p < 0.001$). In fact, non-native performance in quiet was significantly correlated with identification rates in each of the four masking conditions individually.

If the spread of non-native performance in quiet represents different levels of phonetic competence in the foreign language, the interaction between noise condition and nativeness described above could be explained by disproportionately worse performance in noise by the least competent non-natives. If so, the deterioration in performance between quiet and noisy conditions should be negatively-correlated with performance in quiet. Figure 2 (right panel) plots the difference in percentage points between identification rates in quiet and the mean of the four noise conditions as a function of performance in quiet for each non-native listener. Contrary to the prediction that the least competent listeners in quiet conditions would show greater deterioration in noise, the reverse was found (Pearson correlation $= 0.552$, $p < 0.001$). Those listeners who performed well in quiet displayed the greatest absolute drop in identification rates in noise relative to quiet. Significant correlations were found for each of the four noise conditions independently.

C. Language of competing speech

A planned comparison of the effect of language for the competing talker maskers showed no significant difference overall. However, there was a significant interaction of language and nativeness ($F(1,78) = 6.82$, $p = 0.011$, $\eta^2 = 0.08$). Within-group comparisons revealed that this was due to a small but significant difference for the native group ($F(1,78) = 5.23$, $p = 0.025$, $\eta^2 = 0.063$). The language of the competing talker had no effect on the non-native group. This result suggests that English listeners were better able to tune out an unknown language, while both maskers were equally disturbing for Spanish learners of English.

D. Consonant identification

Figure 3 compares native and non-native identification scores for each of the 16 consonants in quiet (upper panel) and averaged over the four masking conditions (lower panel). Figure 4 depicts the native advantage, measured in percentage points, in quiet (upper panel) and in noise (middle panel). The lower panel of Fig. 4 shows the additional native advantage in noise over that in quiet, calculated as the difference between the measures in the middle and upper panels. Tables I to IV show consonant confusions for the two groups in quiet and in noise. Consonants /b v f s z/ are identified much less accurately by non-natives in quiet (Fig. 4, upper panel). However, it is striking that for most of these sounds, non-natives are not further disadvantaged in noise (Fig. 4, lower panel). In fact, most of the increased native advantage in noise comes from poorer identification of the consonants /p d g k l r tʃ f/. The remaining consonants /m n t ʃ/ are identified both in quiet and in noise at similar rates by both groups.

Among the sounds for which non-native listeners' perceptions are considerably worse in quiet conditions, some clear L1 influences can be observed. Sounds such as /v/ and /z/ which are not part of the Spanish phonological inventory,

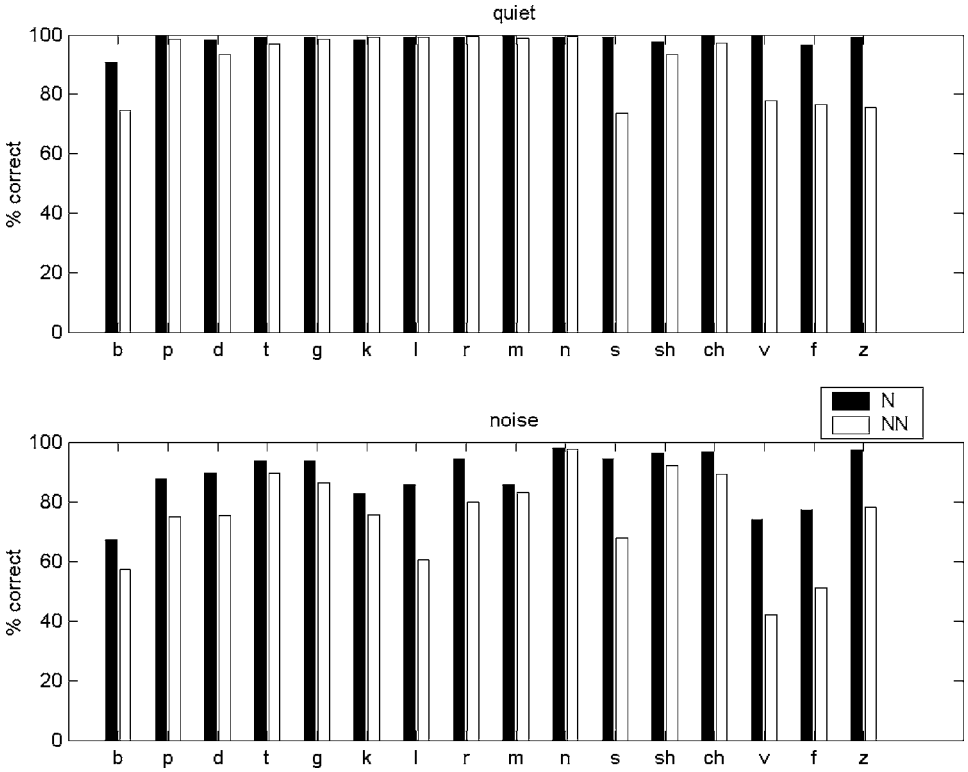


FIG. 3. Native and non-native identification scores for individual consonants in quiet (upper panel) and averaged over the noise conditions (lower panel). Here, and in Fig. 4, labels "sh" and "ch" correspond to the consonants ʃ and tʃ.

are not, nevertheless, "uncategorizable" in L1 terms (Best, 1995). Rather, they represent what PAM would call "deviant" to "acceptable" exemplars of the L1 phonetic categories /b/ and /s/, respectively. The sound [v] is a realization of /b/ in some varieties and, further, orthographically "b" and "v" are realized as /b/. Thus, English /v/ may be perceived by Spanish listeners as a deviant realization of /b/. Indeed, as the confusion matrix of Table II shows, non-native listeners' mainly misidentify /v/ in quiet as /b/. Contrary to appearances, English intervocalic /b/ may be classified as a "poor" exemplar of Spanish /b/ since the phonetic cues that code intervocalic voiced plosives in Spanish (full voicing and lenition) are quite different to those used in English and can even result in an acceptable exemplar of /p/. Table II shows that non-native listeners hear /b/ as /p/ 21% of the time. Regarding /s/ and /z/, the former is a phonetic category in both English and Spanish with realizational differences while [z] is a contextual allophone in Spanish. Consequently, both

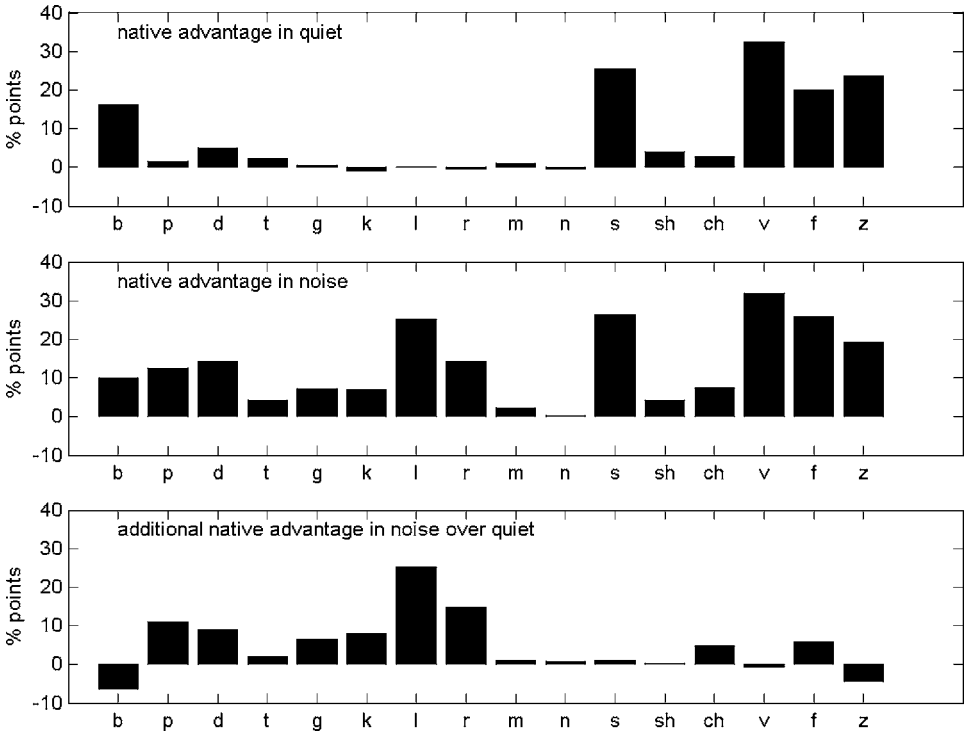


FIG. 4. Differences in native and non-native consonant identification rates, measured in percentage points. The upper panel shows the native advantage in quiet. The middle panel displays the native advantage in noise, where the noise scores are averages over the four masking conditions. In the lower panel, the additional native advantage in noise over quiet is shown. This is simply the difference between middle and upper panels.

TABLE I. Consonant confusions in quiet for the native group, expressed as percentages. Rows represent stimuli presented while columns denote participant responses. Rows may not sum to 100 since percentages are rounded.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	91	3	5		1											
p		100														
d			99	1												
t				99			1									
g		1			99											
k						98							2			
l							99							1		
r		1						99								
m									100							
n					1					99						
s			1								99					
ʃ												98	2			
tʃ													100			
v								1			1			90		9
f			1	1							2				96	
z											1					99

TABLE II. Consonant confusions in quiet for the non-native group.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	75	21	4													
p		98		1												
d			93	6												
t				97									2			
g					98	1										
k						99										
l			1				99									
r								99								
m									99	1						
n										100						
s											74	5				21
ʃ											5	93	1			1
tʃ				2									97			
v	15													78	3	3
f											5			3	77	15
z											18	7				75

FL sounds fall under the one L1 category with different goodness ratings, which accounts for their nonoptimal identification levels (Best, 1995). This is borne out by the confusions in Table II since there are mutual confusions between the two categories: /z/ is misperceived as /s/ on 18% of presentations and /s/ is classified as /z/ 21% of the time. The remaining consonant showing greatest native advantage in the quiet condition is /f/ which is very similar to a NL prototype. Here, the high level of confusion is not predicted by models such as those of Best (1995) and Flege (1995), unless there is another FL sound competing in the same prototype area. Indeed, /v/ could be a candidate. However, an examination of the confusion matrix suggests that /f/ is most often confused with /z/, which points towards an acoustic influence, with friction as the salient perceptual cue.

Most of the sounds which show the greatest native advantage in quiet are not additionally disadvantaged in noise (Fig. 4). In fact, two categories, /z/ and /b/, show *reduced* native advantage in noise. In the case of /z/, as shown by the perception scores in Fig. 4 (upper panel), this could indicate that the voicing feature used to identify it is resistant to noise maskers (Miller and Nicely, 1955; Hazan and Simpson, 2000). For /b/, the decrease in native advantage is due to a large drop in performance for native listeners in noise (Fig. 3, lower panel, and Table III). Confusion patterns in noise (Tables III and IV) show that native listeners most often confuse /b/ with /v/ whereas non-native listeners confuse it predominantly with /p/. This disparity indicates the use of different cues in noise for the two groups. For native listeners, voicing appears to be the most salient cue while non-native listeners seem to employ the same L1 cues as those used in the quiet condition, namely voice onset time and lack of lenition. Therefore, L1 influences appear to determine confusions but this influence is not stronger in noise.

For the sounds /p d g k l r/ the native advantage only becomes apparent in noise. An examination of their confusion patterns shows that both voiceless plosives are mainly confused with other voiceless plosives, a pattern which is also visible for native listeners, suggesting that place of articula-

TABLE III. Consonant confusions in noise (averaged across the four masking conditions) for the native group.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	67	3	2	1			6	2	2				1	14	1	
p	2	88		1		5								1	2	
d			89		7		1			2						
t		3	1	94		2										
g	1		1		94	1				1				1		
k	1	11		1	1	83									2	
l	1						86	3	6	1				2		
r	1	1					1	94						3		
m	1						4		86	7					1	
n									1	98						
s											94				1	2
ʃ			1									96	2			
tʃ						1						1	97			
v	13	2					2	3	3					74	2	
f	5	5	1			3	1							8	77	
z											1			1		97

TABLE IV. Consonant confusions in noise for the non-native group.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	57	19	4	2	3	1	2	1	1				1	7	2	
p	3	75	1	4	2	10			1					1	2	
d	1		75	7	11	1	1	1		2						
t		1	2	89		3					1		1			1
g	1		3	1	86	6		1		1				1		
k	1	11	1	5	2	76									2	
l	7	3	3		2		60	4	12	3	1			4		1
r	5				5		3	80	3					3		
m	1	1			1		3	1	83	7	1			1	1	
n									1	98						
s				1							68	4			1	25
ʃ											4	92	2			1
tʃ			1	2	1	1					1	5	89			
v	29	9	1	1	5	1	1	2	3					42	5	1
f	6	21	1	1	1	3			1		3		1	6	51	4
z											16	4				78

tion information is less resistant to noise than voicing (Miller and Nicely, 1955). Non-native listeners confuse /g/ with its voiceless counterpart, following the same influences observed above for /b/ confusions, i.e., L1 cues for differentiating voiced/voiceless plosives. Finally, the two liquids show the biggest native advantage in noise and the greatest degree of confusion dispersions. In both cases there are prototypes in the learner's L1 system to which the English category could be assimilated as a good or acceptable exemplar, predicting good identifications (Best, 1995): English /ɹ/ could be related to both Spanish /r/ and /ɾ/; English nonvelarized /l/ is similar to Spanish /l/. Indeed, this is the pattern observed in the quiet condition. However, the considerable realizational differences (/l/ velarization and /r/ retroflexion in American English) between the FL and the L1 phonetic categories show up under the strain of unfavourable listening conditions. Some of the non-native confusions for /l/ are also visible to a lesser degree in the native group.

For the sounds /f/ and /v/ which are poorly identified by non-natives in quiet and in noise, native listeners' perceptions also deteriorate considerably in noise. These sounds are confused with similar categories by both groups. Non-natives mainly confuse /f/ with /p/ in noise rather than /s/ in quiet, but in both cases inherent susceptibility to masking rather than L1 influences is the likely cause. The case of /v/ is particularly interesting. In noisy conditions, native and non-native confusions are dominated by /b/, suggesting that friction is masked by noise. This is a natural, perhaps universal, confusion with historical instantiations such as the merging of /v/ and /b/ in Spanish and the change of Latin intervocalic /b/ to /v/ in some Romance languages.

IV. DISCUSSION

A. Effect of different masker types on native versus non-native consonant identification in noise

On a consonant identification task, English outperformed Spanish listeners in quiet conditions by 7 percentage points. The performance differential in all masking conditions was larger than in quiet, ranging from 9 to 18 percentage points. Masking conditions which caused most errors for natives were also those which resulted in larger differences in native and non-native identification rates.

The ranking of masking effectiveness of the three noise types employed in this study was identical for natives and non-natives. Both groups identified consonants in a competing talker background at a higher rate than in a steady noise masker, and both performed worst when the masker was 8-talker babble. This ranking of masking effectiveness has been found in other studies using native listeners (e.g., Simpson and Cooke, 2005) and probably reflects the relative amount of energetic and informational masking produced by the various masker types when added at a fixed SNR.

A single competing talker produces less energetic masking than 8-talker babble, which in turn is a less effective energetic masker than speech-shaped noise. However, speech-shaped noise has no additional informational masking effect, while both babble and competing speech can produce significant amounts of informational masking (Carhart *et al.*, 1969; Brungart *et al.*, 2001; Freyman *et al.*, 2004). The degree of informational masking obtained depends on several factors such as the similarity of target and masker and the number of talkers in the background. For example, Freyman *et al.* (2004) demonstrated a maximal effect of informational masking with two talkers in the background.

The study by Simpson and Cooke (2005) is particularly relevant, since it used the same VCV corpus in a large number of masking conditions, including the three employed here, albeit at a more adverse fixed SNR of −6 dB. Simpson and Cooke demonstrated that a competing talker produces negligible amounts of informational masking for this VCV corpus. Instead, maximal informational masking arises from 8-talker babble. Consequently, the masking effect of babble in the current study may be interpreted as due to the combined effects of energetic and informational masking. The 14 percentage point difference in performance between the two groups in stationary noise is less than the 18 percentage points deficit in babble, which has a smaller energetic masking effect than stationary noise. Consequently, it is reasonable to conclude that non-native listeners are more adversely affected then native listeners by informational masking. Further studies with more confusable competing speech tasks are required before the native informational masking advantage can be quantified.

Regarding the effect of the two competing speech conditions, there was a small effect of masker language for the native group but not for non-native listeners. Native listeners showed better performance when the language of the masker was unknown to them (Spanish) than when it was their L1 (English). In contrast, non-native listeners performed at a similar level irrespective of whether the masker language was English or Spanish, presumably because they spoke both languages. This suggests that the strong attentional component of competing speech had a larger effect when the language was known to the listeners. Although FL listeners commonly report that they are able to tune out FL speech better than their L1, in this case the fact that the task involved FL consonant identification may have made it harder to tune out the FL masker (Cutler, personal communication). Since the effect found was relatively small, further studies with other languages and speech material are necessary. If a language unknown to both groups were to be found to constitute a less effective masker for both, the asymmetry found in the present study could be interpreted as the result of reduced attentional demands. Sentence length speech material in the foreground might be expected to produce stronger language activation for the FL group.

The current study found a significant interaction between masker type and nativeness. The closest work to that reported here is Cutler *et al.* (2004), who found no interaction between noise level and nativeness. There are a number of differences between Cutler *et al.* (2004) and this study. First, Cutler *et al.* (2004) varied masking effectiveness by presenting tokens at a number of SNRs, while the current study employed different masker types known to differ in their degree of energetic masking. There is no direct way to compare the effectiveness of the maskers used in the two studies. However, the most difficult condition in both studies

occurred with babble at 0 dB SNR. Further, it is known that a competing speaker provides 6–8 dB less masking than stationary noise at the same level (Miller, 1947; Festen and Plomp, 1990). Since Cutler *et al.* (2004) used babble noise at 16, 8, and 0 dB and our stationary noise masker produced less masking than the babble condition, there are grounds for arguing that the range of masker effectiveness was broadly similar in the two studies.

A second difference between the two studies concerns the non-native samples (Dutch and Spanish). Dutch learners of English have been shown to display outstanding phonetic performance in English (Bongaerts, 1999; Broersma, 2005). In part, this is due to the quality and quantity of English exposure in Holland compared to that in Spain. In addition, interlanguage phonological distance, which has been considered to be a crucial factor in FL sound acquisition (Bongaerts *et al.*, 2000; Hazan and Simpson, 2000), is considerably smaller between English and Dutch than between English and Spanish.

There are other differences between the two studies. Cutler *et al.* (2004) used CV and VC syllables rather than VCV tokens and asked listeners to identify all phonemes rather than just the consonants. VCV tokens used in the current study were produced by American speakers while the native listeners were British. However, Cutler *et al.* (2005) demonstrated that American and Australian English listeners produced statistically indistinguishable performance on a task involving a subset of the stimuli employed by Cutler *et al.* (2004). Further, the placement of tokens relative to maskers differed in the two studies in such a way that listeners may have been able to better predict the onset of the VCV tokens used in the current study.

B. Fragility of non-native categories

Non-natives were less able to identify consonants in quiet conditions, which agrees with most of the literature on FL sound perception (Pisoni *et al.*, 1994; Best, 1995; Flege, 1995). Clear L1 influences were observed among consonant confusion patterns. The additional native advantage seen in the presence of noise (Fig. 4, lower panel) for sound identification tasks such as that employed here which involved low-level phonetic processing, might be explained by interaction of the L1 and FL phonetic systems (Mayo *et al.*, 1997) compounded by noise degradation (Hazan and Simpson, 2000). Noise may expose the lack of robustness of a non-native listener's FL categories and reveal the use of different phonetic cues for certain FL categories due to L1 interference and incomplete FL sound acquisition. The L1 experience of a native listener will include exposure to adverse conditions, so native identification is more likely to involve the use of multiple, redundant cues and appropriate cue weighting strategies to overcome the effects of energetic masking. In contrast, non-natives may have developed fewer cues and less sophisticated weighting strategies due to limited or faulty exposure to the FL categories and less experience with the FL in noisy conditions. Additionally, non-native listeners may use cues influenced by their L1, and such cues may be affected by noise in a different manner from those used by native listeners. For example, Spanish listeners may make less use of aspiration and listen for lenition to differentiate voiceless/voiced plosives.

Noise constitutes a good testing ground in which to compare native and non-native consonant identification since native performance departs from the near ceiling levels observed in a quiet background. Thus noise can be used to examine whether confusions pattern similarly for both listener groups and hence distinguish between "universal" misidentifications due to acoustic factors such as inherent maskability, and non-native specific confusions ascribable to L1 influences.

The fact that some consonants only show native advantage in noise indicates that the phonetic categories non-native listeners were using in less demanding situations (and which in some cases allowed them to reach the level of native listeners' performance) are too fragile to withstand adverse listening conditions. Therefore, although perception in quiet reveals important information about a non-native listener's level of sound acquisition, it may still be seen as a "performance" measure that only taps indirectly into "competence." As has been suggested (Mayo *et al.*, 1997; Pallier *et al.*, 1997; Bosch *et al.*, 2000), even bilinguals can be shown to differ in competence from monolingual native speakers, although these differences only become apparent under careful testing. Sound perception in noise, in which some of the usual phonetic cues are not reliable due to energetic and informational masking, appears to be a good test of the robustness of FL categories and phonetic competence.

V. CONCLUSIONS

English listeners outperformed Spanish listeners in a task involving the identification of medial consonants in English VCV tokens. The native performance advantage increased when tokens were presented at a fixed SNR of 0 dB in speech-shaped noise, 8-talker babble and in two competing speech maskers. Since stationary noise has no informational masking effect, this result suggests that non-native listeners are more adversely affected by pure energetic masking than are native listeners. The ranking of masking effectiveness across masking conditions was identical for the native and non-native groups, with competing speech being the least effective masker and babble the most challenging. Given that, at a fixed SNR, babble is a less effective energetic masker than stationary noise, these findings also suggest that non-native listeners were more adversely affected by informational masking.

Non-native listener performance in quiet correlated well with performance in noise. However, performance degradation in noise was also positively correlated with identification rate in quiet, suggesting a lack of robustness in non-native FL phonetic categories which is only apparent under adverse conditions such as speech perception in noise.

Spanish listeners performed at the same level in the presence of both English and Spanish competing speech maskers. However, English listeners were slightly better

when the language of the competing speech was Spanish as opposed to English, perhaps due to a reduced attentional engagement for an unknown language.

At the level of individual consonants, those most poorly identified by non-natives in the quiet condition did not suffer any further disadvantage in noise. Most of the additional native advantage was due to a subset of consonants which were identified well in quiet conditions by both native and non-native listeners. A more comprehensive study of consonant identification in noise might reveal important information about the structure of phonetic categories used by non-native learners. Finally, further studies with words and sentences are necessary to discern the effect of competing talkers, masker language, and attentional factors on native and non-native perception in everyday speech conditions.

ACKNOWLEDGMENTS

This work was supported by grants from the Spanish Ministry of Science and Technology and the Basque Government (BFF 2003-04009; 9/UPV 00103.130). The authors thank Sarah Simpson for help in collecting data from the English listeners and Ann Bradlow, Anne Cutler, and two anonymous reviewers for their valuable suggestions.

[1]Second language (L2) learning usually refers to learning a language in the area where it is spoken, after another language (L1) has been acquired. Foreign language (FL) indicates that the language is not a language in the community and thus it has a foreign status (notwithstanding the fact that a FL learner can go the FL country to pursue its learning). In this paper, the term FL is used to refer to learning a language after the L1 has been established.

Assmann, P. F., and Summerfield, Q. (**2004**). "The perception of speech under adverse acoustic conditions," in *Speech Processing in the Auditory System, Springer Handbook of Auditory Research*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, Berlin), Vol. **18**.

Best, C. T. (**1995**). "A Direct Realist View of Cross-Language Speech, Perception," in *Speech Perception and Linguistic Experience*, edited by W. Strange (Timonium, MD), pp. 171–204.

Bongaerts, T. (**1999**). "Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners," in *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong (Erlbaum, Mahwah), pp. 133–159.

Bongaerts, T., Mennen, S., and van der Slik, F. (**2000**). "Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language," Studia Linguistica **54**, 298–308.

Bosch, L., Costa, A., and Sebastián-Gallés, N. (**2000**). "First and second language vowel perception in early bilinguals," European J. of Cognitive Psychology **12**, 189–221.

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., and Pisoni, D. B. (**1999**). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," J. Acoust. Soc. Am. **106**, 2074–2085.

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic màsking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **100**, 2527–2538.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Cooke, M. P. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cutler, A., Weber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Cutler, A., Smits, R., and Cooper, N. (**2005**). "Vowel perception: Effects of non-native language vs non-native dialect," Speech Commun. **47**, 32–42.

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.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings and problems," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium), pp. 233–277.

Florentine, M., Buus, S., Scharf, B., and Canevet, G. (**1984**). "Speech reception thresholds in noise for native and non-native listeners," J. Acoust. Soc. Am. **75**, s84.

Freyman, R. L., Balakrishnan, U., and Heifer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L. (**1992**). "DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM," NIST.

Harris, J. W. (**1969**). *Spanish Phonology* (MIT Press, Cambridge).

Hazan, V., and Simpson, A. (**2000**). "The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects," Lang Speech **43**, 273–294.

Hualde, J. I. (**2005**). *The Sounds of Spanish* (Cambridge University Press, Cambridge).

Imai, S., Walley, A. C., and Flege, J. E. (**2005**). "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners," J. Acoust. Soc. Am. **117**, 896–907.

Ioup, G. (**1984**). "Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition," Lang. Learn. **34**, 1–17.

Leather, J., and James, A. (**1991**). "The acquisition of second language speech," Stud. Second Lang. Acquis. **13**, 305–341.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Miller, G. A. (**1947**). "The masking of speech," Psychol. Bull. **44**, 105–129.

Miller, G. A., and Nicely, P. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Navarro Tomás, T. (**1918**). *Manual de Pronunciatión Española*, 21st ed. (CSIC, Madrid, **1982**).

Pallier, C., Bosch, L., and Sebastián-Gallés, N. (**1997**). "A limit on behavioural plasticity in speech production," Cognition **84**, B9–B17.

Pisoni, D. B., Lively, S. E., and Logan, J. S. (**1994**). "Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception," in *The Development of Speech Perception: The Transition From Speech Sounds to Spoken Words*, edited by J. C. Goodman and H. C. Nusbaum (MIT Press, Cambridge) pp. 121–166.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Scovel, T. (**1969**). "Foreign accents, language acquisition, and cerebral dominance," Lang. Learn. **19**, 245–253.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.

Simpson, S., and Cooke, M. P. (**2005**). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778.

Takata, Y., and Nábělek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.

van Wijngaarden, S. J., Bronkhorst, A. W., Houtgast, T., and Steeneken, H. J. M. (**2004**). "Using the Speech Transmission Index for predicting non-native speech intelligibility," J. Acoust. Soc. Am. **115**, 128l–1291.

van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.

Age-related differences in identification and discrimination of temporal cues in speech segments

Sandra Gordon-Salant and Grace H. Yeni-Komshian
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

Peter J. Fitzgibbons
Department of Hearing, Speech, and Language Sciences, Gallaudet University, Washington, D.C. 20002

Jessica Barrett
Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland 20742

(Received 30 August 2005; revised 10 January 2006; accepted 11 January 2006)

This study investigated age-related differences in sensitivity to temporal cues in modified natural speech sounds. Listeners included young noise-masked subjects, elderly normal-hearing subjects, and elderly hearing-impaired subjects. Four speech continua were presented to listeners, with stimuli from each continuum varying in a single temporal dimension. The acoustic cues varied in separate continua were voice-onset time, vowel duration, silence duration, and transition duration. In separate conditions, the listeners identified the word stimuli, discriminated two stimuli in a same-different paradigm, and discriminated two stimuli in a 3-interval, 2-alternative forced-choice procedure. Results showed age-related differences in the identification function crossover points for the continua that varied in silence duration and transition duration. All listeners demonstrated shorter difference limens (DLs) for the three-interval paradigm than the two-interval paradigm, with older hearing-impaired listeners showing larger DLs than the other listener groups for the silence duration cue. The findings support the general hypothesis that aging can influence the processing of specific temporal cues that are related to consonant manner distinctions. [DOI: 10.1121/1.2171527]

PACS number(s): 43.71.Lz, 43.71.Es, 43.71.Gv, 43.66.Mk [JHG] Pages: 2455–2466

I. INTRODUCTION

In recent years, compelling evidence has accumulated to show that elderly listeners experience difficulties in understanding speech with altered timing characteristics. For example, older listeners exhibit poorer recognition performance compared to younger listeners for sentences that are presented at an increased rate relative to natural-rate speech (Wingfield *et al.*, 1985; Gordon-Salant and Fitzgibbons, 1993; Vaughan and Letowski, 1997). Additionally, older listeners have more difficulty than younger listeners in accurately recognizing sentences with artificial disruptions in overall prosody (Gordon-Salant and Fitzgibbons, 1997).

One form of natural disruption in timing and prosody occurs in speech produced by non-native speakers of English. Accented English is characterized by numerous temporal alterations that modify individual phonemes as well as the overall timing of the sequence of syllables and words in a spoken message. Some specific temporal changes in accented English at the segmental level include variations in voice-onset time (VOT) (Flege and Eefting, 1988; MacKay *et al.*, 2000), syllable stress (Adams and Munro, 1978), and vowel duration (Fox *et al.*, 1995). Another alteration identified in accented English is overall sentence duration (Guion *et al.*, 1997). Disruptions at both the segmental and suprasegmental levels potentially act to vary the global accent of speech. At least one investigation has shown a significant age-related difference between younger (20–39 years) and older (60+years) adults in recognizing English spoken with different degrees of accentedness (Burda *et al.*, 2003). However, this prior study did not elucidate the source of the age-related deficit as deriving primarily from altered segmental characteristics or suprasegmental features.

Converging support for the notion that older listeners have a reduced capacity to process temporal changes in speech may come from psychoacoustic investigations with nonspeech stimuli. Older listeners demonstrate larger gap detection thresholds than younger listeners, particularly for gap intervals placed near the onset or offset of a noise marker (e.g., He *et al.*, 1999; Snell and Hu, 1999). They also show longer duration discrimination thresholds than younger listeners for tonal stimuli and silent intervals between tonal markers (Fitzgibbons and Gordon-Salant, 1994), as well as for silent intervals between noise-band markers of disparate frequencies (Lister *et al.*, 2002). One study has extended these results to show an effect of age on discrimination of silent gaps inserted in synthetic speech signals (Lister and Tarver, 2004). These findings suggest that identification and discrimination of discrete speech signals that are dependent on duration cues should reveal substantial age-related differences.

The overall objective of this preliminary investigation is to identify some age-related perceptual deficits for isolated temporal acoustic cues in consonants and vowels that are necessary for accurate word identification, and that may be characterized by considerable variation and/or deviation in non-native speech. The temporal acoustic cues selected for this study include vowel duration as a cue for final consonant

voicing (Denes, 1955; Peterson and Lehiste, 1960; Luce and Charles-Luce, 1985), voice-onset time (VOT) as a cue for initial stop consonant voicing (Lisker and Ambramson, 1964), glide duration as a cue for the stop-consonant vs glide distinction (Diehl, 1976), and duration of a silent interval to cue the fricative vs affricate distinction (Dorman *et al.* 1979). Relatively little is known about the ability of older listeners to perceive many of these temporal cues in natural speech. For synthetic speech continua with varying VOT, older listeners have been shown to perform differently than younger listeners on identification and discrimination tasks (Strouse *et al.*, 1998). Age-related differences in identification and discrimination of natural speech signals with variation in VOT and other temporal cues have not been reported previously.

A second purpose of the current study is to distinguish age-related difficulties from those attributed to hearing impairment. The primary research questions of interest relate to whether or not elderly listeners have difficulties with temporal cues in speech as a consequence of reduced audibility of high frequency sounds or a diminished capacity to discriminate the duration of speech cues. Additionally, prior studies with nonspeech signals suggest that specific stimulus parameters may be necessary to reveal effects of age and hearing impairment in temporal judgments. For example, older listeners show larger discrimination thresholds for unfilled signals (e.g., silent gaps between tones) than for filled signals (e.g., tones) (Fitzgibbons and Gordon-Salant, 1994). Additionally, hearing loss effects have been observed for stimuli with shorter reference durations but not for stimuli with longer reference durations (Fitzgibbons and Gordon-Salant, 1994). Previous studies have suggested also that identification performance may be predictable from discrimination performance, at least for judgments about tonal temporal order in comparable sequences (Fitzgibbons and Gordon-Salant, 1998). The extent to which these principles apply to older listener's discrimination and recognition of natural speech is unclear at present.

To investigate these issues, duration judgments for selected temporal cues for vowels and consonants were measured. Stimulus parameters include reference duration and type of segmental cue (filled, as in voiced speech segments, vs unfilled, as in affricate silence duration). Discrimination and identification of these temporal cues for single word stimuli were evaluated in separate conditions. The acoustic cues selected for the present investigation include VOT as a cue for initial stop consonant voicing, vowel duration as a cue for final stop consonant voicing, silence duration as a cue to distinguish sibilant from affricate, and transition duration as a cue for the stop vs glide distinction. Listener groups include elderly listeners with normal hearing, elderly listeners with hearing impairment, and noise-masked young listeners in which the noise masking is sufficient to equate thresholds with those of the elderly listeners with normal hearing. Comparisons of performance between the noise-masked young listeners and elderly listeners with normal hearing provide an estimate of age-related effects, and comparisons of performance between the elderly normal-hearing listeners and elderly hearing-impaired listeners provide an estimate of the effects of hearing impairment. Thus, if hearing loss is the primary source of the problem, then the elderly hearing-impaired listeners should perform more poorly than the young and elderly normal-hearing listeners, and if age is the primary source of the problem, then the elderly hearing-impaired and elderly normal-hearing listeners should perform more poorly than the young normal-hearing listeners with equivalent masked thresholds to the elderly normal-hearing group. Based on previous findings of substantial age-related deficits in duration discrimination for non-speech signals, and the importance of processing temporal cues for speech understanding, we hypothesize that older listeners will exhibit a deficit compared to younger listeners, on identification and discrimination of temporal cues in individual speech segments. Moreover, it is expected that for all listeners, perception of single speech segment durations will be poorer for speech segments of shorter reference durations than longer reference durations.

II. METHOD

A. Subjects

One objective of this study was to separate the effects of hearing loss from the effects of age on perception of temporal differences in speech signals. To that end, three listener groups participated in the experiments. The first group was elderly listeners (64–77 years, mean age=71 years; $n=16$) with normal hearing, defined as pure tone thresholds $\leqslant$25 dB HL, from 250 to 4000 Hz (see ANSI, 1996). The second group consisted of elderly listeners (64–80 years; mean age=72.5 years; $n=15$) with mild-to-moderate or mild-to-moderately severe, sloping sensorineural hearing loss and good or excellent monosyllabic word recognition scores ($>80\%$). The third group of listeners was comprised of young listeners (18–33 years, mean age=21.4 years; $n=15$) with normal hearing sensitivity. Because the thresholds of the young listeners were considerably better than those of the older listeners with normal hearing, shaped noise masking was presented to the young listeners to shift their thresholds to be equivalent (within 5 dB), on average, to those of the elderly listeners with normal hearing. The average thresholds that served as targets were based on mean thresholds of elderly normal-hearing listeners who participated in several previous investigations. The software program COOL EDIT PRO (v. 2.0, Syntrillium Software; Graphic Equalizer Filter Option) was used to create the shaped noise, which was then burned onto a compact disc for presentation during the preliminary audiometric tests for verification of the shifted thresholds. This shaped noise also was presented to the young normal-hearing listeners throughout the experiments. Table I presents the average pure tone thresholds (250–4000 Hz) of the three listener groups.

Other preliminary audiometric criteria included normal middle ear function, as assessed by tympanometry, and acoustic reflex thresholds that were within the 90th percentile for a given pure tone threshold (Gelfand *et al.*, 1990). All listeners were tested using transient-evoked otoacoustic emissions (TEOAEs) to document their presence in listeners with normal hearing and their absence in listeners with hear-

TABLE I. Pure tone thresholds in dB HL (see ANSI 1996) in the test ear of the three subject groups. Data shown are group means, with standard deviations in parentheses.

	Frequency (Hz)				
Group	250	500	1000	2000	4000
Young masked	14.7 (5.2)	13.3 (3.6)	12.0 (3.2)	14.7 (3.5)	20.0 (4.6)
Elderly normal	12.8 (7.5)	10.0 (6.3)	10.9 (6.9)	14.4 (7.3)	23.1 (6.0)
Elderly hearing impaired	16.7 (7.9)	19.7 (10.1)	24.3 (11.8)	38.0 (10.5)	52.3 (7.0)

ing impairment. The absence of TEOAEs and the presence of acoustic reflex thresholds at expected levels confirmed a cochlear site of lesion among the listeners with hearing loss.

There were several additional criteria for subject selection. All listeners were required to be native speakers of English. They were in general good health, with no history of stroke, Parkinson's disease, or neurologic impairment, and possessed sufficient motor skills to provide a response using a computer mouse click. Additionally, all listeners passed a screening test for general cognitive awareness (Pfeiffer, 1977).

B. Stimuli

Four sets of contrasting word pairs were selected from the Revised Speech Perception in Noise Test (R-SPIN; Bilger *et al.*, 1984). Each word pair contrast relied on a single acoustic duration cue. A continuum was constructed for each pair that varied along the relevant single acoustic duration cue. An adult American male speaker with a General American dialect recorded the target word pairs in isolation as well as the low probability R-SPIN sentences that contained the eight target words. Recordings were made in a quiet room directly onto a PC, using a microphone and sound card with 16-bit resolution (Audigy Sound Blaster). The stimuli were analyzed and edited from the original recordings using waveform editing software (COOL EDIT PRO 2.0 at 44.1 kHz sampling rate). Additionally, the sound spectrogram display of a second waveform editing program, WEDW (available online at www.asel.udel.edu), was used to verify the temporal characteristics of the modified stimuli.

In this study, four pairs of contrasting words produced in isolation were used to generate the target speech continua. The four sets of contrasting words and the associated single acoustic duration cue included: (a) BUY/PIE, varying in VOT; (b) WHEAT/WEED, varying in vowel duration; (c) DISH/DITCH, varying in closure duration between the initial CV and the final consonant; and (d) BEAT/WHEAT, varying in transition duration of the initial consonant. Continua of each of these contrasting word pairs were created from one of the endpoint natural productions, by varying the relevant duration cue in equal intervals. Pilot data using these stimuli (described below) were collected from 11 young normal-hearing listeners, and verified that the endpoint signals from each continuum were clearly identified as the intended targets.

For the BUY/PIE continuum, the unmodified natural endpoint of the continuum was BUY (0 ms VOT, Stimulus #1). To create Stimulus #2, 10 ms of aspiration was excised from the natural PIE sample and inserted at a zero-crossing between the burst release and the onset of voicing of the original BUY stimulus. Aspiration was used, rather than a silent gap, because natural productions of voiceless plosives in English are aspirated (Klatt, 1975). Stimulus #3 was created by inserting an additional 10 ms of aspiration to stimulus #2 immediately after the burst and contiguous with the onset of aspiration from stimulus #2. This process continued until the creation of Stimulus #7, which had a VOT of 60 ms. Thus, stimuli on this continuum varied in VOT in 7 steps separated by 10 ms intervals, from 0 ms (BUY) to 60 ms (PIE). The duration of the endpoint stimulus, BUY, was 310 ms. The duration of Stimulus #7 endpoint, PIE, was 370 ms.

The WHEAT/WEED continuum is based on vowel duration. Several steps were taken in creating this continuum. The endpoint stimulus was a hybrid, consisting of a natural token of WEED, in which the /d/ release was excised and replaced by a high amplitude release taken from the final burst of a natural token of WHEAT. This stimulus was perceived as WEED, and represents the token with the longest vowel duration of the continuum. The next stimulus in the continuum had a shorter vowel duration and was created by cutting a single pulse between two zero-crossings from the steady-state portion of the preceding stimulus. Additional stimuli were created by excising single pulses from the steady-state portion of the preceding stimulus on the continuum. Each excised pulse was 7–8 ms in duration. Thus, the WHEAT/WEED continuum varied vowel duration in discrete steps corresponding to one glottal pulse ($\Delta t = 7-8$ ms). A 9-step continuum, rather than a 7-step continuum, was created to establish reliable endpoints in the pilot study. The endpoint signal with the shortest vowel (perceived as WHEAT) had a total duration of 249 ms (93 ms steady-state vowel duration), whereas the endpoint signal with the longest vowel (perceived as WEED) had a total duration of 311 ms (steady-state vowel duration of 155 ms).

The DISH/DITCH continuum is based on the duration of the silent interval between the vowel and the fricative /ʃ/. The endpoint stimulus was a hybrid in which the initial stop, vowel, and closure duration, were taken from the natural token, DITCH. The burst and fricative of /tʃ/ were excised and replaced by the fricative /ʃ/ of a natural token of DISH. Thus, this stimulus, which was perceived as DITCH, consisted of an initial stop, a vowel, closure, and a final fricative, and represents the longest stimulus on the continuum. Subsequent stimuli were created by excising 10 ms of silence from the closure period. There were seven steps on this con-

tinuum: the shortest stimulus, DISH, had a 0 ms closure duration, whereas the longest stimulus, DITCH, had a closure duration of 60 ms. The total duration of the shortest stimulus, DISH, was 483 ms, and that of DITCH was 543 ms.

The final continuum, BEAT/WHEAT, varied the transition duration of the initial consonant. The natural production of WHEAT, minus the first 50 ms following stimulus onset, constituted one endpoint of the continuum. This stimulus was clearly perceived as "Wheat" by a group of pilot listeners. The duration of the transition, measured from the onset of the stimulus to the steady-state location of the vowel, was 51 ms. The next stimulus was created by excising one glottal pulse of 7–8 ms, taken from the onset of the preceding stimulus. Subsequent stimuli had additional pulses of 7–8 ms removed from the onset of the preceding stimulus, to shorten the duration of the initial transition. There were 7 steps in this continuum. The transition durations of BEAT and WHEAT were 7 and 51 ms, respectively. The endpoint signal, BEAT, had a total duration of 247 ms, and the endpoint signal, WHEAT, had a total duration of 289 ms.

After all of the stimuli were created for each continuum, the rms levels of each stimulus were analyzed. The stimuli were then scaled in level to be equivalent while avoiding any peak-clipping of any single stimulus. A calibration tone was also created that was equivalent in level to the rms levels of the stimuli. Each stimulus from each continuum was stored on the computer for later retrieval during online stimulus presentation and data collection.

C. Procedure

The four stimulus continua were presented to the listeners using three different experimental paradigms: identification, two-interval discrimination, and three-interval discrimination. The identification and two-interval (same/different) discrimination tasks were designed to be somewhat comparable to standard identification and discrimination experiments using speech continua (e.g., Pisoni and Lazarus, 1974; Strouse *et al.*, 1998). The three-interval discrimination task was intended to be comparable to the paradigm typically used in psychoacoustic experiments conducted in our laboratory.

Each of the experimental paradigms was implemented using experiment generator and controller software (ECos2a, Avaaz Innovations). The identification paradigm followed a standard procedure in which the stimuli within each continuum were presented in 10 blocks, with a random order presented in each block, for a total of 10 presentations of each stimulus. Following each stimulus presentation, the computer monitor displayed the written words of the two endpoint stimuli, with each written word occupying one half of the screen in a vertical display. Listener responses were self-paced, and consisted of placing a mouse arrow on either the left or right half of the monitor to identify the stimulus word (using the left mouse click). Listeners were encouraged to guess if they were unsure of the stimulus perceived. The interval between the listener's response and the subsequent stimulus presentation was 4 s. Pilot testing with several older listeners indicated that the computer response was accomplished easily. There were four identification conditions, corresponding to the four stimulus continua.

The two-interval discrimination paradigm (A-X) presented two stimuli on each trial with a 500-ms interobservation interval, and required listeners to judge whether the two stimuli were the same or different. The first stimulus was the standard endpoint stimulus and the second stimulus was the comparison stimulus. The standard stimulus was chosen to be the shortest endpoint stimulus for each continuum. The comparison stimulus presented in the initial trial was the longest possible comparison stimulus relative to the standard (i.e., the opposite endpoint stimulus on the continuum). The listener's task was to identify whether the two stimuli were the same or different, by clicking on one of these two terms displayed on a computer monitor using the computer mouse. Feedback was not provided. Following two correct responses, the comparison stimulus selected was one step closer to the standard stimulus in the continuum. The procedure continued following the adaptive rule, which stipulated a decrease in stimulus duration following two consecutive correct responses and an increase in stimulus duration following each incorrect response (Levitt, 1971). Each change in duration was always one step, with the step size fixed for each continuum, as described previously. This adaptive procedure continued until there were eight reversals in stimulus selection. The mean of the midpoint of the final four reversals was calculated to determine the discrimination threshold of 70.7% correct discrimination. The stimulus pairs included two identical stimuli (i.e., "catch" trials) in order to facilitate threshold determination for listeners who displayed perfect discrimination performance. If the listener displayed perfect performance (i.e., they identified Stimuli 1 and 2 as different, and correctly identified the two identical endpoint stimuli on the catch trials as the same), then a nominal discrimination threshold of 1 (as in step #1, JND=0 steps) was assigned to that run. Three of these trial-run threshold estimates were collected from each subject for each of the four, two-interval discrimination conditions. An analysis was made of the threshold results to determine if stable performance was achieved between trial runs 2 and 3. Stable performance was defined as two threshold estimates that varied by less than 1 step. If stable performance was not observed, additional trial runs were conducted in an effort to observe stable performance, with a maximum of five trial runs presented. Thresholds values, collected in relative step sizes, were later converted to ms, based on the duration corresponding to the step size for each continuum.

The adaptive three-interval paradigm is a three-interval cued two-alternative forced-choice discrimination procedure. Each listening trial consisted of three observation intervals with an interobservation interval of 500 ms. The first interval on each trial contained the reference signal for a given condition, and the second and third intervals contained the reference and comparison signals in either order with equal probability. The listening intervals of each trial were marked with a visual display. Subjects responded with a mouse click to identify the interval (either 2 or 3) corresponding to the stimulus that was different from the standard stimulus pre-

sented in the first interval. The adaptive rule described for the two-interval discrimination paradigm was followed for the three-interval discrimination paradigm to derive the 70.7% correct discrimination threshold. A minimum of three and a maximum of five trial-run threshold estimates were presented to enable individual listeners to reach stable performance.

The order of the three paradigms (identification, two-interval discrimination, three-interval discrimination) was randomized across subjects. A strategy was followed to facilitate the listener's retention of the task associated with each paradigm, and to avoid confusion. To that end, all of the speech continua were presented in a given paradigm before introducing the next paradigm. Additionally, the order of presentation of the four speech continua was randomized for each paradigm, and with a different randomization for each subject. Prior to commencing each experimental paradigm, listeners were provided with some listening practice with the same paradigm, using a novel continuum of speech stimuli. During the experiments, the stimuli were routed from the laboratory computer and sound card to an amplifier (Crown D75A), an audio mixer-amplifier (Colbourn S82-24), and delivered to the listener's better ear through a single insert earphone (Etymotic ER3A) at 85 dB SPL. For the noise-masked young listeners, the CD containing the shaped noise was played back on a compact disk player (Tascam CD-160), amplified (Crown D-75A), attenuated (Hewlett-Packard 350 D), and mixed through the audio-mixer amplifier with the speech signal. The masking noise was presented at an overall level of 50 dB SPL, which was determined to be sufficient to shift the listener's thresholds to the desired levels on the basis of pilot testing. The stimuli and masking noise were calibrated daily.

Listeners were seated in a sound-attenuating chamber for all of the experiments. They were tested in 2 h sessions, with frequent breaks. The entire procedure, including preliminary audiometric assessment and practice sessions, was completed in approximately 6 h. Listeners were reimbursed for their participation in the experiment.

III. RESULTS

Each listener's identification functions were reviewed to ensure that both target endpoint stimuli for each continuum were identified at a probability higher than chance level. For three of the continua, a few listeners failed to identify both endpoint stimuli on the continuum. The data for these listeners in these conditions were removed from subsequent data analysis (2 listeners for BUY/PIE, 3 listeners for WHEAT/WEED, and 3 listeners for BEAT/WHEAT; these listeners

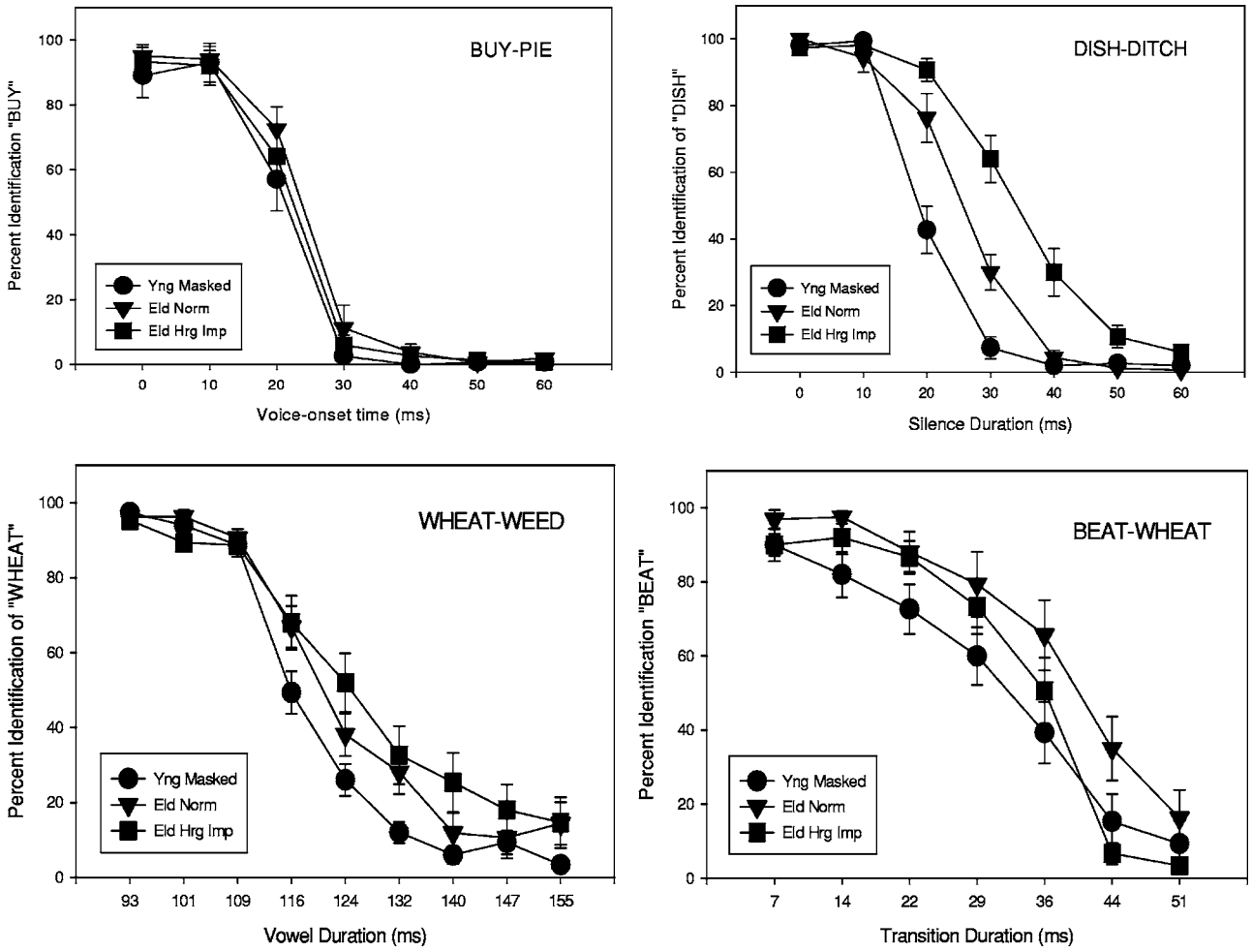


FIG. 1. Identification functions for the three listener groups for the BUY-PIE continuum (top left panel), DISH-DITCH continuum (top right panel), WHEAT-WEED continuum (bottom left panel), and BEAT-WHEAT continuum (bottom right panel). Error bars show one standard error of the mean.

TABLE II. Crossover points for the relevant acoustic cue, in ms, of the three subject groups for the four speech continua. Data shown are group means, with standard deviations in parentheses.

	Speech continuum			
Group	BUY/ PIE	DISH/ DITCH	WHEAT/ WEED	BEAT/ WHEAT
Young masked	21.39 (4.37)	20.00 (4.15)	118.78 (3.34)	28.56 (10.92)
Elderly normal	22.91 (4.43)	25.31 (6.24)	121.57 (5.56)	37.33 (7.69)
Elderly hearing impaired	22.36 (3.43)	34.77 (6.00)	122.97 (7.83)	33.87 (6.69)

were distributed across the three groups). The average identification functions of the remaining listeners from the three groups for each of the four speech continua are shown in Fig. 1. Individual listener identification functions for each continuum were analyzed separately for the slope (a) and y-intercept (b) values by performing a linear regression using data points on the linear portion of the function (between 80% and 20% correct, approximately). The number of points used to perform the linear regression varied between subjects both within and across continua. The 50% crossover point, x, for each function was subsequently calculated using the formula

$$x = (50 - b)/a.$$

The average crossover points for the three subject groups across the four speech continua are shown in Table II. Separate one-way analyses of variance (ANOVA) were conducted for each speech continuum to determine the effect of listener group on the crossover points. Significant group effects were observed for the DISH/DITCH ($p < 0.01$) and BEAT/WHEAT ($p < 0.05$) continua, but not for the BUY/PIE or the WHEAT/WEED continua ($p > 0.05$). Follow-up analyses of the significant group effects were conducted with the Bonnferoni method. The crossover values were significantly different between each pair of groups for the DISH/DITCH continuum. For the BEAT/WHEAT continuum, the crossover values were significantly different for the young noise-masked listeners and the elderly normal-hearing listeners.

The slope data for the three subject groups, derived from the identification functions, are shown in Table III. A higher slope value suggests a clearer distinction between one speech sound and the contrasting speech sound in a continuum. A review of Table III suggests that large standard deviations are observed in the slope data, particularly for the BUY/PIE and BEAT/WHEAT continua. ANOVAs were conducted on the slope values, using a one-way design, and the results failed to reveal a significant main effect of listener group for the BUY/PIE, WHEAT/WEED, and BEAT/WHEAT continua. However, the group effect was significant for the DISH/DITCH continuum, with the young noise-masked listeners showing a larger slope value than the elderly hearing-impaired listener group ($p < 0.05$). In addition, the overall rank order across the BUY/PIE and WHEAT/WEED continua follows a similar pattern in which the young noise-masked group had steeper slopes than the two elderly groups.

Initial analysis of the discrimination results was directed toward evaluation of the reliability of the data. To that end, thresholds of all subjects were compared across the trial blocks for each condition. For most subjects, no significant changes in discrimination performance were observed after three trial blocks. A few subjects required up to five practice blocks before stable performance was observed, but there were no observable differences in the training required of young and elderly subjects. The discrimination threshold for each subject was taken as the average threshold from the final three trial blocks, which is a comparable procedure to that used in our psychoacoustic experiments. Discrimination thresholds were converted to difference limens (DLs), relative to the duration of the reference signal.

As data were collected, it became apparent that the discrimination thresholds were quite different for the two- vs the three-interval procedures, for all listener groups. Figure 2 presents the DLs obtained in the two procedures for the three listener groups, across the four different speech continua. Repeated measures ANOVAs were conducted separately for each continuum using the DL data, with one within-subjects factor (discrimination procedure) and one between-subjects

TABLE III. Average absolute slope values (% change in recognition performance/step; step size varies for each continuum) for the three subject groups for the four speech continua. Standard deviations are shown in parentheses.

	Speech continuum			
Group	BUY/ PIE	DISH/ DITCH	WHEAT/ WEED	BEAT/ WHEAT
Young masked	69.1 (24.1)	59.0 (18.1)	39.2 (17.6)	36.5 (21.7)
Elderly normal	67.7 (26.8)	48.0 (12.4)	29.6 (14.8)	47.2 (15.7)
Elderly hearing impaired	57.6 (25.2)	40.8 (19.6)	26.8 (10.8)	48.7 (25.7)

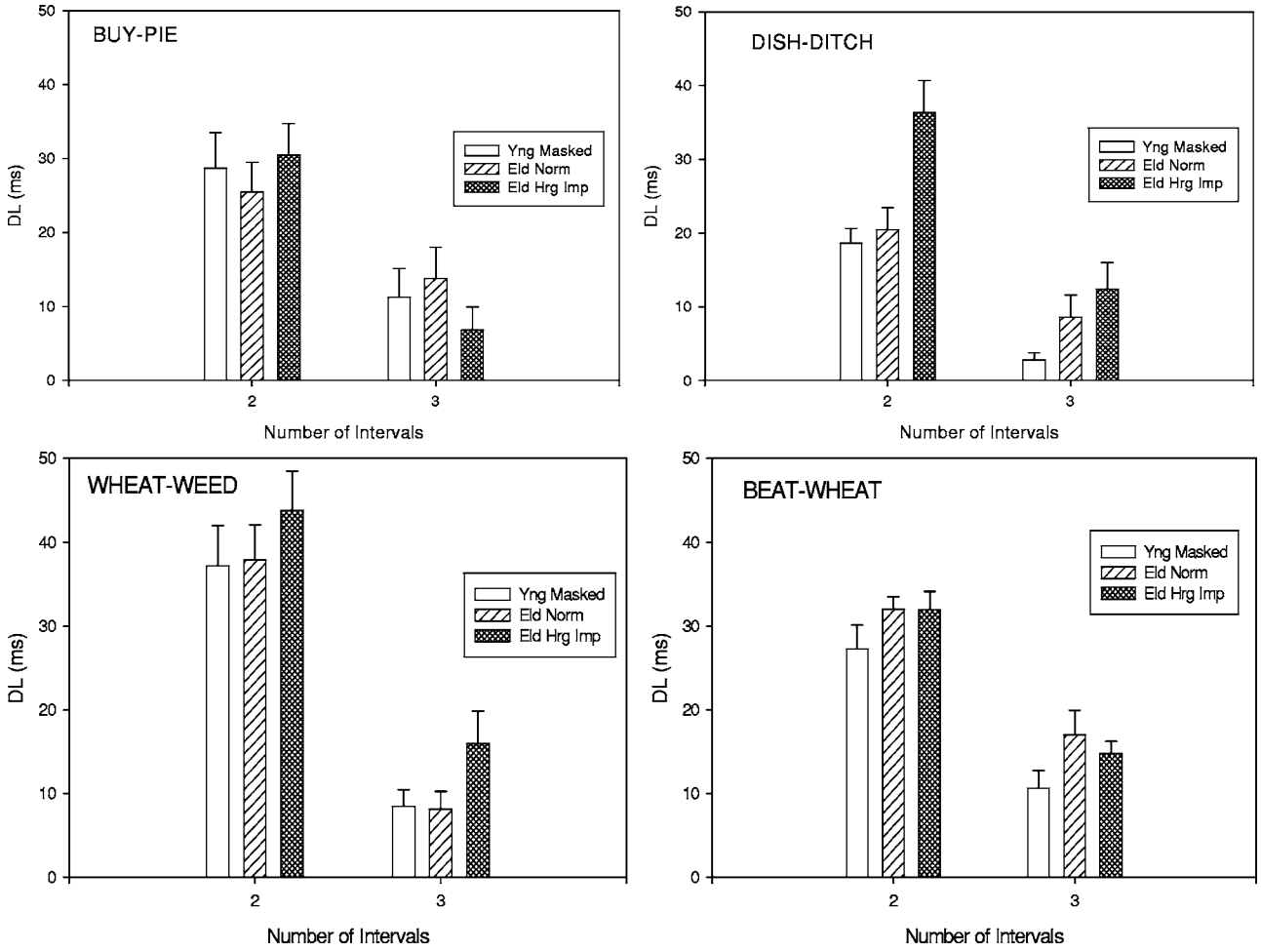


FIG. 2. Mean discrimination thresholds (DLs), in ms, for the three listener groups in the two-interval same-different paradigm and the three-interval, cued, two-alternative forced choice paradigm, for the BUY-PIE continuum (top left panel), DISH-DITCH continuum (top right panel), WHEAT-WEED continuum (bottom left panel), and BEAT-WHEAT continuum (bottom right panel). Error bars show one standard error of the mean.

factor (listener group). Results revealed a significant main effect of procedure (number of intervals) for all four speech continua ($p < 0.01$), with smaller DLs measured consistently for the three-interval procedure compared to the two-interval procedure. The main effect of listener group was significant for the DISH/DITCH continuum only ($p < 0.01$). There were no interactions between listener group and procedure for this continuum. Multiple comparison testing indicated that the elderly hearing-impaired group had higher average DLs than the young noise-masked group and the elderly normal-hearing group, on this continuum.

IV. DISCUSSION

A. Identification performance

The principal experimental question concerned whether or not age-related differences would be observed on identification judgments of modified natural speech tokens that varied along a single temporal parameter to represent two phonetic categories. The tokens of four different contrasting speech continua were presented in an effort to identify specific temporal cues that are most sensitive to auditory aging. Results from the varying experiments generally support the notion that age does influence perception of specific temporal acoustic cues in natural speech tokens, but these findings were not uniform across the different speech continua.

The DISH/DITCH continuum revealed the strongest age-related findings. The variable duration cue in this continuum was the silent interval that preceded the final sibiliant /ʃ/, as a cue for the sibilant/affricate distinction, ranging from 0 ms (/ʃ/) to 60 ms (/tʃ/). The total duration of the endpoint stimulus DISH was 483 ms, and was the longest reference duration of all of the continua presented in this experiment. Crossover values for the young masked, elderly normal-hearing, and elderly hearing-impaired listeners were 20 ms, 25.31 ms, and 34.77 ms, respectively. These findings provide additional information regarding phonetic boundaries, and are consistent with the 30 ms boundary reported previously for young normal-hearing listeners, in judging stimuli from a SHOP/CHOP continuum that varied the duration of a silent interval inserted between the offset of the carrier phrase "please say" and the onset of the sibilant /ʃ/ in "shop" (Dorman *et al.*, 1979). Statistical analyses of the current data showed that the crossover values for the young listeners were significantly different from those of the two elderly groups, confirming the presence of a significant age effect. Thus, elderly listeners required a longer silent interval than younger listeners to identify the stimulus as DITCH. This

finding supports the hypothesis that elderly listeners require longer duration acoustic cues than younger listeners for accurate recognition of target speech signals. One possible reason for the identification performance differences between the younger and older listeners was the use of noise masking for the younger listeners, which might have affected their perception of the silent interval cue distinguishing DISH from DITCH. To address this issue, an analysis was conducted to compare the identification performance of the 15 young listeners with noise masking to that of 11 young listeners with normal hearing (without masking) who participated in pilot testing with the same stimuli and task. The results revealed no significant differences in either crossover points ($t=0.803, p>0.10$) or slopes ($t=0.14, p>0.10$). Thus, it appears that the low level of noise masking presented to the young listeners did not affect their perception of the stop closure that distinguished stimuli along the DISH/DITCH continuum. Additionally, the crossover values for the elderly hearing-impaired group were significantly larger than those for the elderly normal-hearing group, indicating that the presence of hearing loss among elderly listeners also affected performance. A temporal processing deficit was not predicted for the hearing-impaired listeners, particularly for relatively long baseline duration stimuli. However, the target acoustic cues in the DISH/DITCH distinction involve relatively high frequency, weak energy signals that often are not perceived accurately by hearing-impaired listeners (Owens *et al.*, 1972). It is possible that the distorted perception of the final sibilant by the elderly hearing-impaired listeners obscured their ability to utilize the silent interval cue effectively for the sibilant/affricate distinction. This suggestion is supported somewhat by the finding of significant correlations between the DISH/DITCH crossover points and the listeners' high frequency thresholds ($r=0.558, p<0.01$ at 2000 Hz; $r=0.718, p<0.01$ at 4000 Hz). The slope data for the DISH/DITCH continuum suggest that the young listeners demonstrated steeper identification function slopes (59.0) than those observed for the elderly normal-hearing listeners (48.0) and elderly hearing-impaired listeners (40.8). Statistical analyses partially supported this observation, and revealed a significant group effect in which the slopes of the young noise-masked listeners were significantly steeper than those observed for the elderly hearing-impaired listeners. The group effects shown for both the crossover points and the slopes for this continuum strongly suggest that older listeners with and without hearing impairment have difficulty perceiving the silence duration cue.

The present finding of an age effect for phoneme category boundaries on a continuum that varies silence duration is somewhat different from results reported by Dorman and co-workers (1985). In that investigation, listeners identified stimuli in a "slit-split" continuum in which the duration of a silent interval inserted between the sibilant noise and the vocalic portion was varied from 20 to 120 ms. Young normal-hearing listeners showed a shorter category boundary than elderly hearing-impaired listeners; however, the variability in performance of the elderly normal-hearing and elderly hearing-impaired listener groups obscured possible age effects in that investigation. In another investigation, which made use of a "rabid" to "rapid" 4-step continuum, older listeners required longer silence durations than younger listeners to perceive the "rapid" tokens (Price and Simon, 1984). The range of silence closure employed in this 4-step continuum was 35–125 ms. Thus, there appears to be converging evidence from several different investigations and different speech samples that elderly listeners require longer silent intervals to perceptually classify stimuli that are cued by a relatively longer silent interval.

Age-related differences were observed for the BEAT/WHEAT continuum. The varying acoustic parameter in this continuum was the duration of the formant transitions, which serves as a cue for the voiced stop/glide distinction. Liberman *et al.* (1956) showed that listeners perceived a voiced stop for formant transitions <40 ms, and perceived a glide for formant transitions >40 ms, in a continuum of synthetic tokens of /bɛ/ to /wɛ/. In the current experiment, the formant transition duration for the natural token BEAT was 7 ms, and that of WHEAT was 51 ms. Crossover points were 28.56 ms (young noise-masked listeners), 37.33 ms (elderly normal-hearing listeners), and 33.87 ms (elderly hearing-impaired listeners), suggesting that the change in percept from the voiced stop to the glide occurred at relatively brief formant transition durations for the young listeners, and that the elderly listeners required a longer formant transition duration to perceive a glide. Statistical analyses indicated a significant age effect, with the young noise-masked listeners exhibiting smaller crossover values than the elderly normal-hearing listeners. This finding also tends to support the hypothesis of an age-related deficit in processing brief temporal cues for speech. The slope data for the BEAT/WHEAT continuum did not show significant group effects, suggesting that the change in perceived category boundaries occurred at a similar rate for the three groups, despite a difference in the crossover value required by younger and older listeners to alter their stimulus identification. Dorman *et al.* (1985) did not observe an age effect for identification judgments of synthetic stimuli that constituted a /bæ/ to /wæ/ continuum, created by varying the duration of the initial formant transitions from 40 to 90 ms. The stimuli in the Dorman *et al.* synthetic continuum varied in several important ways from the natural stimuli used in the current experiment, including the type of stimuli (synthetic vs natural), number of formants (two formants in the Dorman *et al.* study vs three or more formants in the present study) and the range in duration of the formant transitions (40–90 ms in the Dorman *et al.* study vs 7–51 ms in the present study).

Group effects were not observed for the BUY/PIE and the WHEAT/WEED continua. The temporal cue that varied in the BUY/PIE continuum was VOT, or the interval from the burst release to the onset of voicing, which ranged from 0 to 60 ms. Identification functions of the three listener groups revealed crossover points of approximately 21 ms, 23 ms, and 22 ms for the young noise-masked, elderly normal, and elderly hearing-impaired listeners, respectively, which are generally consistent with those observed in other studies with young normal-hearing listeners where the VOT continuum consisted of bilabial stops (Lisker and Abramson, 1967; Pisoni and Lazarus, 1974). Statistical analyses showed

no group differences for the identification function crossover points and the slopes. Strouse *et al.* (1998) also did not find an age effect for phonetic boundaries in identification functions for a synthetic /ba/ to /pa/ continuum varying in VOT, although slope values were significantly different between younger and older groups. The divergent findings regarding the slope values between the Strouse *et al.* investigation and the present study could be attributed, at least in part, to the use of synthetic vs natural speech in creating the VOT continua. Additionally, the large variance in slope values for the BUY/PIE continuum in the present investigation may have been great enough to obscure possible differences in the means of these groups.

The WHEAT/WEED continuum was comprised of stimuli with varying vowel duration to cue final stop consonant voicing. Peterson and Lehiste (1960) reported that vowel duration in naturally produced English words averaged 197 ms preceding a voiceless final stop consonant and 297 ms preceding a voiced stop, with a ratio of vowel before voiceless consonant to vowel before voiced consonant of 2:3. The perceptual relevance of vowel duration as a cue to word final voicing has been shown for word-final stops, fricatives, and consonant clusters (Raphael, 1972). The absolute vowel durations required for a change in the percept of final consonant voicing depend on the specific vowel used; hence, the vowel durations in the current continuum are not readily comparable to those reported in other studies. For example, the vowel in "kit" vs "kid" is shorter than the vowel in "bat" vs "bad." In the present experiment, crossover values corresponded to vowel durations of 118.78, 121.57, and 122.97 for the young masked, elderly normal, and elderly hearing-impaired listener groups, respectively. Although the average data suggest a tendency for young listeners to show a shorter crossover point and steeper slope for this continuum than the two older groups, statistical analyses failed to reveal any significant differences in the crossover points or the slope values between groups for the vowel duration cue. The slope values for the WHEAT/WEED continuum ranged from 26.8 to 39.2, suggesting that listeners did not perceive a strong contrast between the two tokens in this continuum. Indeed, our pilot testing during stimulus development dictated that we employ a 9-step continuum, rather than a 7-step continuum, to ensure that listeners identified the endpoint stimuli with reasonable accuracy. The identification function for WHEAT/WEED, shown in Fig. 1, suggests that the young noise-masked listeners achieved nearly perfect accuracy in identifying the endpoint stimuli; however, the two older listener groups did not show perfect identification of the endpoint stimulus "WEED." This lack of a clear percept of the exemplar word stimulus by older listeners may have contributed to more variable results and a resulting inability to observe significant group effects for this continuum.

A comparison of the identification data across the different speech continua reveals some interesting observations. First, significant age effects, when observed, are associated with smaller crossover values for younger listeners than for older listeners. This finding tentatively indicates that young listeners require relatively brief alterations in target stimulus duration to alter their phonetic percept. Second, the strongest age-related differences are observed for the continuum with the longest baseline duration (483 ms). All other continua were characterized by baseline reference durations of 310 ms or less. Third, the type of duration cue, filled or unfilled, does not appear to be linked consistently with the presence or absence of age effects in the identification data. For example, age effects were observed for identification of the BEAT/WHEAT continuum which requires processing of a filled cue (transition duration), but age effects were not observed for the WHEAT/WEED continuum which requires processing of another filled cue (vowel duration). Moreover, age effects were observed for the DISH/DITCH continuum which requires processing of the duration of an unfilled cue (silence). Fourth, the location of the target cue within the word is not clearly associated with age-related effects. The target cues for DISH/DITCH and WHEAT/WEED are located in the medial position of the words, but age-related effects were observed for the former continuum only. The target cues for BUY/PIE and BEAT/WHEAT are located in the initial position of the words, but age-related effects were observed for the latter continuum only. Taken together, the significant group findings in the present experiment were related to temporal cues for consonant manner and not to temporal cues for consonant voicing. Other factors, such as cue position, reference duration, and filled vs unfilled cue interval, are not consistently related to the observation of age effects in identification of the limited set of speech contrasts examined in the present experiments.

B. Discrimination performance

Discrimination performance was measured with two different paradigms to facilitate comparison of the current results with those obtained by other investigators. The two-interval paradigm was intended to be relatively comparable to a standard same-different speech discrimination paradigm, whereas the three-interval paradigm was designed to be analogous to psychoacoustic measurement with nonspeech signals. Discrimination thresholds obtained with the two paradigms were quite different; statistical analyses revealed that the thresholds measured with the three-interval paradigm were significantly smaller than those measured with the two-interval paradigm for all listener groups across all four speech continua (Fig. 2). Comparable findings were reported by Pisoni and Lazarus (1974) in a study comparing performance on two different discrimination tasks intended to elicit categorical vs noncategorical perception. Listeners in the current experiments provided anecdotal reports suggesting that same-different procedures are performed on the basis of stimulus identification whereas three-interval forced choice procedures are based on acoustic differences between speech stimuli. Additionally, the discrimination thresholds measured for the two-interval paradigm are generally similar in value to the crossover points measured in the identification paradigm, lending further support to the notion that listeners judged the identity of the two stimulus words while performing this discrimination task.

The three-interval results tend to approximate some of the relevant findings from psychoacoustic experiments. A possible psychoacoustic analog for the DISH/DITCH continuum is the measurement of the threshold of a silent gap inserted between two markers. Detection thresholds for a silent gap inserted in the midpoint of a broadband marker are approximately 2–3 ms for young listeners with normal hearing (Plomp, 1964). The average threshold value measured for young listeners for the DISH/DITCH continuum in the current experiment is 2.77 ms, which has the same approximate value as that observed for nonspeech broadband stimuli.

The three-interval discrimination task for the WHEAT/WEED and BEAT/WHEAT continua is similar to a duration discrimination measure for filled (tonal or noise) stimuli. Duration discrimination thresholds for signals of approximately 250 ms are about 25 ms, or 10% of the reference duration on average, for young listeners (Small and Campbell, 1962; Abel, 1972). The DLs measured for the WHEAT/WEED continuum range from 8.15 ms to 15.95 ms, and those measured for the BEAT/WHEAT continuum range from 10.63 ms to 14.78 ms. Given that the reference duration of the endpoint speech stimulus for each of these continua is approximately 250 ms, the listeners in these experiments demonstrate better DLs than those that have been observed for nonspeech signals of equivalent reference duration. For the WHEAT/WEED continuum, the relevant reference duration may be the vowel duration cue (93 ms) rather than the entire word stimulus, in which case the results generally conform to the expected DL of 10%. The reference transition duration cue for BEAT/WHEAT is only 7 ms. Duration discrimination for such brief reference signals are not commonly examined in psychoacoustic studies but can be expected to be considerably larger than the 10% value associated with longer references (Abel, 1972). Generally, these findings suggest that in the three-interval paradigm, listeners based their judgments on discrete acoustic differences in the stimuli, rather than on labels assigned to the three stimuli presented on a trial. Moreover, their performance appears to follow trends observed in psychoacoustic data for analogous stimuli.

Performance patterns for the different listener groups derived from the two different discrimination paradigms were somewhat consistent with those observed in the identification paradigm. An examination of Fig. 2 suggests that the elderly hearing-impaired group showed larger discrimination thresholds than the other groups for the DISH/DITCH continuum. There also appears to be an age-related trend for larger discrimination thresholds in the BEAT/WHEAT continuum. Statistical analyses confirmed a significant main effect of listener group for the discrimination thresholds measured for the DISH/DITCH continuum, but not for any other continuum. Multiple comparison testing of the group effect revealed that the DLs for the elderly hearing-impaired listeners were greater than those for the two normal-hearing groups for both the two- and three-interval discrimination procedures. This finding agrees with the hearing loss effect observed in the analysis of the identification functions obtained for the DISH/DITCH continuum. The absence of pure age effects in the discrimination data differed from results observed in gap detection studies conducted with nonspeech stimuli (e.g., Schneider *et al.*, 1994; Snell, 1997). However, the age effects observed previously for gap detection are relatively small (<3 ms difference) and require small stimulus changes in the measurement procedures to be observed. It is possible that the larger step size employed in the present discrimination procedures obscured observation of possible age effects in these conditions. Other stimulus variables differentially influence gap detection performance by younger and older listeners, such as location of the gap in the stimulus (He *et al.*, 1999) or the location of the stimulus itself within a sequential context (Fitzgibbons and Gordon-Salant, 1995). In these more complex conditions, older listeners generally exhibit substantially larger temporal discrimination thresholds than those measured for stimuli presented in isolation. Further evaluation of younger and older listener's discrimination performance for temporal acoustic cues in speech may also reveal substantial age-related differences when the speech targets are embedded within extended sentence-length stimuli.

V. SUMMARY AND CONCLUSIONS

This investigation of identification and discrimination for natural speech stimuli varying in temporal acoustic cues by young and elderly listeners revealed the following:

(1) Older listeners show different identification functions than younger listeners for some, but not all, temporally based speech continua. Age effects were observed for the speech continua that varied acoustic temporal cues underlying the perception of consonant manner of articulation. These continua included DISH/DITCH, in which silence duration as a cue for the sibilant/fricative distinction was varied, and BEAT/WHEAT in which the transition duration for the voiced stop/glide distinction was varied. The present experiment did not reveal age effects for speech continua that varied acoustic cues for stop consonant voicing in either the initial or final position of a word. These continua included BUY/PIE, in which VOT served as the primary cue for initial consonant voicing, and WHEAT/WEED, in which vowel duration was the principal cue for final consonant voicing.
(2) For speech continua in which an age-related difference was observed, younger listeners demonstrated shorter duration crossover points than older listeners. These findings suggest that older listeners require longer target temporal acoustic cues to form judgments about stimulus identity.
(3) Discrimination thresholds for the three-interval task were dramatically lower than for the two-interval task. These results indicate that listeners may be able to distinguish fine acoustic differences in speech that are not perceptible to them in a same-different paradigm. This increased acuity may be due, in part, to the listener's knowledge that one of the two comparison stimuli must be different from the initial reference stimulus.
(4) Discrimination thresholds for the temporal acoustic cues in speech agree reasonably well with discrimination thresholds obtained for nonspeech signals, when similar

paradigms are used. The results suggest that all listener groups are quite sensitive to subtle temporal changes in speech when a labeling judgment is not required and the target speech signals are presented in isolation. Although age-related effects were not observed in these preliminary discrimination data, a more perceptually relevant test would involve discrimination judgments for target speech signals presented in the context of ongoing sequential speech.

(5) Effects of hearing loss were observed consistently for the DISH/DITCH speech continuum, when presented in the identification paradigm and both discrimination paradigms. The source of the hearing loss effect for the silence duration cue may be associated with the hearing-impaired listeners' diminished capacity to detect the onset of the final sibilant in these speech tokens, because of the loss of high frequency audibility.

In conclusion, the findings support the hypothesis that older listeners experience a subtle deficit in processing changes in some acoustic temporal cues for discrete speech signals. For these signals, the data presented in this report provide further evidence that older listeners exhibit deficits in auditory temporal processing, and that these deficits contribute to alterations in speech identification.

ACKNOWLEDGMENTS

This research was supported by an individual research grant (R37AG09191) from the National Institute on Aging. The authors are grateful to Michele Spencer for her assistance in data collection and analysis, and to two anonymous reviewers for their valuable comments on an earlier version of this manuscript.

Abel, S. (**1972**). "Duration discrimination of noise and tone bursts," J. Acoust. Soc. Am. **51**, 1219–1223.

Adams, C., and Munro, R. R. (**1978**). "In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterance of some native and non-native speakers of English," Phonetica **35**, 125–156.

ANSI (**1996**). ANSI S3.6-1996, "American National Standard Specification for Audiometers" (American National Standards Institute, New York).

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Burda, A. N., Scherz, J. A., Hageman, C. F., and Edwards, H. T. (**2003**). "Age and understanding speakers with Spanish or Taiwanese accents," Percept. Mot. Skills **97**, 11–20.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Diehl, R. (**1976**). "Feature analyzers for the phonetic dimensions 'stop vs continuant,' " Percept. Psychophys. **19**, 267–272.

Dorman, M. F., Marton, K., Hannley, M. T., and Lindholm, J. M. (**1985**). "Phonetic identification by elderly normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 664–670.

Dorman, M. F., Raphael, L. J., and Liberman, A. M. (**1979**). "Some experiments on the sound of silence in phonetic perception," J. Acoust. Soc. Am. **65**, 1518–1532.

Fitzgibbons, P., and Gordon-Salant, S. (**1994**). "Age effects on measures of auditory duration discrimination," J. Speech Hear. Res. **37**, 662–670.

Fitzgibbons, P. J., and Gordon-Salant, S. (**1995**). "Duration discrimination with simple and complex stimuli: Effects of age and hearing sensitivity," J. Acoust. Soc. Am. **98**, 3140–3145.

Fitzgibbons, P., and Gordon-Salant (**1998**). "Auditory temporal order perception in younger and older adults," J. Speech Lang. Hear. Res. **41**, 1052–1060.

Flege, J. E., and Eefting, W. (**1988**). "Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation," J. Acoust. Soc. Am. **83**, 729–740.

Fox, R. A., Flege, J. E., and Munro, J. (**1995**). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," J. Acoust. Soc. Am. **97**, 2540–2551.

Gelfand, S., Schwander, T., and Silman, S. (**1990**). "Acoustic reflex thresholds in normal and cochlear-impaired ears: Effect of no-response rates on 90th percentiles in a large sample," J. Speech Hear Disord. **55**, 198–205.

Gordon-Salant, S., and Fitzgibbons, P. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.

Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res. **36**, 1276–1285.

Guion, S. G., Flege, J. E., Liu, H., and Yeni-Komshian, G. (**1997**). "Sentence duration as an index of overall proficiency in an *L2*," J. Acoust. Soc. Am. **101**, 3128.

He, N.-J., Horwitz, A. R., Dubno, J. R., and Mills, J. (**1999**). "Psychometric functions for gap detection in noise measured from young and aged subjects," J. Acoust. Soc. Am. **106**, 966–978.

Klatt, D. H. (**1975**). "Voice onset time, frication, and aspiration in word initial consonant clusters," J. Speech Hear. Res. **18**, 686–706.

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

Liberman, A. M., Delattre, P. C., Gerstman, L. J., and Cooper, F. S. (**1956**). "Tempo of frequency change as a cue for distinguishing classes of speech sounds," J. Exp. Psychol. **52**, 127–137.

Lisker, L., and Abramson, A. (**1967**). "Some effects of context on voice onset time in English stops," Lang Speech **10**, 1–28.

Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.

Lister, Besing, and Koehnke (**2002**). "Effects of age and frequency disparity on gap discrimination," J. Acoust. Soc. Am. **111**, 2793–2800.

Lister, J., and Tarver, K. (**2004**). "Effect of age on silent gap discrimination in synthetic speech stimuli," J. Speech Lang. Hear. Res. **47**, 257–268.

Luce, P. A., and Charles-Luce, J. (**1985**). "Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production," J. Acoust. Soc. Am. **78**, 1949–1957.

MacKay, I. R. A., Flege, J. E., and Piske, T. (**2000**). "Persistent errors in the perception and production of word-initial English stop consonants by native speakers of Italian," J. Acoust. Soc. Am. **107**, 2802.

Owens, E., Benedict, M., and Schubert, E. D. (**1972**). "Consonant phonemic errors associated with pure tone configurations and certain kinds of hearing impairment," J. Speech Hear. Res. **15**, 308–322.

Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 268–277.

Pfeiffer, E. (**1977**). "A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients," J. Am. Geriatr. Soc. **23**, 443–441.

Pisoni, D. B., and Lazarus, J. H. (**1974**). "Categorical and noncategorical modes of speech perception along the voicing continuum," J. Acoust. Soc. Am. **55**, 328–333.

Plomp, R. (**1964**). "Rate of decay of auditory sensation," J. Acoust. Soc. Am. **36**, 277–282.

Price, P. J., and Simon, H. J. (**1984**). "Perception of temporal differences in speech by "normal-hearing" adults: Effects of age and intensity," J. Acoust. Soc. Am. **76**, 405–410.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., and Lamb, M. (**1994**). "Gap detection and the precedence effect in young and old adults," J. Acoust. Soc. Am. **95**, 980–991.

Small, A. M., and Campbell, R. A. (**1962**). "Temporal differential sensitivity for auditory stimuli," Am. J. Psychol. **75**, 401–410.

Snell, K. B. (**1997**). "Age-related changes in temporal gap detection," J. Acoust. Soc. Am. **101**, 2214–2220.

Snell, K. B., and Hu, H.-L. (**1999**). "The effect of temporal placement on gap detectability," J. Acoust. Soc. Am. **106**, 3571–3577.

Strouse, A., Ashmead, D. H., Ohde, R. N., and Grantham, D. W. (**1998**). "Temporal processing in the aging auditory system," J. Acoust. Soc. Am. **104**, 2385–2399.

Vaughan, N., and Letowski, T. (**1997**). "Effects of age, speech rate, and type of test on temporal auditory processing," J. Speech Lang. Hear. Res. **40**, 1192–1200.

Wingfield, A., Poon, L. W., Lombardi, L., and Lowe, D. (**1985**). "Speed of processing in normal aging: Effects of speech rate, linguistic structure, and processing time," J. Gerontol. **40**, 579–585.

Synchronization of organ pipes: experimental observations and modeling

M. Abel and S. Bergweiler
Institute of Physics, University of Potsdam, Potsdam, 14469, Germany and UP Transfer GmbH at the University of Potsdam, Potsdam, 14469, Germany

R. Gerhard-Multhaupt
Institute of Physics, University of Potsdam, Potsdam, 14469, Germany

(Received 25 May 2005; revised 20 December 2005; accepted 9 January 2006)

We report measurements on the synchronization properties of organ pipes. First, we investigate influence of an external acoustical signal from a loudspeaker on the sound of an organ pipe. Second, the mutual influence of two pipes with different pitch is analyzed. In analogy to the externally driven, or mutually coupled self-sustained oscillators, one observes a frequency locking, which can be explained by synchronization theory. Further, we measure the dependence of the frequency of the signals emitted by two mutually detuned pipes with varying distance between the pipes. The spectrum shows a broad "hump" structure, not found for coupled oscillators. This indicates a complex coupling of the two organ pipes leading to nonlinear beat phenomena. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2170441]

PACS number(s): 43.75.Np, 43.28.-g [NHF] Pages: 2467–2475

I. INTRODUCTION

Sound production in organ pipes is traditionally described as a generator-resonator coupling. In the last decades, research has been concerned with the complex aeroacoustic processes that lead to a better understanding of the sound generation in a flue organ pipe. The process of sounding a flue-type organ pipe employs an airstream directed at an edge, the labium of an organ pipe. An oscillating "air sheet" is used to describe the situation in which the oscillations of the jet exiting from the flue are responsible for the creation of the pipe sound.[1,2] Using the "air sheet" terminology, it is pointed out that the oscillation is controlled not by pressure, as in earlier investigations,[3–7] but by the airflow.[8–13]

The situation becomes more involved if two flue organ pipes are close in sounding frequency and in spatial distance. Then a synchronization of the pipes, a frequency locking, occurs.[14–17] This has a direct importance for the arrangement of organ pipes in a common Orgelwerk. The effect has been known by organ builders for a long time and taken into account intuitively in the design of organs.[18] Some measurements of the acoustic field or of its dependence on the mutual coupling or on the distance between the pipes have been reported first in the early 20th century,[15] but no theoretical explanation was given. In this article, we report acoustic measurements and give a possible explanation by means of modern synchronization theory.[19] To a certain approximation, an organ pipe can be considered as a self-sustained oscillator, explained in detail later. In this context, the synchronization of a pipe by an external, acoustical force is important, such that we carried out additional measurements, where a pipe stands aside a loudspeaker whose frequency is tuned around the pipes pitch. As well in this case, the pipe is found to be synchronized.

As a side result of these measurements, we can clarify a discussion about the nature of the strong amplitude decrease for two coupled pipes, already observed by Lord Rayleigh[14] and later by very detailed measurements of Bouasse.[15] As a possible interpretation, the so-called oscillation death has been given in Ref. 19, which means that there is an oscillation breakdown at the pipe mouth, and all the energy is dissipated. From our results, we rule out such a scenario and suggest an antiphase oscillation that yields destructive interference of the emitted acoustic waves. This holds for two neighboring pipes, as well as for the pipe standing aside a loudspeaker.

This article is structured as follows: In Sec. II we explain briefly the generation of sound in an organ pipe and review basic concepts from synchronization theory. In Sec. III we report on detailed measurements of the synchronization of a loudspeaker positioned directly on the side of a pipe and observe two adjacent pipes. The results are interpreted within the frame of synchronization theory. We give an explanation of why a model of two oscillators works out so well, as indicated in Ref. 16. Further, the dependence of the frequency spectrum on the distance between two detuned pipes is investigated for a fixed amount of detuning. Finally, we conclude with Sec. IV.

II. BASIC PRINCIPLES

A. Sound generation in organ pipes

Sound generation in organ pipes has been repeatedly investigated.[8,12,20] Here, the beauty of musical sound generation is paired with complex aerodynamical phenomena; their coupling to the acoustic field has been understood to a reasonable degree in the last 30 years (see the review[1]).

Let us consider a single organ pipe: The wind system blows with constant pressure producing a jet exiting from the pipe flue. Typical Reynolds numbers, corresponding to a free jet, are of the order of 10^3, depending on the pressure sup-

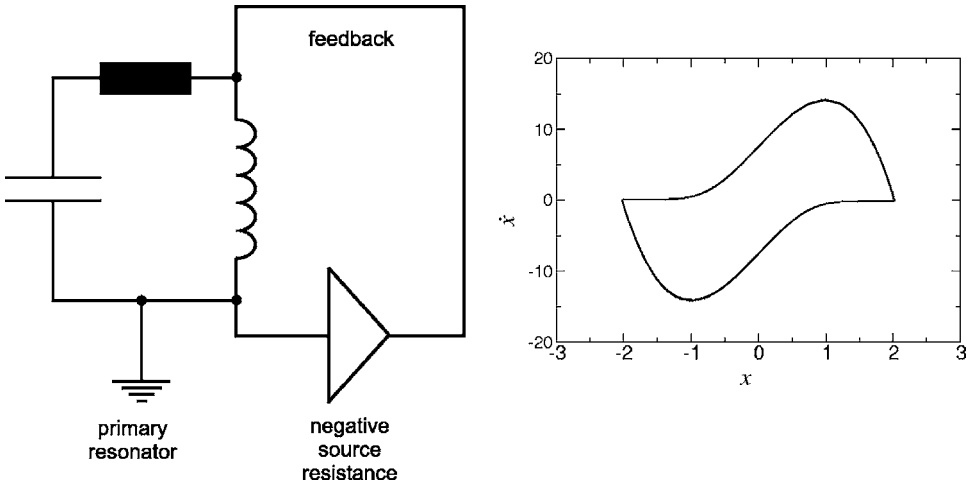


FIG. 1. Left: Schematic diagram for a self-sustained oscillator. Right: resulting limit cycle behavior.

plied and the pipe dimensions; see, Refs. 12 and 21. The jet exiting from the flue generates a pressure perturbation at the labium of the pipe that travels inside the pipe resonator and is reflected at the end of the resonator. This pressure wave returns after time T to the labium, where it in turn triggers a change of the phase of the jet oscillations. After a few transients, a stable oscillation of an "air sheet" at the pipe mouth is established. This oscillation of the wind field couples to the acoustic modes, and a sound wave is emitted. Thus, the system can be considered as a damped linear oscillator (the resonator) that controls the periodic energy supply of the air jet by the air column oscillation. We want to adapt this idea of an organ pipe as a self-sustained oscillator[22,23] to understand the mechanism of mode or frequency locking,[24] or synchronization.

One feature of a self-sustained oscillator is the occurrence of oscillations, even for constant driving. The physical mechanism is the balance of energy losses and energy input. The classical example is the van der Pol oscillator (Fig. 1), where a Triode (or more modern, a tunnel diode) acts as "negative resistance" whose response is fed back to an electrical oscillator. In the case of an organ pipe, energy is supplied by the wind system, losses are due to internal dissipation in the spatially extended resonator, and radiation. The negative resistance is represented by the airstream at the pipe mouth and the feedback coupling is given by the boundary conditions between the "air sheet" wind field and the air-column oscillations in the resonator. It can be described by impedances,[25] detailed results for two coupled Helmholtz resonators are given in Ref. 26. Of course, the exact form of the limit cycle and thus the acoustic oscillations depend on the details of energy losses, as the quality factor of the resonators, the radiation of sound, and the understanding of the energy supply by the coupling to the jet. The general mechanism, however, remains untouched, since only a nonlinear growth of losses and supply is needed for a self-sustained oscillator to work. Here, we do not want to investigate these details, but rather emphasize the general character of our investigations.

The sound emitted by this complex system can be evaluated by the coupling of the flow and the acoustic modes in the framework of aeroacoustic modeling on the basis of Lighthills analogy. According to Howe,[27–29] the sound generation is dominated by the singularity at the edge of the labium. Numerical approaches still suffer from the very expensive runs needed to resolve the large range of excited scales and the proper choice of boundary conditions.[30,31]

B. Synchronization of self-sustained oscillators

In his book, Lord Rayleigh states "When two organ pipes of the same pitch stand side by side…it may still go so far as to cause the pipes to speak in absolute unison in spite of inevitable small differences."[14] He describes the so-called "Mitnahme-Effekt" (loosely the takealong effect), known by organ builders.[16,17] In Ref. 19 the phenomenon is interpreted in the frame of synchronization theory. Hereafter we will use the terms frequency locking or synchronization to be synonymous with the "Mitnahme-Effekt." In Ref. 17 the effect has been analyzed heuristically, while Stanzial has investigated the dependence of the frequency locking on the detuning or frequency difference of two single pipes in Ref. 16 by acoustic measurement. He modeled the effect by two coupled oscillators, without giving a physical reason of the coupling. A treatment of the mode-locking phenomenon, specific for musical instruments is very nicely given in Ref. 23. In the following we will rely on synchronization theory to show that *any* self-sustained coupled oscillator generically shows synchronization.

Probably the most important feature of self-sustainment is the occurrence of an attracting limit cycle. It appears due to two properties: nonlinearity and the energy balance between losses and driving. For a linear, damped oscillator a limit cycle solution does not exist; the only possible attractor is the trivial solution. Nonlinearity allows for the dependence of frequency on amplitude, which constitutes a mechanism to drive the system toward an amplitude at which the regular oscillations are established. Since the amplitude corresponds directly to the mechanical energy of the system, this is right at the point of equality of losses and supply—the limit cycle.

We want to sketch the principles for synchronization of an oscillator with an external driving and then explain the interaction of two oscillators. In the following we will discuss some equations in terms of the angular frequencies and return later to frequencies when the measurements are concerned.

1. Single oscillator

Let us consider the equation for a harmonic oscillator with negative, nonlinear resistance:

$$\ddot{x} = -\omega_0^2 x - \frac{d}{dt} f(x) = -\omega_0^2 x - \dot{x}\frac{df}{dx}, \tag{1}$$

where ω_0 is the angular frequency of the harmonic system and $f(x)$ is a nonlinear function with at least one part of positive and another part of negative slope. Now we discuss the basic principles by the concrete example of the van der Pol oscillator and give general results without a further derivation; for details see Ref. 19. In the case of the van der Pol oscillator $f(x)=\alpha x-\frac{1}{3}\beta x^3$, and $\omega_0^2=(LC)^{-1}$. The energy supply is accounted for by α, the losses by β. If the inharmonicity is not too great, the amplitude on the limit cycle differs little from the harmonic oscillator and the phase coincides approximately with the harmonic one. To get rid of fast oscillations, one applies the method of averaging:[32] One inserts for the amplitude $x(t)=A(t)\sin(\Theta_0)=A(t)\sin(\omega_0 t)$ and collects terms of the different time scales 1 and ω_0. So, one obtains an equation for the slowly varying amplitude $A(t)$, and for the phase Θ_0 here for the van der Pol oscillator:

$$\dot{A} = 1/2\left(\alpha - \frac{1}{4}\beta A^2\right)A, \tag{2}$$

$$\dot{\Theta}_0 = \omega_0. \tag{3}$$

Equation (3) describes the slow relaxation of the amplitude toward the steady state. A limit cycle exists for $|A|^2=4\alpha/\beta$. The solution of the phase equation allows for the addition of an arbitrary constant phase. Thus, one can shift the state along the cycle without changing the energy. This allows for the adjustment of the phase, when the system is driven, or coupled to another oscillator. This adjustment is exactly what we understand as synchronization, or mode locking.

2. Driven, single oscillator

If the system is driven externally by a harmonic force with angular frequency $\omega_1 \simeq \omega_0$, two time scales are present in the system: a fast one $t_f=2\pi/\omega_1 \simeq =2\pi/\omega_0$, and a slow one $t_s=2\pi/(\omega_1-\omega_0)$, and $t_s \ll t_f$. The corresponding equation reads,

$$\ddot{x} + \omega_0^2 x - \dot{x}\frac{df}{dx} = \omega_1^2 R\cos(\omega_1 t). \tag{4}$$

We want to investigate now the dependence on the slow time and average over the fast one. To do so, we use the ansatz $x(t)=A(t)\sin(\omega_1 t+\phi)$, with ϕ the slow phase (it is a bit easier working with a complex amplitude, putting $x=Ae^{i\omega_1 t+\phi}$). By using again the averaging method one obtains, after a few straightforward steps,[19,33]

$$\dot{A} = \frac{1}{2}\left(\alpha - \frac{1}{4}\beta A^2\right)A - \frac{\omega_1 R}{2}\cos\phi, \tag{5}$$

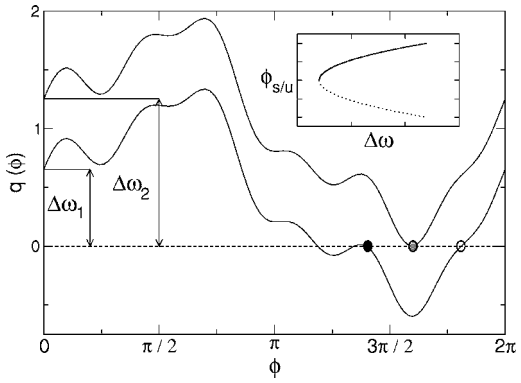


FIG. 2. Stable (filled circle) and unstable (open circle) fixed point Θ_s, Θ_u for the Adler equation to illustrate the mode locking. If detuning or coupling is varied, the phase is adjusted accordingly by the system to the stable fixed point. For $\Delta\omega-\epsilon=0$, both fixed points merge; this results in a saddle-node bifurcation, as sketched in the inset; the straight/dotted line corresponds to the stable/unstable fixed point.

$$\dot{\phi} = \frac{1}{2}\frac{\omega_0^2-\omega_1^2}{\omega_1} + \frac{\omega_1}{2}\frac{R}{A}\sin\phi. \tag{6}$$

The second equation is known as the Adler equation;[19,34] the general form for the phase equation depends on the nonlinear function $f(x)$ and can be seen as the difference of the equations for the two fast variables, ω_1 and ω_0:

$$\begin{aligned} \dot{\Theta}_1 &= \omega_1, \\ \dot{\Theta}_2 &= \omega_0 + \epsilon q(\omega_0-\omega_1), \end{aligned} \qquad \dot{\phi} = \dot{\Theta}_2 - \dot{\Theta}_1 = \Delta\omega + \epsilon q(\phi), \tag{7}$$

with a 2π-periodic function q, and the two parameters detuning $\Delta\omega=\frac{1}{2}[(\omega_0^2-\omega_1^2)/\omega_1] \simeq \omega_0-\omega_1$ and locking term $\epsilon=(\omega_1/2(R/A))$, which is proportional to the driving strength. To zeroth order, one can assume $A \simeq$ const and the phase equation effectively decouples from the amplitude equation, which in turn is driven by the phase; similar holds for a first order.[19] The Adler equation $[q(\phi)=\sin(\phi)]$ has two stationary solutions, $\dot{\phi}=0$, for $|\Delta\omega| \leqslant \epsilon$. The stability of these fixed points, ϕ_s and ϕ_u, is determined by linear stability analysis. One puts $\phi=\phi_{s/u}+\delta\phi$, inserts this expression into (7) and Taylor expands the expressions. The stable point is given by $\cos\phi_s<0$, the unstable one by $\cos\phi_u>0$, corresponding to exponential decay or growth of the perturbation $\delta\phi$. In general, if higher harmonics enter in $q(\phi)$, several fixed points can exist, and the positions can be strongly asymmetric; cf. Fig. 2.

3. The transition to synchronization

If one varies the detuning from zero to ϵ, the phase varies from zero to (at most) π. At the merging, or bifurcation point, $\Delta\omega_{\max}-\epsilon=0$, and the stable and unstable fixed point annihilate (opaque in Fig. 2). This type of bifurcation is known as a saddle-node bifurcation.[35] In the case of an unstable saddle, one attracting and one repelling direction to/from the periodic limit cycle exist; for a stable cycle two attracting directions exist, as for a stable node.

Inside the synchronization region, the phase difference $\phi=\phi_0=$const, and the phase of the observed oscillation is

$\omega_1 t+\phi_0$. Let us turn to the phase dynamics outside the synchronization region. For $\dot{\phi}\neq 0$, one can formally integrate Eq. (7) to obtain

$$t=\left|\int_{\phi_0}^{\phi}\frac{d\phi'}{\epsilon q(\phi')-\Delta\omega}\right|. \tag{8}$$

The period $T=2\pi/\omega_b$ results with the integration bound $\phi=\phi_0+2\pi$. The total phase rotates uniformly $\varphi=\omega_1 t+\phi(t)=\omega_1 t+\omega_b t=\Delta\Omega t$, and one recognizes the typical beat phenomenon, as for the superposition of two harmonic waves; only that here the beating is strongly nonlinear. For harmonic oscillations the phase slip of the beating is distributed over the whole interval. Close to the bifurcation from the synchronized region, one observes long epochs of nearly constant phase $\phi\simeq\phi_{\max}$ broken by relatively short intervals in which the phase rotates by 2π. We observe this behavior as well for coupled organ pipes; this is illustrated next by a plot of the measured time signal close to the synchronized region.

For a saddle-node bifurcation, one expects a square root dependence of $\Delta\Omega$ on the detuning $\Delta\omega$. This indeed by expansion of the denominator in (8):

$$|\Delta\Omega|\simeq 2\pi\left|\int_{-\infty}^{\infty}\frac{d\phi}{\epsilon/2q''(\phi_{\max})\phi^2-(\Delta\omega-\Delta\omega_{\max})}\right|^{-1}, \tag{9}$$

$$=\sqrt{\epsilon|q''(\phi_{\max})|\cdot(\Delta\omega-\Delta\omega_{\max})}\simeq\sqrt{\Delta\omega-\Delta\omega_{\max}}. \tag{10}$$

A very detailed description, including many examples and generalizations, is given in Ref. 19.

4. Two coupled oscillators

Now, we will turn our attention to the case of two oscillators and focus to the equations for the phases. From the above, it is now clear that any pair of uncoupled, self-sustained oscillators close to a limit cycle can be written in terms of phase, (Θ_1,Θ_2), and amplitude, (A_1,A_2), in the following form:

$$\dot{\Theta}_i=\omega_i, \tag{11}$$

$$\dot{A}_i=-\gamma(A_i-A_{i,0}), \tag{12}$$

with $i=1,2$. For weak coupling, one can again perform an expansion and apply the method of slowly varying amplitude such that the phase equations are, analogous to (7),

$$\dot{\Theta}_1=\omega_1+\epsilon G_1(\Theta_1,\Theta_2), \tag{13}$$

$$\dot{\Theta}_2=\omega_2+\epsilon G_2(\Theta_1,\Theta_2). \tag{14}$$

For $\omega_1\simeq\omega_2$ the phase difference $\phi=\Theta_1-\Theta_2$ is a slow variable. The discussion can now follow the above, with the difference that now the oscillators have to adjust themselves mutually, whereas above one oscillator adjusted its frequency to an external driving.

We have to distinguish very carefully and clearly between the *single* angular frequencies ω_i, measured for a *decoupled* system, and the angular frequencies $\dot{\Theta}_i$ measured for the *coupled* system. For a clear nomenclature, we call the difference of the uncoupled frequencies *detuning*, and denote it by the symbol $\Delta f=(2\pi)^{-1}(\omega_2-\omega_1)$. The *frequency difference* of the coupled system is always denoted by $\Delta\nu$. A typical plot visualizes $\Delta\nu$ against the detuning Δf. One observes a plateau $\Delta\nu=0$ for a synchronized region and a linear dependence for uncoupled oscillations. Below, we will use the symbol f for uncoupled frequencies, and ν for measurements of the coupled pipes; analogously we use Δf and $\Delta\nu$.

At the end of this section, we would like to hint to two facts: (i) the theory presented above holds only for small deviations from harmonic oscillations and for small coupling. For large parameters, in principal, the considerations still hold, but quantitatively deviations are expected. (ii) Even though it seems appealing to explain the frequency locking of organ pipes in such a simple way, it is not completely satisfactory because organ pipes are extended systems and such a simple description does not take into account the aeroacoustics that is eventually needed for a complete understanding.

TABLE I. Pipe geometry.

Pipe body	
Overall length (without foot)	530 mm
Clear length (adjustable)	370···450 mm
Wall thickness	6 mm
Cross section	
Shape clear	Quadratic
Clear width	37 mm
Clear depth	48 mm
Mouth	
Cutup height	7 mm
Cutup width	37 mm
Flue-exit height	0.4 mm
Foot	
Length	320 mm
Toe hole diameter	15 mm

III. MEASUREMENT AND RESULTS

A. Measurement setup and signal analysis

Measurements were carried out on a miniature organ especially made by Alexander Schuke GmbH for physical investigations. Just like a real organ, it consists of a blower connected by a mechanical regulating valve to the wind belt and finally a wind chest upon which the pipes are positioned. Since only stationary behavior was of interest, we joined the two pipes directly to the wind belt with flexible tubes. The wind pressure was set by the organ builder to a mean value of 160 Pa. Over 10 s we measured a standard deviation value of 6 Pa in the pipe foot. The wind may be considered as very stable.

The pipes used are stopped; a detailed description of the pipe-geometry is given in Table I. The resonator top of the pipes was movable in order to adjust resonator length and pitch. To minimize losses, a gasket made of felt was applied. The pipes as delivered from the manufacturer were tuned to

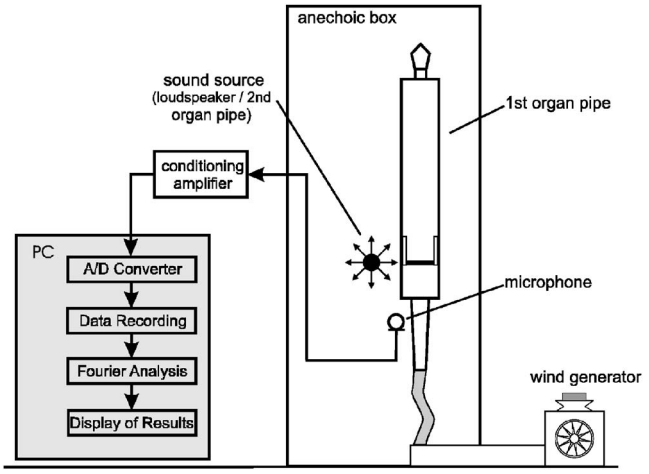


FIG. 3. Sketch of the experimental setup. The microphone is in plane with the pipe mouth, the view is at an angle from above.

e and f in German notation, that is, E3 and F3 in American notation. The applied driving pressure roughly yields a Reynolds number Re=435.

The acoustic signal was measured by a B&K 4191 condenser microphone positioned at the centerline between the sound sources. Further information on positioning of the acoustic sources and the microphone is given in Secs. III B and III C. Measurements took place inside the anechoic chamber of the Technical University of Berlin. The room's size is $W \cdot L \cdot D = 20\ \mathrm{m} \cdot 17\ \mathrm{m} \cdot 7\ \mathrm{m}$ giving a volume of 2000 m^3. The low cutoff frequency of this room is 63 Hz. Temperature, humidity, and pressure conditions were kept stable to keep their influence on the measurements negligible. A sketch of the experimental setup is given in Fig. 3.

The recording of the sensor signal as a function of time was achieved via hard-disk recording with an M-Audio 2496 Soundboard at a sampling rate of 44.1 kHz and a resolution of 16 bit and saved in PCM format. Data analysis was performed with MATLAB. All spectra shown in the results are Fourier transformed out of 10 s (441 000 samples) of the time signal, giving a frequency resolution of 0.1 Hz. During Fourier transformation, a symmetric Hann window was applied to time data. All spectra have been averaged. From these spectra the amplitude of the harmonics was read out by a routine searching for the maximum signal level within a frequency range of ±0.3 Hz.

B. Organ pipe driven externally by a loudspeaker

To investigate the influences of an external sound source onto the organ pipe, we positioned a loudspeaker directly beside a single organ pipe. The loudspeaker was a 2-Way Active System MS 16 from Behringer. The woofers cone diameter is 10 cm. The speaker was connected to a HP 33120A waveform generator, delivering a sinusoidal signal. The sound pressure level emitted by the loudspeaker was set equivalent to the volume of the fundamental organ tone to within 81.5±0.25 dB. At this level the distortion of the speaker was at THD=1.5%. The experiment started with the pipe at a fundamental frequency of 168.3 Hz and the loudspeaker at 165.8 Hz. During the measurement the organ pipe

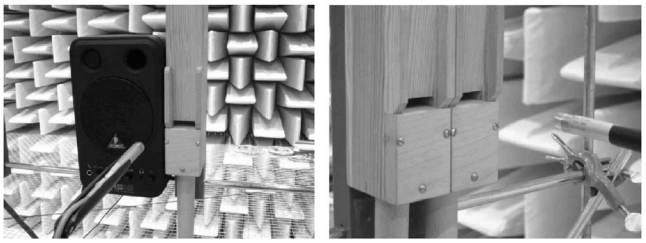


FIG. 4. Photograph of the measurement setup. Left: loudspeaker and pipe sounding together, right: two "coupled" organ pipes.

remained unchanged but the frequency of the loudspeaker was very slowly, every 40 s, increased by 0.1 Hz. This gives the possibility to average ten seconds of signal four times to suppress noise. The microphone was positioned at the centerline between the pipe and the loudspeaker at a distance of 15 cm. A picture of the setup is given in Fig. 4.

The synchronization can be observed clearly by the plateau in the plot of frequency difference, $\Delta\nu$, against detuning, Δf, as well the predicted square root function is observed at the transition to synchronization; cf. Fig. 5. The amplitude recorded with the microphone shows a positive and negative peak, at the beginning and end of the synchronization, respectively. Invoking the explanations of synchronization theory, cf. Eq. (7), this amounts to a change of the relative phase from 0 to π. If the loudspeaker and the pipe are considered monopole sources, their field at the centerline, where the signal is recorded, can be obtained as a superposition of two fields with equal, measured amplitude and phase difference ϕ_0, decaying as $1/r{:}A(r)=A_0/r + A_0/r \cos\phi_0$ (remember that we adjusted the sound pressure level of the loudspeaker and the pipe, so $A_{\mathrm{speaker}}=A_{\mathrm{pipe}}=A_0$). For $\phi_0=0$ interference is positive, and $\phi_0=\pi$ implies negative interference. This is exactly seen in Fig. 5(b), where we plot the peak amplitude of the loudspeaker's frequency (open circle), and the first harmonic of the organ pipe (filled circles), respectively, over the detuning. For zero detuning, $\phi_0=\pi/2$. This corresponds to an interval of π within the synchronization region; i.e., from positive to negative superposition. One notices that the phase $\phi_0=\pi/2$ right at $A=A_0$. This corresponds to an equal distribution of energy at this point. The increase of the amplitude at the loudspeaker's frequency at the left hand of the synchronization region comes from the above-mentioned effect that the pipe sounds for long periods with the frequency of the speaker to undergo a fast phase slip, when close to being synchronized. This results in a higher peak amplitude at the loudspeaker's frequency (cf. the discussion in Sec. III D).

C. Two pipes standing side by side

To measure the dependence of frequency locking on the detuning of the two pipes, both pipes were installed on a common bar directly side by side (see Fig. 4). One of the pipes has been kept at the fixed frequency of 170.1 Hz; the other pipe was detuned in variable steps. With two pipes connected to the wind chest, one cannot detect the fundamental frequency of one pipe by simply turning of the other through closing the pipes valve. This would raise the wind

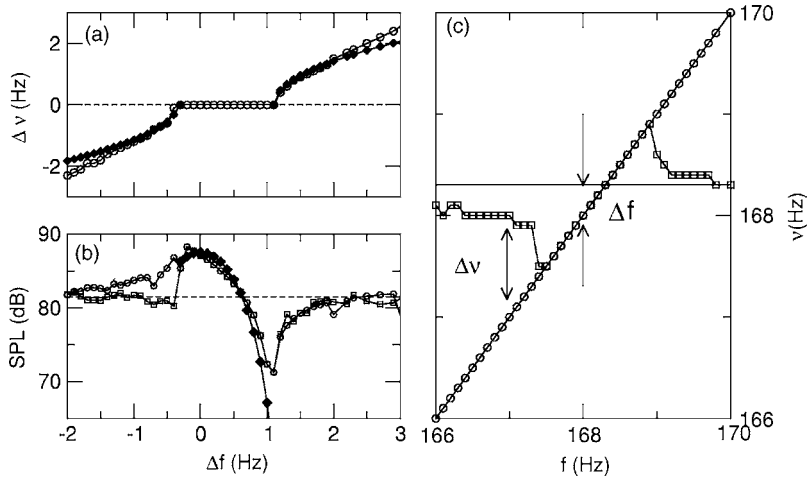


FIG. 5. Upper left (a): Frequency locking of organ pipe and loudspeaker. Top: observed frequency difference $\Delta\nu$ ascertained with pipe and loudspeaker sounding together versus detuning Δf (open circles). In the synchronization region, a very clean plateau is observed. A saddle-node bifurcation (filled diamonds) occurs at the edges, as predicted by Eq. (9). Bottom left (b): Measured amplitude signal in dB. The phase varies from 0 to π over the synchronization plateau. The region is asymmetric; $\Delta\nu \in (-0.3, 1.1)$ within the frequency resolution. The circles indicate the peak amplitude at the loudspeaker's frequency, and the pipe's frequency (squares). The superposition of the signals including the phase shift along the plateau is plotted by filled diamonds, the agreement is quite good. Right (c): absolute measured signals for pipe and loudspeaker sounding together (ν, open circles) and the signals for the speaker alone (squares) and pipe alone (straight line). The detuning is indicated by Δf, as the difference of the "uncoupled oscillators," pipe and speaker alone; the frequency difference, $\Delta\nu$, results from the subtraction of the measured signal for the pipe and speaker sounding simultaneously, coupled and the varied frequency, in our case of the loudspeaker.

pressure for the investigated pipe and consequently also the frequency itself. In order to acquire the value precisely the pipe at a fixed frequency was made soundless by putting a little absorber wedge into the mouth. Doing so, the air sheet oscillation and therefore sound generation is suppressed, but the air consumption remains the same, keeping the wind pressure stable.

In Ref. 16 the frequencies ν_1, ν_2 of the coupled pipes were plotted against the frequency detuning of the uncoupled pipes, Δf. As explained above, a clearer characterization is achieved when the frequency difference $\Delta\nu$ is plotted versus Δf.[19,32] We plot this quantity in Fig. 6. Clearly, the typical behavior of frequency locking can be observed. As synchronization theory predicts, the bifurcation close to the ends of the locking region is of the saddle node type[35] [filled diamonds in Fig. 6(a)].

Now, stressing the analogy with two coupled oscillators, a change of the relative phase from $\phi_0-\pi/2$ to $\phi_0+\pi/2$ is expected[19] ($\phi=2\pi f$). This has consequences for the amplitude of the emitted sound. The radiated sound wave in the far field can be roughly approximated by a spherical wave, and the field at the measurement point $\vec{r}$ is obtained by superposition of the two waves emitted at the pipe mouth's positions. Assuming the sources to be point-like and considering that the pipes are situated directly side by side, the amplitude along the centerline is $A=2A_0/r(1+\cos\phi_0)$, this has already been described in Ref. 15. The notation is as above; please note that here $A_{\text{pipe1}}=A_{\text{pipe2}}=A_0$, because both pipes have approximately the same amplitude; cf. Fig. 6.

Varying the frequency mismatch Δf, we observe a gradual increase of the amplitude at the centerpoint, where the microphone is situated. In Fig. 6(b), the result of the measurement is displayed: We see an excellent agreement of our estimate with the measurement. This behavior is observed as well in the spectral plot, Fig. 7, where one can observe the collapse of all sidebands in the measured spectrum to one single frequency as a very sharp transition.

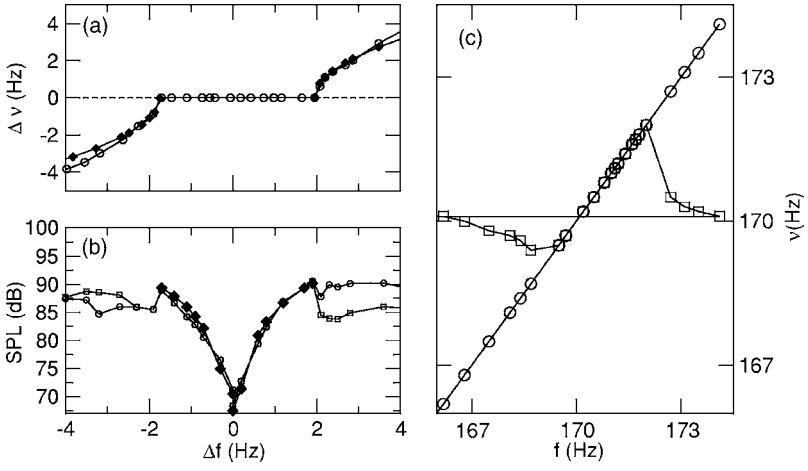


FIG. 6. Frequency locking of two organ pipes. The plot shows the observed frequency difference $\Delta\nu$ versus the detuning of the uncoupled pipes, Δf (circles). Upper left (a): In the synchronization region, a very clean plateau is observed. As in the case of external driving, the saddle-node bifurcation is observed at the edge to the synchronization region (filled diamonds). The agreement is very good. Lower left (b): Amplitudes at the first harmonics of the pipes (circles, variable-frequency pipe; squares, fixed-frequency pipe) against detuning; the sharp decrease of the amplitude at $\Delta f=0$ indicates an antiphase oscillation at the pipe mouth. The analytical curve obtained using synchronization theory is shown with filled diamonds, assuming $\phi_0=\pi$—the agreement is excellent. Right (c): Plot of the absolute frequencies (circles/squares as above); the straight line shows the 170.1 Hz of the fixed-frequency pipe.

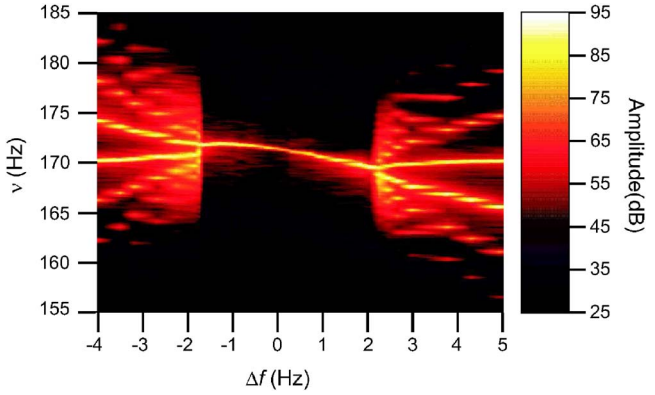


FIG. 7. (Color online) A sharp transition to synchronization is observed and the sidebands from beating in the measured spectrum collapse to a single frequency in the synchronization region. The amplitude is encoded according to the levels on the right.

For higher harmonics, the phase difference by ϕ_0 does not imply destructive interference because they are slaved and follow the first harmonic. In other words, the relative phase for the nth harmonic in the synchronization region lies in the interval $[n(\phi_0-\pi/2), n(\phi_0+\pi/2)]$. For the second harmonic, this implies in-phase oscillations at $\Delta f=0$. This is heard acoustically by a dominance of the octave around $\Delta f=0$. At the edge of the synchronization region, however, there is a phase difference of π and destructive interference is expected. This is confirmed well by the quantitative measurement of the amplitude dependence of the second harmonic, shown in Fig. 8. The agreement with the estimate from the superposition of two monopoles is still good, but deviations occur. Higher harmonics are in qualitative accordance with the above ideas, but quantitatively more and more differences are observed. This can be expected, because the approximation by a monopole source does not hold.

When two pipes are coupled diffusively, a so-called oscillation death can occur, leading to a complete silence of both pipes. The whole input power would then be converted to heat without radiation. We thus observe a strong impact of the synchronization of the pipes on the acoustic properties of the emitted sound. From our results, we can rule out an oscillation death for two coupled pipes with zero detuning, as suggested in Ref. 19. Rather, the two oscillators radiate two antiphase waves; this results in a vanishing acoustic signal.

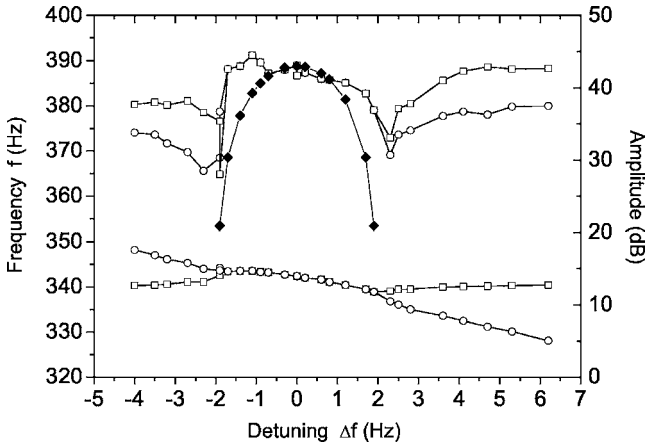


FIG. 8. Amplitude and frequency difference for the second harmonic. Here, the phase is doubled and an annihilation occurs right at the edge of the synchronization region, seen as a sharp decrease. The curve predicted by theory shows deviations for the second harmonic.

D. The coupling of the pipes

As in the case of the external driving by a loudspeaker, the two pipes interact by the pressure perturbation generated at the pipes' mouths. But now, there are two sources of pressure contributions: the acoustic and the aerodynamic part from the vortices itself. Which one is dominant depends on the ratio of their amplitudes. In the case of the organ pipes, the acoustic pressure is very large due to the selected amplification of the resonator frequencies. However, the wind field at one pipe can literally "blow away" an adjacent vortex. On the basis of our observations, we can only state that definitely, the acoustical field is sufficient to give raise of synchronization. The influence of the wind field must be subject of careful future work, including flow visualization on the scale of several vortex diameters. At the edge of the synchronization region, there must be a phase difference of $\phi_0 \pm \pi/2$, which corresponds to a shift of a quarter of a jet oscillation. The jets then undergo a mutual reordering into a complex spatiotemporal three-dimensional pattern. Again, the observation of the pattern was beyond the facilities of the current experimental setup.

All our observations are consistent with an interpretation in the frame of synchronization theory of two oscillators. A full description of the physics has to take into account the origin of the sound generation—the oscillations of resonator and the vortices generated by two jets. Most likely, there is a competition of the interaction between jets and resonators. Which mechanism wins depends then on the ratio of the respective amplitudes. A comment should be made on why the synchronization description by two oscillators holds so well. In the measured Reynolds number regime, there might be three-dimensional patterns at the pipe mouth, if the aspect ratio is large enough. To check this in our particular case requires detailed experiments. Neglecting three-dimensional structures, and assuming that the air sheet is homogeneously oscillating, a description by oscillator models like Stuart-Landau or a van der Pol equation could be sufficient to describe the two-dimensional oscillations and the transition to turbulence.[36] Since the main sound production is due to the interaction with the labium, an oscillator model suits well. From synchronization theory it follows that *any* self-sustained oscillators will generically follow the synchronization scenario when coupled, leading to the observed behavior.[19,37] The simulations in Ref. 16 are completely in agreement with this fact. We are convinced that a full understanding may be achieved by studying the details of the coupling mechanism.

To investigate the sensitivity to coupling strength, we measured the frequency spectrum about the first harmonic while varying the interpipe distance (measured between the outer walls of the pipes). The detuning was fixed to 0.7 Hz. The corresponding plot is depicted in Fig. 9. For large dis-

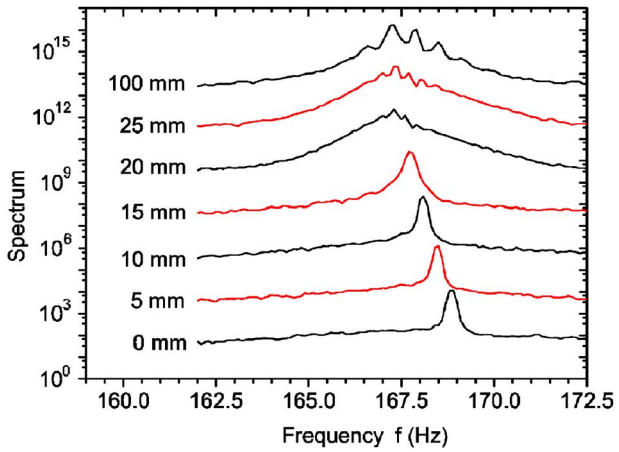


FIG. 9. (Color online) Frequency spectrum of the coupled pipes in dependence on the distance, given in the legend. The pipes decouple more and more when farther apart. For small distances, a sharp peak at the synchronization frequency is observed, for large distances, the peak broadens more and more, at very large distances the spectrum of the uncoupled pipes is recovered with the typical beat phenomenon. To distinguish different graphs, an offset has been added to each curve.

tances, the typical interference pattern of two noninteracting oscillators is observed. As the pipes come closer to each other, the individual, sharp peaks of half-width $\simeq 0.1$ Hz reduce in amplitude and at the same time the spectrum broadens to a "hump" with half-width $\simeq 1$ Hz. Peaks typical of the beating from linear superposition sit on the hump. Coming even closer, full synchronization is observed with one single, very sharp peak of again half-width $\simeq 0.1$ Hz. That means the hump is about ten times broader than either of the peaks for the uncoupled system or the synchronized one.

The basic observations can be explained by synchronization theory. In Eq. (9), the square root behavior of the frequency difference at the transition to synchronization (the bifurcation point) is derived. Lowering the coupling strength is equivalent to shift the bifurcation point toward zero. At a certain distance, the detuning leaves the synchronization region, and one should observe a (strongly nonlinear) beat. The peaks of the beat are observable as small side peaks on top of the broad peak. For a pure beat, only sharp peaks should be observed. Here, the duration of the large time interval—we do not want to call it period—varies slightly, such that the sidebands wiggle a bit; altogether this leads to a broad peak. To illustrate this behavior, we plot in Fig. 10 three different time signals for the distances 10 mm (synchronized), 15 mm (shortly before synchronization), and 25 mm. The fast oscillations look very similar for all of the distances, as exemplarily given by Fig. 10 (bottom, right).

A complete understanding of why such a broad hump appears could not be achieved. There are several possible explanations. According to classical synchronization theory the frequency fluctuation cannot be explained but by noise, e.g., in the air supply in the wind system. Another option are the small fluctuations in the vortices at the pipe mouth that can be sufficient to cause the very small variation in the beat frequency. Finally, one can consider the bifurcation to synchronization as a kind of phase transition, where typically large fluctuations occur, including chaotic behavior. By means of our measurement we cannot decide which of these scenarios is true. Finally, we want to mention recent results concerning whistling of two nearby Helmholtz resonators show a similar transition in the phase of the acoustical signals emitted by the resonators.[38] Interestingly, the phase relation is lost at distances of the order of the ones we find (25 mm, if the wall thickness of the pipes is included). Thus, this observation might be explained by synchronization as well.

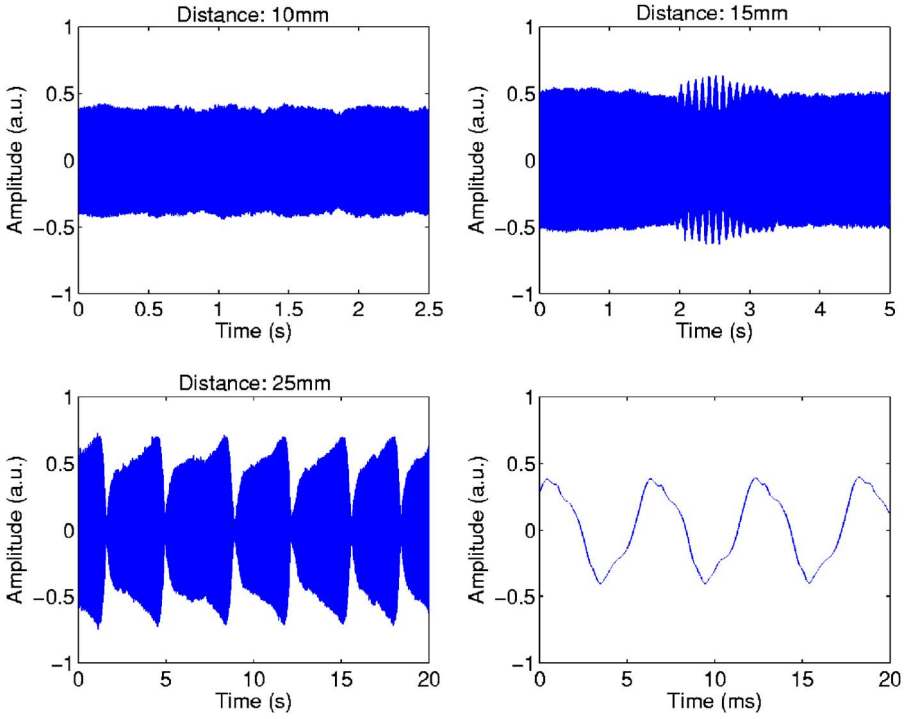


FIG. 10. (Color online) The time signal for three different distances. Upper left: 10 mm; lower left: 25 mm; upper right: 15 mm. The lower right graph shows a typical fast oscillation, as found for all three distances. The transition from synchronization to beat occurs between 15 and 25 mm. For 10 mm the amplitude is constant and the phase rotates with the fast frequency. At 25 mm the beat phenomenon is observed with a constant amplitude over relatively long times and a quick phase slip (see Sec. II B, 3). At 15 mm, the system shows some intermittent bursts at irregular times.

IV. CONCLUSION

We presented measurements on an external driving of an organ pipe by a loudspeaker and on the mutual influence of two organ pipes. The observed behavior is completely consistent with an explanation in the frame of synchronization theory. For the dependence of the synchronization on the "coupling strength," measurements determining the dependence of the frequencies on the interpipe distance have been carried out. They reveal a broadening of the peak at the synchronization frequency with increasing distance. This broadening can be explained by quick phase slips, as predicted by theory close to a saddle-node bifurcation. The smearing out of the side peaks of this beat can be either due to an inconstant wind supply or due to the fluctuations caused by the oscillating air sheet at the pipe mouth. To clear this question further measurements have to be carried out.

From an acoustic point of view, one can address the question of how to position two organ pipes close in frequency. This has been intuitively solved by organ builders by trial and error in the last centuries.[17] Our work might give quantitative hints on how large the interpipe distance needs to be to suppress mutual influence, and on details of the coupling mechanism. For example, avoiding an amplitude minimum for the first harmonic is highly desirable.

From an aerodynamical point of view, the above scenario requires more detailed investigations to understand the full dynamics of the coupled resonator/jet system, although a model consisting of two mutually coupled oscillators seems to be sufficient for all qualitative questions. In addition the more involved setup of more than two organ pipes is an interesting subject for further investigations. From our measurements, one can clearly say that the pipes do not show an oscillation death; rather, antiphase sound radiation yields the observed weakening of the amplitude.

ACKNOWLEDGMENTS

We acknowledge fruitful discussion with M. Rosenblum and A. Pikovsky about synchronization theory, with D. Lohse and J. F. Pinton, and thank J. Ong for careful reading of the manuscript. We especially thank the unknown referees for many constructive comments. M. Abel acknowledges support by the DFG (Deutsche Forschungsgemeinschaft). We thank the organ manufacturer Alexander Schuke Potsdam Orgelbau GmBH for kindly providing the organ pipes for our measurements.

[1]B. Fabre and A. Hirschberg, "Physical modeling of flue instruments: A review of lumped models," Acust. Acta Acust. 86, 599–610 (2000).
[2]N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments* (Springer-Verlag, New York, 1998).
[3]L. Cremer and H. Ising, "Die selbsterregten Schwingungen von Pfeifen," (The self-sustained oscillations of pipes) Acustica **19**, 143–153 (1967).
[4]J. W. Coltman, "Sounding mechanism of the flute and organ pipe," J. Acoust. Soc. Am. **44**, 983–992 (1968).
[5]J. W. Coltman, "Jet behavior in the flute," J. Acoust. Soc. Am. **92**, 74–83 (1992).
[6]N. H. Fletcher, "Autonomous vibration of simple pressure-controlled valves in gas flow," J. Acoust. Soc. Am. **93**, 2172–2180 (1993).
[7]A. Nolle, "Flue organ pipes: Adjustments affecting steady waveform," J. Acoust. Soc. Am. **73**, 1821–1832 (1983).
[8]J. W. Coltman, "Jet drive mechanism in edge tones and organ pipes," J. Acoust. Soc. Am. **60**, 725–733 (1976).
[9]M. P. Verge, B. Fabre, W. E. A. Mahu, A. Hirschberg, R. van Hassel, A. Wijnands, J. d. Vries, and C. Hogendoorn, "Jet formation and jet velocity fluctuations in a flue organ pipe," J. Acoust. Soc. Am. **95**, 1119–1132 (1994).
[10]M. P. Verge, B. Fabre, A. Hirschberg, and A. Wijnands, "Sound production in recorder like instruments. I. dimensionless amplitude of the internal acoustic field," J. Acoust. Soc. Am. **101**, 2914–2924 (1997).
[11]M. P. Verge, A. Hirschberg, and R. Caussé, "Sound production in recorder like e instruments. II. a simulational model," J. Acoust. Soc. Am. **101**, 2925–2939 (1997).
[12]B. Fabre, A. Hirschberg, and A. P. J. Wijnands, "Vortex shedding in steady oscillation of a flue organ pipe," Acust. Acta Acust. 82, 863–877 (1996).
[13]C. Segoufin, B. Fabre, and L. de Lacombe, "Experimental investigation of the flue channel geometry influence on edge-tone oscillations," Acta. Acust. Acust. **90**, 966–975 (2004).
[14]J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. 2.
[15]H. Bouasse, *Instruments a Vent* (Lib. Delagrave, Paris, 1929).
[16]D. Stanzial, D. Bonsi, and D. Gonzales, "Nonlinear modeling of the mitnahme-effect in coupled organ pipes," in *International Symposium on Musical Acoustics (ISMA) 2001*, Perugia, Italy, pp. 333–337, 2001,
[17]J. Angster, J. Angster, and A. Miklós, "Coupling between simultaneously sounded organ pipes," *AES preprint 94th Convention Berlin 1993*, March, 1993.
[18]D. Zscherpel (private communication, 2004).
[19]A. Pikovsky, M. Rosenblum, and J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, MA, 2001).
[20]J. W. S. Rayleigh, "On the pitch of organ-pipes," Philos. Mag. **XIII**, 340–347 (1882).
[21]S. Pitsch, J. Angster, M. Strunz, and A. Miklós, "Spectral properties of the edge tone of a flue organ pipe," *Proceedings International Symposium and Musical Acoustics, 1997*, Edinburgh **19**(5), 339–344 (1997).
[22]N. H. Fletcher, "Air flow sound generation in musical wind instruments," Annu. Rev. Fluid Mech. **11**, 123–146 (1979).
[23]N. H. Fletcher, "The nonlinear physics of musical instruments," Rep. Prog. Phys. **62**, 723–764 (1999).
[24]N. H. Fletcher, "Mode locking in nonlinearly excited inharmonic musical oscillators," J. Acoust. Soc. Am. **64**, 1566–1569 (1978).
[25]N. H. Fletcher, "Sound production by organ flue pipes," J. Acoust. Soc. Am. **60**, 926–936 (1976).
[26]T. A. Johansson and M. Kleiner, "Theory and experiments on the coupling of two Helmholtz resonators," J. Acoust. Soc. Am. **110**, 1315–1328 (2001).
[27]M. S. Howe, "Contribution to the theory of aerodynamic sound, with application to excess jet noise and theory of the flute," J. Fluid Mech. **71**, 625–673 (1975).
[28]M. S. Howe, *Theory of Vortex Sound*, (Cambridge University Press, Cambridge, UK, 2003).
[29]M. S. Howe, *Acoustics of Fluid-Structure Interactions* (Cambridge University Press,Cambridge, UK, 1998).
[30]S. K. Lele, in *Proceedings of the 35th Aerospace Sciences Meeting and Exhibit*, AIAA Paper 97-0018, Reno, Nevada, 1997.
[31]C. K. W. Tam, "Advances in numerical boundary conditions for computational aeroacoustics," J. Comput. Acoust. **6**, 377–402 (1998).
[32]N. N. Bogoliubov and V. A. Mitropolsky, "Asymptotic methods in the theory of non-linear oscillations," *International Monographs on Advanced Mathematics and Physics* (Gordon and Breach Science, New York, 1961) [translation from Russian original (1955)].
[33]P. S. Landa, *Nonlinear Oscillations and Waves in Dynamical Systems* (Kluwer Academic, Dordrecht, 1996).
[34]R. Adler, "A study of locking phenomena in oscillators," Proc. IRE **34**, 351–357 (1946).
[35]E. Ott, T. Sauer, and J. Yorke, *Coping with Chaos*, Series in Nonlinear Science (Wiley, New York, 1994).
[36]M. Provansal, C. Mathis, and L. Boyer, "Bénard-v on Karman instability: transient and forced regimes," J. Fluid Mech. **182**, 1 (1987).
[37]D. G. Aronson, G. B. Ermentrout, and N. Koppel, "Amplitude response of coupled oscillators," Physica D **41**, 403–449 (1990).
[38]M. M. G. Derks and A. Hirschberg, "Self-sustained oscillation of the flow along Helmholtz resonators in a tandem configuration," in *8th International Conference on Flow Induced Vibration, FIV*, July 2004, pp. 435–440.

Noncontact modal analysis of a pipe organ reed using airborne ultrasound stimulated vibrometry

Thomas M. Huber
Physics Department, Gustavus Adolphus College, Saint Peter, Minnesota 56082

Mostafa Fatemi, Randy Kinnick, and James Greenleaf
Ultrasound Research Laboratory, Mayo Clinic and Foundation, Rochester, Minnesota 56082

(Received 21 June 2005; revised 10 January 2006; accepted 10 January 2006)

The goal of this study was to excite and measure, in a noncontact manner, the vibrational modes of the reed from a reed organ pipe. To perform ultrasound stimulated excitation, the audio-range difference frequency between a pair of ultrasound beams produced a radiation force that induced vibrations. The resulting vibrational deflection shapes were measured with a scanning laser vibrometer. The resonances of any relatively small object can be studied in air using this technique. For a 36 mm $\times$ 6 mm brass reed, displacements and velocities in excess of 5 μm and 4 mm/s could be imparted at the fundamental frequency of 145 Hz. Using the same ultrasound transducer, excitation across the entire range of audio frequencies was obtained. Since the beam was focused on the reed, ultrasound stimulated excitation eliminated background effects observed during mechanical shaker excitation, such as vibrations of clamps and supports. The results obtained using single, dual and confocal ultrasound transducers in AM and two-beam modes, along with results obtained using a mechanical shaker and audio excitation using a speaker are discussed. [DOI: 10.1121/1.2171516]

PACS number(s): 43.75.Np, 43.35.Zc, 43.40.Cw [NHF] Pages: 2476–2482

I. INTRODUCTION

In a variety of academic and industrial applications, modal analysis is an important technique for studying systems. The goal is to excite mechanical vibrations of the system with a driver, and measure the response using some measuring transducer. In the past, the primary instrument used for measuring vibration was the accelerometer. One drawback related to their usage is the fact that they place a mass burden on the object; if the object is small, the mass of the accelerometer can distort the vibrational modes. In recent years, the laser Doppler vibrometer[1] has become the instrument of choice in many applications requiring measuring vibration, since there is no contact between the vibrometer and the device under study. The reed in reed organ pipes is one of numerous systems that has been studied using laser vibrometry.[2]

While the laser vibrometer has allowed noncontact measurements of vibration, there are fewer noncontact methods that can be used for excitation. The primary excitation transducers used in modal analysis are the mechanical shaker and the impact hammer. Especially when a mechanical shaker is used for a relatively small object such as a reed, one area of concern is the distortion of resonances due to mass loading by the shaker. After an impact hammer leaves the surface of an object, there is no mass loading. With appropriate care in setting up a miniature hammer as an excitation source, it would be possible to measure the frequency response and deflection shape of an object such as the organ reed. Other techniques, such as electromagnetic driving or using an acoustical speaker, are sometimes used for noncontact excitation. However, these are limited in terms of the frequency response, or the composition of the object of interest.

Described below is a technique using the radiation force resulting from the interference of ultrasound beams[3] to induce audio-range vibrations of an organ reed. These vibrations were then detected using a laser vibrometer. Both the excitation method, using ultrasound radiation force, as well as the detection method are noncontact. There have been several studies which used ultrasound radiation force for modal analysis in water,[4] however the current study is the first which demonstrates the feasibility of using this technique in air.

A. Previous studies of vibrational modes of pipe organ reeds

The current study is a practical application of the ultrasound radiation force in air to validate results obtained in a recent study of the resonances and deflection shapes of an organ reed.[5] In an organ reed pipe, the air flows past a thin brass reed, through the shallot (a tube with a slot to which the reed is clamped), and into a resonator. For the study of Ref. 5, the reed was excited both by a conventional air blower, as well as with a mechanical shaker, and the resonance frequencies and deflection shapes were measured using a scanning laser vibrometer.

When an organ reed pipe is excited with an air blower, similar to the method that would be used in a conventional pipe organ, the acoustic spectra measured using a microphone as well as the velocity spectra of the reed measured using a vibrometer shows a series of integer-multiple harmonics of the fundamental.[2,5,6] Measurements with a scanning vibrometer revealed that the deflection shapes measured at these integer-multiple harmonics of the fundamental

 0001-4966/2006/119(4)/2476/7/$22.50

showed complicated deflection shapes that included torsional and higher-order transverse deflection shapes.[5]

To understand the deflection shapes observed for the air-driven reed, the simplest model that one might consider for an organ reed is a cantilever fixed at one end and free at the other end. For this type of cantilever, theory predicts a set of nonharmonic modes that include both transverse and torsional mode shapes. To verify this, the resonance frequencies and deflection shapes of the reed excited with a mechanical shaker were measured.[5]

One concern about using a shaker to measure the vibrational modes of the reed was that the driving is indirect. The mechanical shaker, attached to the back of the shallot, caused a small vibration (<1 μm) of the entire pipe; when the driver was at a resonant frequency of the reed, it led to an increased signal from the laser vibrometer. However, since the driving was indirect, vibrations of any portion of the pipe, as well as supports or clamps were also detected by the vibrometer. As is the case with many relatively small objects, it was not possible to attach a shaker directly to the reed. Since the mass of the shaker is significantly larger than the reed, it would distort the reed's vibrational modes. The ultrasound excitation method described in the current paper is a direct method to drive the reed in a noncontact manner; the results obtained are compared with those obtained with a more conventional mechanical shaker and excitation using a speaker.

B. Outline of the paper

The following section discusses the physical mechanism that leads to audio-range excitation of structures using ultrasound stimulation. In Sec. III, the apparatus and techniques used in this study are described. These are followed by results of the experiment in Sec. IV, and discussion of these results in Sec. V.

II. THEORY: RADIATION FORCE DUE TO ULTRASOUND STIMULATION

Previous papers have described in detail the mechanism for ultrasound stimulated audio-range excitation, particularly in water.[3] The object of study, in the current experiment an organ reed, was excited by two different frequencies f_1 and $f_2=f_1+\Delta f$ where f_1 and f_2 are ultrasound frequencies, and Δf is the audio-range frequency of interest. When the two frequencies impinge on the reed, interference between the two frequencies produces a radiation force[3] that results in a vibration of the reed at the audio-range frequency Δf.

The radiation force[7] is caused by changes in the energy density of an acoustic field. The total ultrasound pressure field at a point from a pair of CW ultrasound sources of frequency f_1 and f_2, with pressure amplitudes $P_1(\bar{r})$ and $P_2(\bar{r})$ and phases $\varphi_1(\bar{r})$ and $\varphi_2(\bar{r})$ may be written as

$$p(\bar{r},t) = P_1(\bar{r})\cos[2\pi f_1 t + \varphi_1(\bar{r})] + P_2(\bar{r})\cos[2\pi f_2 t + \varphi_2(\bar{r})]. \quad (1)$$

This causes an instantaneous energy density given by $e(\bar{r},t)=p(\bar{r},t)^2/\rho c^2$; this energy density will have a time-independent component, a component at the difference frequency Δf, and high-frequency components at multiples of f_1 and f_2. The radiation force of interest for the current technique is the energy density component at the difference frequency, which can be written as

$$e_{\Delta f}(\bar{r},t) = \frac{P_1(\bar{r})P_2(\bar{r})}{\rho c^2}\cos[(2\pi\Delta f)t + \Delta\varphi(\bar{r})]. \quad (2)$$

Assuming that $P_1(\bar{r})$ and $P_2(\bar{r})$ are plane waves, this will impart a force in the beam direction on an object of area dS with drag coefficient $d_r(\bar{r})$ given by[3]

$$F_{\Delta f}(\bar{r},t) = e_{\Delta f}(\bar{r},t)d_r(\bar{r})dS = \frac{P_1(\bar{r})P_2(\bar{r})}{\rho c^2}\cos[(2\pi\Delta f)t + \Delta\varphi(\bar{r})]d_r(\bar{r})dS. \quad (3)$$

The total radiation force as a function of time is the integral of Eq. (3) over the entire surface of the object; this radiation force can cause a vibration of the object at a frequency Δf. Object vibration due to this radiation force is a function of the size, shape and mechanical impedance of the object. If the radiation force at frequency Δf corresponds to one of the resonant frequencies of the object, it will induce a larger amplitude vibration. The difference frequency Δf is swept through a range of frequencies, and the response is measured at each frequency.

III. EXPERIMENTAL SETUP

A. Scanning laser doppler vibrometer

To measure the vibration of the organ reeds, a Polytec PSV-300 scanning laser vibrometer was used.[8] A vibrometer uses the Doppler shift of reflected laser light to determine the speed of the vibrating reed. The tremendous advantage in using a laser vibrometer is that, unlike a conventional accelerometer, no physical contact is needed to measure the vibration of the object. This makes a laser vibrometer ideal for measuring the motion of a small object such as an organ reed.

With a scanning laser vibrometer, the laser can be deflected across the surface to measure the amplitude and phase at many points on the surface of the reed. Instead of measuring the motion of a single point, a scanning vibrometer and its software can reconstruct the vibrational deflection shapes of the surface. To determine these deflection shapes, the software calculates the phase shift between the electrical signal creating the driving force and the vibrational response at many positions on the reed. A primarily transverse mode will have a constant phase across the entire width of the reed, whereas a torsional mode will have a 180-degree phase shift across width of the reed.

B. Reed organ pipes

The apparatus used in this experiment is shown schematically in Fig. 1. The reed pipe that was the primary focus of the current study, as well as a previous study,[5] was a $G_5^{\#}$ trumpet pipe with a conical resonator 14 cm in length with a diameter that went from 1.1 cm to 4.6 cm. In order to enable imaging of a larger fraction of the reed, the 1.45 mm diameter tuning wire was bent in such a fashion that it did not

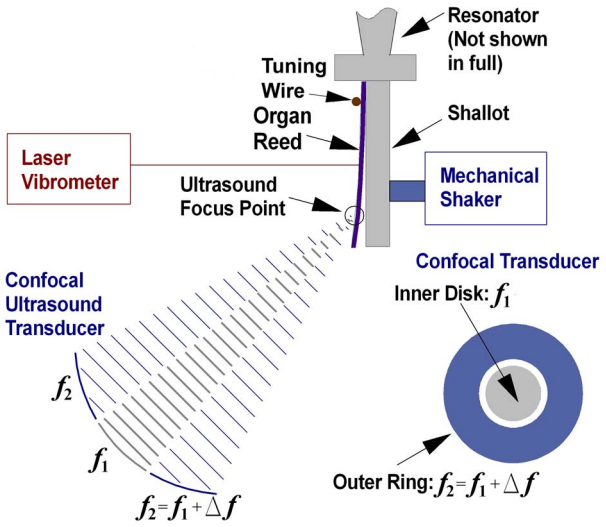


FIG. 1. (Color online) Apparatus used to compare ultrasound stimulated excitation with other excitation methods. The inner disk and outer annulus of the confocal ultrasound transducer, shown in this illustration, could be driven at two different frequencies. The vibrational deflection shapes of the reed were measured using a scanning laser vibrometer. For comparison, the system could be excited using a mechanical shaker placed in contact with the back of the shallot.

obscure the bottom section of the reed. The tuning wire was adjusted by blowing the pipe at its nominal blowing pressure of 3 in. of water and tuning it to a $G_5^{\#}$ with a Korg chromatic tuner. The reed in this pipe is 0.091±0.02 mm thickness brass with a width of 4.57±0.02 mm, and a length of 12.2±0.1 mm from the wedge to the semicircular free end of the reed. After tuning the pipe, the center of the tuning wire was 9.2±0.2 mm from the free end of the reed. Two other pipes were also used, including one with a significantly larger reed that was 6 mm×36 mm.

For the $G_5^{\#}$ trumpet pipe used in both the current study and Ref. 5, the three lowest modes that are expected for a clamped-free cantilever are a simple transverse cantilever mode, a torsional mode, and a second-transverse cantilever mode that has a nodal line roughly 1/3 of the way from the free end of the reed.[9] To estimate the modal frequencies for this reed, the tables of Ref. 9 were used for a rectangular cantilever with an effective length of 8.7±0.3 mm, which accounts for the semicircular end of the reed; this results in predictions of 670±50 Hz for the fundamental mode, 2.8±0.2 kHz for the torsional mode, and 4.2±0.3 kHz for the second-transverse cantilever mode. The modal frequencies for the reed, including its semicircular end, were also calculated using the finite element modeling package[10] in ME'Scope; the first three modal frequencies predicted using this program were 730±20 Hz, 3.2±0.2 kHz, and 4.5±0.2 kHz, respectively. These predictions are in good agreement with the measured frequencies of 726±2 Hz, 2.95±0.01 kHz, and 4.54±0.02 kHz for the fundamental, first torsional, and second transverse modes when the pipe was driven with a mechanical shaker in the absence of air in a vacuum chamber.[5] The goal of the current study was to compare these results to the measurements obtained using noncontact ultrasound excitation.

C. Excitation using confocal transducer

There were two different types of ultrasound transducers used in the current study. As shown in Fig. 1, the primary transducer used was a custom-made confocal ultrasound transducer for operation in air (MicroAcoustics Instruments, Broadband Air-Coupled Transducer sBAT-5). This transducer produces a focused ultrasound spot with a roughly Gaussian beam profile 1 mm in diameter with a focal length of 7 cm. This transducer has an inner disk that can be driven at one frequency, and an outer annulus that can be driven at another frequency; both elements have broadband performance, with a central maximum located near 600 kHz and a bandwidth of over 200 kHz. As shown in Fig. 1, in order to avoid blocking the beam from the scanning vibrometer, this transducer was mounted below the reed and pointed upwards at a roughly 45° angle. When used in dual-beam mode, the inner disk was driven at one frequency f_1, and the outer annulus was driven at a different frequency $f_2=f_1+\Delta f$; this requires a separate function generator and amplifier for both of the two transducer elements.

Another method of producing the difference frequency Δf was to emit a single waveform that was a superposition of the two frequencies, i.e., using a double-sideband suppressed-carrier amplitude modulated (AM) waveform.[11,12] In the case of AM excitation, both the inner disk and outer annulus of the confocal transducer were driven with the same AM waveform. This method has the distinct advantage of requiring only a single ultrasound transducer and amplifier.

D. Diverging (ex-focal) excitation

In addition to using the confocal ultrasound transducer, another excitation method involved using a pair of diverging ultrasound transducers (Prowave 328ST180) which were directed towards the organ reed. These transducers had a central frequency of 32.8 kHz and a bandwidth of about 1 kHz; instead of emitting a focused ultrasound beam, these transducers are diverging with a full beam angle (at 6 dB below the maximum) of 45°. As shown in Fig. 2, a pair of these transducers were placed 5 cm from the reed surface and oriented at a 45° angle of incidence relative to the surface of the reed. This leads to an interaction region about 3 cm wide where there was significant overlap between the two ultrasound beams; it is only within this region that excitation at the difference frequency Δf can occur. This overlap region was centered on the reed. Because they were mounted to the left and right of the reed surface, the transducers did not block the beam from the scanning vibrometer.

As shown in Fig. 2, there is also a region behind the reed where the ultrasound beams could have overlapped, which might induce a radiation force on the supports that held the organ pipe. In that region, the intensity of each beam separately was 3 dB or more below the maximum, so the ultrasound radiation stress in that region would be a small fraction of the radiation stress arriving at the reed. Also, the supports were positioned such that the mechanical shaker placed directly behind the reed shielded this region of over-

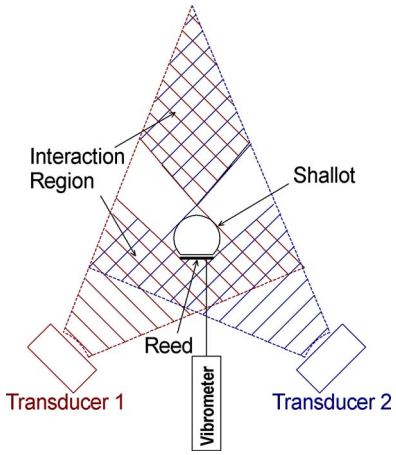


FIG. 2. (Color online) Excitation of reed using a pair of ex-focal (diverging) ultrasound transducers as viewed from above; the vibrometer is not drawn to scale. Excitation can only occur in the cross-hatched interaction region where there is an overlap between the output of the two ultrasound transducers. In the region behind the reed, the ultrasound radiation force is much smaller than the region in the vicinity of the reed, since the intensity from each transducer separately was at least 3 dB smaller. Not shown is the mechanical shaker, that was attached to the shallot opposite from the vibrometer. This shaker blocked the interaction region behind the reed, and thus prevented ultrasound from reaching any supports behind the reed pipe.

lap. Thus, even though the ultrasound beams were diverging, essentially only the organ reed experienced the ultrasound radiation force directly.

E. Waveform generation

To generate waveforms, a pair of Hewlett Packard 33120A function generators were used. For the pair of transducers or when the two elements of the confocal transducer were excited separately, one of the function generators generated a sine wave with a fixed frequency f_1 and the other was set to sweep a chirp sinusoidal waveform of frequency $f_2=f_1+\Delta f$. To generate a reference signal, that was needed to determine the vibrometer phase, the two signals were mixed and filtered using a bandpass filter at $f_2-f_1=\Delta f$. The signals from the two function generators were directed to a pair of power amplifiers; for the 32.8 kHz transducers a Crown 150A stereo amplifier was used, and when the confocal transducer was used, a pair of ENI 240L RF amplifiers served to amplify the signals.

To generate an amplitude modulated waveform, one of the function generators generated an audio-range waveform; this was connected to the second function generator to produce a double-sideband, suppressed carrier amplitude modulated signal. Because the generator produced two sidebands, the difference frequency produced is twice the frequency of the audio-range input frequency.

F. Shaker and audio excitation

The organ reed could also be excited using two conventional methods: a mechanical shaker, and an audio speaker. First, similar to previous studies, a Brüel-Kjær 4810 mechanical shaker was attached to the back of the shallot with a small piece of wax at the point of contact. This shaker caused a very small vibration of the entire pipe. These vibrations

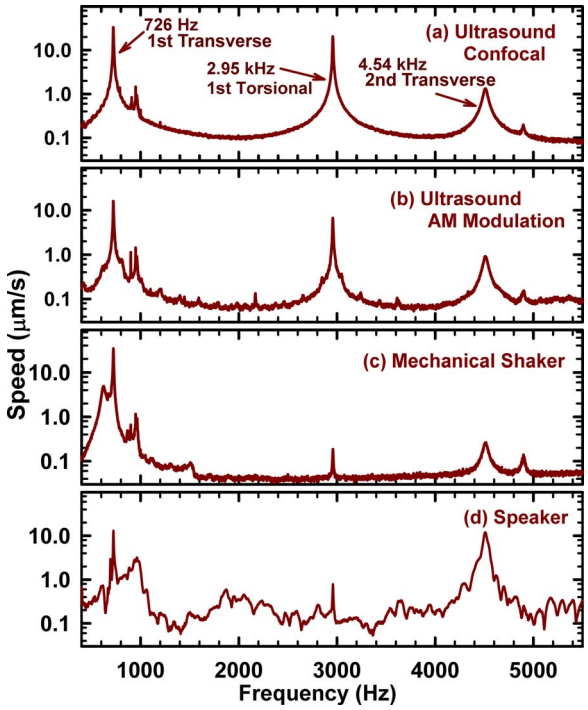


FIG. 3. (Color online) Vibrometer measurements of frequency response of the reed showing the first three resonances. (a) Ultrasound stimulated excitation using confocal transducer; (b) Ultrasound stimulated excitation using the same confocal transducer in AM mode; (c) excitation using a mechanical shaker; (d) audio excitation by nearby speaker. The peaks observed near 900 Hz and 4.8 kHz are from a previously observed noise source in the system.

were transmitted to the reed by its contact with the shallot, causing a larger amplitude output from the vibrometer when the driving frequency matched one of the reed's resonance frequencies. The shaker and reed pipe were attached by a series of rods and clamps to a vibration isolation table. The clamp system has its own resonant frequencies; since the shaker vibrated the entire system, resonances in this support system would also be observable in the vibrometer spectra.

The other method used to excite the reed was acoustic excitation using an audio speaker. To perform this excitation, an Altec Lansing speaker was placed roughly 10 cm from the pipe.

IV. RESULTS

A. Modal analysis using ultrasound stimulated vibrometry

Figure 3 shows the vibrometer spectra, averaged over the 50 points distributed over the surface of the reed, obtained for a $G_5^{\#}$ trumpet pipe; this was the same pipe characterized in a previous study.[5] The ultrasound transducer used for excitation was a confocal transducer. The position of the transducer was adjusted using a translation stage to obtain the maximum signal from the vibrometer; the maximum signal was obtained when the transducer was about 3 cm from the reed, which is consistent with its focal length. Figure 3(a) shows the spectrum obtained when the inner disk was driven with a frequency $f_1=550$ kHz and the outer annulus was driven with a frequency f_2 which was varied with a sinu-

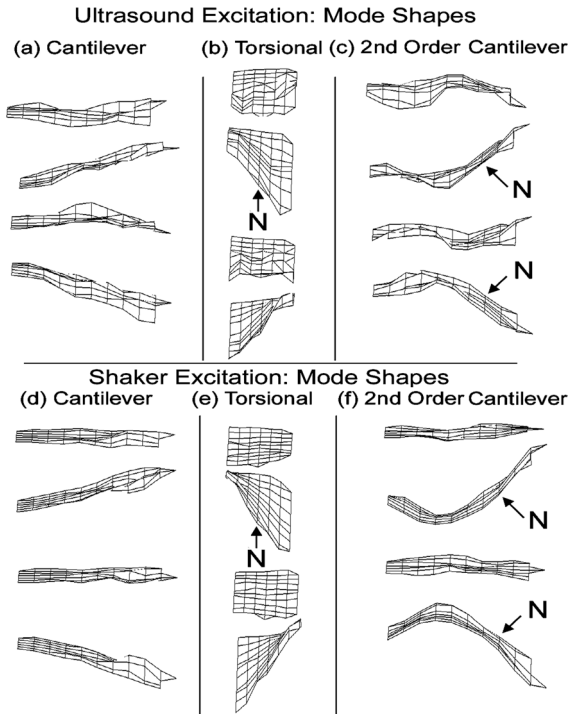


FIG. 4. Scanning vibrometer measurements of the displacement of the reed from its equilibrium position. Successive images downwards show displacement after 90° phase shifts during the cycle; the N in the figures illustrates the location of the nodal line for the torsional and second-transverse modes. (a) and (d) are the first transverse mode of the cantilever at 726 Hz; (b) and (e) are the first torsional mode at 2.95 kHz, and (c) and (f) are the second transverse mode of the cantilever at 4.54 kHz. The deflection shapes of (a)–(c), using ultrasound stimulated excitation, were nearly identical to (d)–(f) that were obtained using a mechanical shaker.

soidal chirp waveform from 550.5 kHz to 556 kHz over a 1.25 s period. This gave a difference frequency that ranged from 500 Hz to 6 kHz that excited the reed. The vibrometer spectra clearly shows peaks at 726 Hz, 2.95 kHz, and 4.54 kHz; these are in agreement with the predicted frequencies for a clamped-free cantilever.

Figures 4(a)–4(c) show the deflection shapes measured by the scanning vibrometer for these three resonances; these shapes are consistent with the mode shapes expected for the first three modes of a clamped-free cantilever. In particular, the peak at 2.95 kHz has a very clear torsional motion, and the peak at 4.54 kHz has a nodal line about 1/3 of the distance from the free end of the reed, which is expected for the second transverse mode.

There are two other regions of interest that can be observed in the spectra of Fig. 3(a); the region near 900 Hz, and the region near 4.8 kHz also show evidence of a resonance. These same peaks have been observed in spectra of other experiments using the same apparatus including vibrometer, electronics, and isolation table. Thus, they appear to be a noise source in either the electrical system, or an unrelated resonance in some portion of the mechanical system. Looking at the deflection shapes at 900 Hz and 4.8 kHz, the reed does not show a collective vibrational shape similar to those in Fig. 4. Instead, the phase at these frequencies for each measurement point on the reed appears to be uncorrelated to the phase of the driving force, which is consistent with a noise source.

Figure 3(b) shows the vibrometer spectrum when the reed was excited with the same ultrasound transducer using amplitude modulation. Both the inner disk and outer annulus were driven using an amplitude-modulated waveform with a carrier frequency of 550 kHz and a modulation frequency that was swept from 250 Hz to 3 kHz. Thus the difference frequency ranged from 500 Hz to 6 kHz. The spectrum of Fig. 3(b), with excitation by a single amplitude modulated ultrasound signal, is nearly identical to the spectra observed in Fig. 3(a) with the two beams from the confocal transducer. This is consistent with previous comparisons of confocal and amplitude modulated ultrasound excitation in water.[12] In contrast to the silent operation when the two elements of the transducer were driven with different frequencies, there was some small emission of audible sound because of recombination of the two waveforms in the transducer itself.

To demonstrate the ability to excite a larger reed, the 1 mm diameter ultrasound beam driven with an AM signal was focused on a much larger 36 mm $\times$ 6 mm brass reed. Even for this larger reed, the expected transverse and torsional deflection shapes of the entire reed were excited. Displacements and velocities in excess of 5 μm and 4 mm/s could be imparted at the fundamental frequency of 145 Hz.

B. Comparison with mechanical shaker and acoustical speaker

Figure 3(c) shows the vibrometer spectrum when a mechanical shaker was used as an excitation source. This shaker, in contact with the back of the shallot of the pipe, was driven with a chirp excitation from 500 Hz to 6 kHz. The amplitude of the driving force was attenuated to produce a vibrometer spectrum with about the same amplitude as for the ultrasound excited case; the amplitudes used in this experiment were roughly 1/100 of the maximum amplitude that could be obtained using this mechanical shaker. The frequencies of the three primary peaks in the shaker spectrum are similar to the peaks observed in the ultrasound excited spectra. As expected for an uncorrelated noise source, the background peaks at about 900 Hz and 4.8 kHz were observed with roughly the same amplitude in this spectra as in the other spectra. Figures 4(d)–4(f) show the deflection shapes measured for the three primary resonances. Figures 3(a)–3(c) and Fig. 4 demonstrate that noncontact ultrasound stimulation produces the same resonance frequencies and deflection shapes as are produced using a mechanical shaker.

One significant difference between the shaker spectrum of Fig. 3(c), and the ultrasound spectra of Figs. 3(a) and 3(b) was the relative amplitude of the peaks. In the shaker spectrum, the torsional peak was significantly smaller in amplitude than the two transverse peaks. One possible explanation was that the shaker was more efficient at exciting transverse modes than torsional modes. Since the shaker was moving the entire shallot, the driving force on the two sides of the reed were in phase instead of a 180° phase difference needed for a torsional mode. When the shaker was used, there was little lateral asymmetry in the driving force that would be needed to excite a torsional mode. In contrast, the ultrasound source that was used comes to a relatively small focus. Since the ultrasound spot was not focused along the axis of the

reed, this causes a left-right asymmetry in the driving force which may be more efficient at exciting torsional vibrations.

Figure 3(d) shows results of another noncontact method for exciting the reed, namely using a nearby audio speaker. Because of the poor low-frequency response of the speaker used, there was very little excitation of the fundamental mode. It would be difficult to obtain reasonably uniform frequency response across the entire audio spectrum using a single speaker element. The output from an audio speaker, especially for low frequency, cannot be focused. This means that only a small fraction of the acoustical power needed to excite a small object, such as a reed, will be incident on the object. The remainder of the acoustical output is emitted in other directions.

C. Localized nature of the ultrasound stimulated excitation using divergent beams

In addition to being noncontact, another unique characteristic of ultrasound stimulated excitation is that it can be localized. By using a focused ultrasound transducer, the driving force is applied to a small region; another way to achieve a localized excitation is to use a pair of diverging ultrasound transducers oriented such that the interaction region of the beams from the two transducers is relatively small. Oscillations of support structures were highly suppressed by using ultrasound stimulation instead of a mechanical shaker.

As shown in Fig. 2, the ultrasound transducers used were a pair of diverging transducers that were driven at two different frequencies; one was fixed at 32.4 kHz, and the other was swept from 32.5 kHz to 33.2 kHz. The region of overlap from the two transducers was sufficiency large to ensonify the reed, but did not ensonify the clamps or supports that were holding the pipe. Figure 5(a) shows the vibrational spectra of a reed that was excited using ultrasound stimulation; the reed of the pipe used for this test had a resonance frequency of 580 Hz. The major peak observed at 580 Hz was the fundamental transverse resonance of this reed. There were also some small peaks observed between 200 and 450 Hz.

In Fig. 5(b), the same system was driven with a mechanical shaker using a swept sine wave from 100 Hz to 800 Hz. In addition to the main peak, the peaks between 200 and 450 Hz were much larger than the ultrasound stimulated case. These peaks appear to originate in resonances of the supports and clamps used to hold the reed pipe in place; there was very little damping material that was used in setting up the apparatus for this experiment. By making adjustments such as moving or tightening the clamps and supports, these resonances changed. For example, the only difference between Figs. 5(b) and 5(c) was that a 100 g mass was placed on one of the clamps. This caused a shift of roughly 25 Hz in the 240 Hz resonance. These results indicate that these additional peaks between 200 and 450 Hz originate from the fixtures and support structures holding the reed pipe.

Since the overlap of ultrasound does not include the supports holding the pipe, very little energy was transferred to the vibrational modes of the clamps and supports. This is in contrast to the shaker, which vibrated the entire system, and thus drove a significant amount of energy into these modes. Since the energy is proportional to the square of velocity, squaring the spectrum for the ultrasound stimulated excitation shown in Fig. 5(a), the area under the main peak region from 520 to 640 Hz is 270 times the area in the region due to the support clamp resonance from 200 to 300 Hz. A similar analysis of the square of the spectrum in Fig. 5(b), produced using the mechanical shaker, shows that the primary peak contains only six times the area of the region from 200 to 300 Hz. This demonstrates that ultrasound radiation force is much more effective than a mechanical shaker for exciting the reed without exciting the corresponding background vibration of the supports.

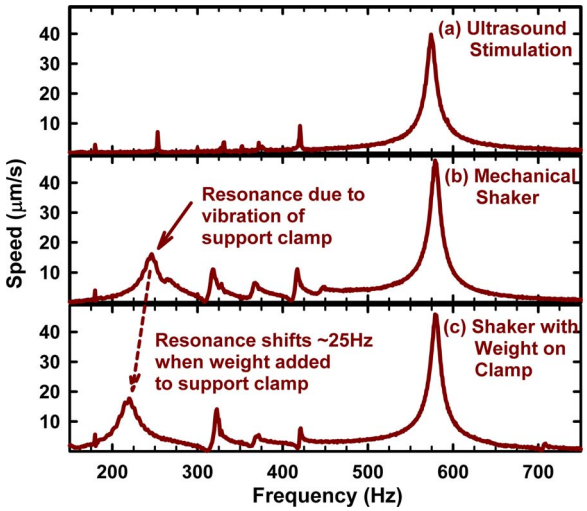


FIG. 5. (Color online) Illustration of efficacy of ultrasound stimulated excitation in suppressing background resonances. (a) Shows vibrometer spectrum for ultrasound stimulated excitation; (b) and (c) show excitation using a mechanical shaker. The background resonances of support structure in region between 200 and 450 Hz were strongly suppressed in the ultrasound stimulated spectrum. The difference between spectra (b) and (c) was that a 100 g weight was placed on one of the support clamps to demonstrate that the support system was origin of these resonances. The ~10 Hz shift in the primary resonance at 580 Hz between Figs. 4(a) and 4(b) was caused because the tuning wire was bumped slightly between these two runs, not because of difference in the resonances for ultrasound and shaker excitation.

V. CONCLUSIONS

In Ref. 5, the vibrational modes of an organ reed were excited using a mechanical shaker, and measured with a scanning laser vibrometer. The resonance frequencies measured in the current study, using ultrasound stimulation excitation are consistent with those previous measurements, as well as theoretical predictions. In the previous study, there was some concern about the validity of using a mechanical shaker to determine the vibrational modes, since the driving mechanism was indirect and there was some mass-load when the shaker was attached. The current study, which demonstrated that the modal frequencies were the same for both shaker and ultrasound stimulated excitation, appears to validate the previous measurements.

More importantly, the current study has demonstrated the efficacy of using airborne ultrasound stimulated vibrometry for modal analysis. Ultrasound stimulated excitation appears to have many advantages, including the noncontact na-

ture of the interaction, relatively large bandwidth, and the ability to focus the excitation into a relatively small region, in contrast to a shaker which is much more homogeneous. The combination of an ultrasound excitation source and a vibrometer allows for completely noncontact modal testing. This method may be advantageous over conventional techniques, particularly for items that would be significantly affected by the mass loading of a mechanical shaker. Also, if a broadband ultrasound transducer is used, it may be possible to excite structures at much higher frequencies than can be obtained using conventional mechanical shakers which are generally limited to less than about 20 kHz.

ACKNOWLEDGMENTS

The authors would like to thank Charles Hendrickson, president of the Hendrickson Organ Company, for providing the organ reed pipes used in this study, and Mario Pineda of Polytec for his useful discussions.

[1]Polytec Incorporated, Laser Doppler Vibrometer, http://www.polytec.de/polytec-com/I_vib/vib_uni_vib.html

[2]T. D. Rossing, J. Angster, and A. Miklós, "Reed vibration and sound generation in lingual organ pipes," Paper 2pMU6, 136th Meeting of Acoustical Society of America (1998); T. D. Rossing, J. Angster, and A. Miklós, "Reed vibration and sound generation in lingual organ pipes," ISMA2001 Proceedings (2001), p. 313.

[3]M. Fatemi and J. F. Greenleaf, "Ultrasound stimulated vibro-acoustic spectrography," Science **28**, 82 (1998); M. Fatemi and J. F. Greenleaf, "Vibro-acoustography: An imaging modality based on ultrasound-stimulated acoustic emission," Proc. Natl. Acad. Sci. U.S.A. **96**, 6603 (1999); J. F. Greenleaf and M. Fatemi, "Acoustic force generator for detection, imaging and information transmission using the beat signal of multiple intersecting sonic beams," U.S. Patent No. 5 903 516 (1999).

[4]M. Fatemi and J. F. Greenleaf, "A novel method for modal analysis of fine structures," Proceedings of the 2000 IEEE International Ultrasonic Symposium Short Courses (Oct 2000), p. 252; M. Fatemi, M. R. Zeraati, X. Zhang, and J. F. Greenleaf, "Comparative study of three mode shape measurement techniques using the radiation force of ultrasound," Proceedings of the 4th GRACM Congress on Computational Mechanics (June 2002); X. Zhang, M. Fatemi, and J. F. Greenleaf, "Vibro-acoustography for modal analysis of arterial vessels," Proceedings of the 2002 IEEE International Symposium on Biomedical Imaging (July 2002), p. 513; M. Fatemi and J. F. Greenleaf, "Mode excitation and imaging by the radiation force of ultrasound," J. Acoust. Soc. Am. **111**, 2472 (2002); X. M. Zhang, M. Fatemi, R. R. Kinnick, and J. F. Greenleaf, "Noncontact ultrasound stimulated optical vibrometry study of coupled vibration of arterial tubes in fluids," *ibid.* **113**, 1249 (2003); F. G. Mitri, P. Trompette, and J.-Y. Chapelon, "Detection of object resonances by vibro-acoustography and numerical vibrational mode identification," *ibid.* **114**, 2648 (2003); M. Fatemi, X. Zhang, and J. F. Greenleaf, "Mode shape measurement and imaging by ultrasound," SIAM Conference on the Life Sciences (July 2004).

[5]T. M. Huber, B. A. Collins, M. Pineda, and C. Hendrickson, "Vibrational modes of the reed in a reed organ pipe," J. Acoust. Soc. Am. **114**, 2348 (2003); available online at http://physics.gustavus.edu/~huber/presentations/

[6]A. Miklós, J. Angster, S. Pitsch, and T. D. Rossing, "Reed vibration in lingual organ pipes without resonators," J. Acoust. Soc. Am. **113**, 1081 (2003).

[7]P. J. Westervelt, "Theory of steady force caused by sound waves," J. Acoust. Soc. Am. **23**, 312 (1951); F. E. Borgnis, "Acoustical radiation pressure of plane compressional waves," Rev. Mod. Phys. **25**, 653 (1953).

[8]Polytec PSV-300 Scanning Vibrometer, www.polytec.de/polytec-com/l_vib/scanning_vib.html

[9]M. V. Barton, "Vibration of rectangular and skew cantilever plates," J. Appl. Mech. **18**, 129 (1951); A. W. Leissa, Vibration of Plates, NASA SP-160, Washington, D.C. (1969), reprinted by Acoustical Society of America (1993).

[10]Computer Program ME'Scope VT-860SL Visual STN Pro Version 4.0.0.72 (Vibrant Technology, 2005).

[11]M. Fatemi and J. F. Greenleaf, "Acoustic force generation by amplitude modulating a sonic beam," U.S. Patent No. 5 921 928 (1999).

[12]S. Chen, M. Fatemi, R. R. Kinnick, and J. F. Greenleaf, "Comparison of stress field forming methods for vibro-acoustography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**(3), 313 (2004).

Effect of the glottal source and the vocal tract on the partials amplitude of vibrato in male voices

Ixone Arroabarren[a] and Alfonso Carlosena[b]
Department of Electrical and Electronic Engineering, Universidad Pública de Navarra, E-31006 Pamplona Spain

(Received 7 October 2004; revised 22 January 2006; accepted 23 January 2006)

In this paper the production of vocal vibrato is investigated. The most relevant features of the acoustical vibrato signal, frequency and amplitude variations of the partials, will be related to the voice production features, glottal source (GS) and vocal tract response (VTR). Unlike previous related works, in this approach, the effect on the amplitude variations of the partials of each one of the above-mentioned voice production features will be identified in recordings of natural singing voice. Moreover, we will take special care of the reliability of the measurements, and, to this aim, a noninteractive vibrato production model will be also proposed in order to describe the vibrato production process and, more importantly, validate the measurements carried out in natural vibrato. Based on this study, it will be shown that during a few vibrato cycles, the glottal pulse characteristics, as well as the VTR, do not significantly change, and only the fundamental frequency of the GS varies. As a result, the pitch variations can be attributed to the GS, and these variations, along with the vocal tract filtering effect, will result in frequency and amplitude variations of the acoustic signal partials. [DOI: 10.1121/1.2177584]

PACS number(s): 43.75.Rs, 43.75.Bc, 43.75.−z, 43.70.−h [NHF] Pages: 2483–2497

I. INTRODUCTION

Singing voice constitutes an interesting challenge to the study of voice quality because of its differences from every day speech. Among these, the most interesting one is perhaps the vocal vibrato. This is a specific musical feature not present in speech and has been by itself the topic of interest of many researchers in distinct areas such as physiology and musicology.

It is quite easy to describe vocal vibrato from the acoustical point of view. Borrowing Sundberg's definition:[1] "vibrato is a regular fluctuation in pitch, timbre and/or loudness;" however, some of its basic aspects remain hidden still today.

From this definition, it is clear that vibrato is a regular pitch variation and, in fact, this is the most widely studied aspect from the pioneering work of Seashore[2] to our day. Several features related to the pitch variation, or, in parallel, fundamental frequency variation, have been analyzed, and the most relevant works are those dealing with the parametrization of this variation.[3–5] This is usually characterized by three parameters: the intonation, which represents slow variations around the average frequency value of the note; the vibrato extent, which represents the amplitude of fundamental frequency variations; and the vibrato rate, which represents the frequency of the fundamental frequency variations. Making use of these parameters, differences among singers have been measured and studied,[6,7] and even links between vibrato rate and physiological features have been explored based on a reflex resonance model.[8] In particular, this latter model is proposed to describe the relationship between the muscular activation and the typical values of vibrato rate.

Going back to Sundberg's definition, it is obvious that it does not specify what happens with timbre and loudness during vibrato. Both concepts are not unambiguously defined since they depend on the acoustic signal as well as the listener. Focusing on the acoustical signal, it can be seen as a particular combination of the amplitude of the different partials (a chord of sine tones of different frequencies) composing the voiced sound. The position of these partials is harmonically related to the fundamental frequency, and in the case of vibrato, it is expected that harmonicity is preserved. On the other hand, the amplitude of the partials also shows temporal variations during vibrato, but its origin and characterization still remains unclear, as will be shown later in more detail. These variations also reflect in the timbre and loudness temporal evolution, and, since the amplitude variations are not well understood, neither are the timbre and loudness variation.

Regarding the amplitude variation of the partials, several works have dealt with this topic during the last decades, but apart from minor differences among them, almost all simply measure these variations and argue about their perceptual relevance. In particular, the first step on this direction was taken by McAdams and Rodet,[9] who demonstrated that the amplitude behavior of a partial was coupled with its frequency behavior according to a given spectral envelope, which could be used by the auditory system to discriminate the timbre of a complex sound. Later, Horii[10] proposed a resonance-harmonic interaction to explain the phase relationship between fundamental frequency and amplitude modulation of vibrato. A parallel work was carried out by Maher and Beauchamp,[11] who concluded that the amplitude fluctuation

[a]Electronic mail: ixone.arroabarren@unavarra.es
[b]Electronic mail: carlosen@unavarra.es

of a partial during vibrato varies in form, amplitude, and phase according to the position of the partial within the vocal tract resonances. However, in neither case was it demonstrated how the vocal tract, as well as the glottal source (GS), behaves such that those variations are generated. In the same way, Imazumi *et al.*[12] investigated the sources of vibrato by considering the correlation between the amplitude and frequency variations of the partials, and argued about the interrelationships between the fundamental frequency, the level of the harmonics, the formant frequencies, and the overall amplitude of the singing voice. The main limitation of their approach is that they did not identify the individual contribution of the voice production elements, GS and vocal tract response (VTR), on the amplitude and frequency of the partials, making it difficult to demonstrate the precise behavior of these elements in vocal vibrato production.

The correlation between the amplitude and frequency variations of the partials associated with vocal vibrato has been subsequently analyzed by other authors,[13,14] and the main conclusion in all cases is that, while this correlation is determined by the vibrato production mechanisms, the exact role of each voice production element cannot be identified based only on the direct observation of the amplitude versus frequency representation, since this represents a global information where the individual contributions of the voice production elements are combined.

By analyzing the different efforts devoted to the study of amplitude variation of the partials, it can be seen that all the works have in common the use of the sinusoidal model[15,16] to represent the acoustic signal, since the parameters of this model are the amplitude and frequency variations of the partials. However, in the mentioned works no voice production description is considered, which precludes any statement concerning the vibrato production mechanisms.

Additionally, regarding the voice production process, different mathematical models have been proposed during the last decades in order to describe it. For instance, physical models are closer to the voice production system,[17,18] since the GS and VTR are obtained based on some physical features of the voice organ elements. Additionally, the source-filter model proposed by Fant[19] (also known as the noninteractive source-filter model) can be qualified as a signal voice production model, since it does not contain any physical parameters, but it directly depends on a waveform representing the GS and a transfer function corresponding to the VTR. Each model represents a tradeoff in the voice production descriptions, for instance physical models provide a more realistic description, but is not easy to extract the values of the model parameters from a given acoustic signal. On the contrary the source-filter model constitutes a more simplified description, but it makes it easier to extract the GS and VTR for a given recorded sound.

With this in mind, in this paper the vibrato production will be investigated, but, unlike earlier works, the individual contribution of the GS and VTR on the amplitude variation of the partials will be identified. For this purpose, the acoustical features of the vibrato signal, amplitude, and frequency variations, as well as the voice production process, will be jointly considered by combining two different signal descriptions: the sinusoidal model and the noninteractive source-filter model. At this point, it is important to emphasize how different the two approaches are, which makes it very important to assess the reliability of the measurements that will help to combine such different approaches.

For this purpose, we will propose a vibrato production signal model, the noninteractive vibrato production model, that will allow us to validate the measurements made in natural singing voice recordings and will describe, and thus help to understand, how the voice production elements, GS and VTR, behave such that the specific correlation between the amplitude and frequency of the partials is produced. Finally, a model evaluation process will allow us to examine if the behavior of the proposed model fits well with natural vocal vibrato production. It is important to note that the voice material considered for this study corresponds to male recordings and, thus, the conclusions derived from it are of application strictly to this kind of voice.

The organization of the work is as follows. In Sec. II the acoustical properties of vibrato are described, where the sinusoidal model is used to describe the acoustic signal, and, based on its parameters, the amplitude variations of the partials, as well as the intensity of the sound, is analyzed. In Sec. III the voice production process is briefly described by the noninteractive source-filter model, defining its main features (GS and VTR) and briefly reviewing the analysis techniques associated to this particular model.

In Sec. IV the vibrato production is examined, and its noninteractive vibrato production model will be proposed. Finally, in Sec. V, this model is evaluated, and the conclusions that follow are discussed in Sec. VI.

II. ACOUSTICAL CHARACTERIZATION OF VIBRATO

Before starting with the acoustical analysis, there are some details that must be clarified: First, it has to be recalled that vibrato is a time-dependent musical effect and, thus, its features can be affected by its musical context (pitch transition,[20] crescendo, decrescendo,[21] etc.). For this particular study, a simple musical context has been selected in such a way that other musical effects are avoided or at least minimized. Thus, the specifications of the voice material used for this study will be briefly described.

Second, as already mentioned, the sinusoidal model has been described as the most suitable signal model for vocal vibrato description. However, there are different possibilities for the calculation of its parameters from a given recorded signal,[11,13,14] which might affect the final result. Therefore, it is necessary to provide a brief description of the particular analysis procedure adopted for this study.

Third, from Sundberg's definition of vocal vibrato mentioned above, the main conclusion is that the less-understood acoustical aspects are the timbre and loudness. In particular, regarding the pitch variations characterization, the authors refer the readers to a recent work of Arroabarren *et al.* concerning pitch variations.[5] On the other hand, timbre and loudness are also dependent on the perceptual system of the listener,[22] thus it is important to note that in this particular

study we will only focus on the objective acoustic signal features, frequency and amplitude variations, as well as the intensity variations of the sound.

A. Voice material

During a short musical excerpt (a few seconds) of natural singing voice performance many acoustical changes may take place, such as, for instance, pitch and intensity variations. Therefore, for a sensible acoustical characterization of vocal vibrato a very simple musical context has been selected, in order to make easier the isolation of vibrato out of other musical effects, such as those mentioned.

For this study three semi-professional male singers were enlisted, two tenors and one baritone. Each singer was asked to sing an exercise of three notes, the second one corresponding to a long sustained note separated three semitones from the others (Do-Mi-Do). The first and third notes behave as a musical support for the second one, so that the singer can sing a long sustained vowel, and the sung vowel of these notes was the vowel [i]. This exercise was repeated by each singer, increasing the frequency of each note in one semitone, so that their whole frequency range was covered. Regarding the sung vowel of the central note, the whole exercise was repeated for the five Spanish vowels, [a], [e], [i], [o], [u], in order to have available several vocal tract configurations.

Recordings were made in a professional recording studio, in such a way that reverberations were reduced considerably though not completely eliminated as in an anechoic chamber.[14] The signals were sampled at a standard rate of 44.1 kHz, with 16-bit linear encoding on a single channel, and the recording levels were adjusted on the DAT machine during the recording process to achieve a (roughly) constant output level. These changes were necessary to account for the large dynamic range produced by the singer across his tessitura. In this way, the full 16 bits of resolution were used for each note produced. This means that the relative power of the notes sung is not preserved in our recordings; however, we deemed it more important to avoid quantization error than to preserve relative power.

B. Sinusoidal modeling

Concerning the acoustic signal analysis, the additive analysis-synthesis approach has been selected. It was developed during the 1980's by J. O. Smith III[23] in the context of computer music, and, in parallel but independently, McAulay and Quatieri proposed a similar approach for speech applications.[15] Later, a more complete sinusoidal model was proposed by Serra and Smith,[16] avoiding the problems of the preliminary versions in nonharmonic sounds. In particular, in the preliminary versions of the model, the whole sound was modeled by a set of sinusoids. However, in the case of nonharmonic sounds the number of sinusoids required to represent the sound was too high. Therefore, in the sinusoidal model proposed by Serra and Smith, an additional term was included to represent the nonharmonic part of the sound, and only the harmonic part of the sound was represented by a set of sinusoids.

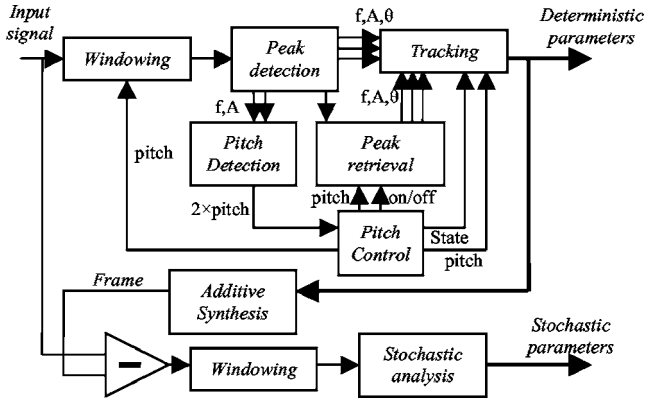


FIG. 1. Block diagram of the analysis part of the additive analysis-synthesis approach.

The continuous time sinusoidal model, which this analysis-synthesis scheme is based on, assumes the following expression for the acoustic signal $s(t)$:

$$s(t) = \sum_{i=1}^{M} a_i(t)\cos\,\theta_i(t) + e(t), \tag{1}$$

$$\theta_i(t) = 2\pi \int_{-\infty}^{t} f_i(\tau)d\tau \tag{2}$$

$$e(t) = \int_{0}^{t} h(t,\tau)u(\tau)d\tau, \tag{3}$$

where $a_i(t)$ and $f_i(t)$ are the instantaneous amplitude (IA) and instantaneous frequency (IF) of partial i, respectively, which characterize the deterministic component of the signal, and $e(t)$ is the stochastic one. The stochastic component $e(t)$ is usually described as a filtered white noise $u(\tau)$ by a time-varying filter characterized by its impulse response $h(t,\tau)$. In natural sound analyses it is seldom clear which spectral components are associated with each part, hence, the estimation algorithm has to be flexible enough.

To carry out our particular analysis we have implemented a tool based on the sinusoidal model that is described by the block diagram of Fig. 1.

According to the block diagram of Fig. 1, the input signal is analyzed frame by frame, with a typical frame duration of 6 ms. Then, in the first step a short time window length of the signal (three or four fundamental periods) is selected. The short-term spectrum of the signal is calculated and the spectral peaks detected, estimating their frequencies, amplitudes, and phases. This information is conveyed to a pitch detection block, where the fundamental frequency (F_0) of the frame is estimated. As shown, this is a pitch synchronous analysis, because the pitch information is used to select the most appropriate window length. Once the spectral peaks are detected, and the pitch estimated, a peak retrieval routine is applied, in order to recover nondetected peaks. Considering the spectral peaks of the frame, they are linked to earlier frame analyses in the peak tracking step, generating the trajectories of each partial, and obtaining the sinusoidal parameters of (1), namely, $a_i(t)$ and $f_i(t)$, the IA and IF of the signal

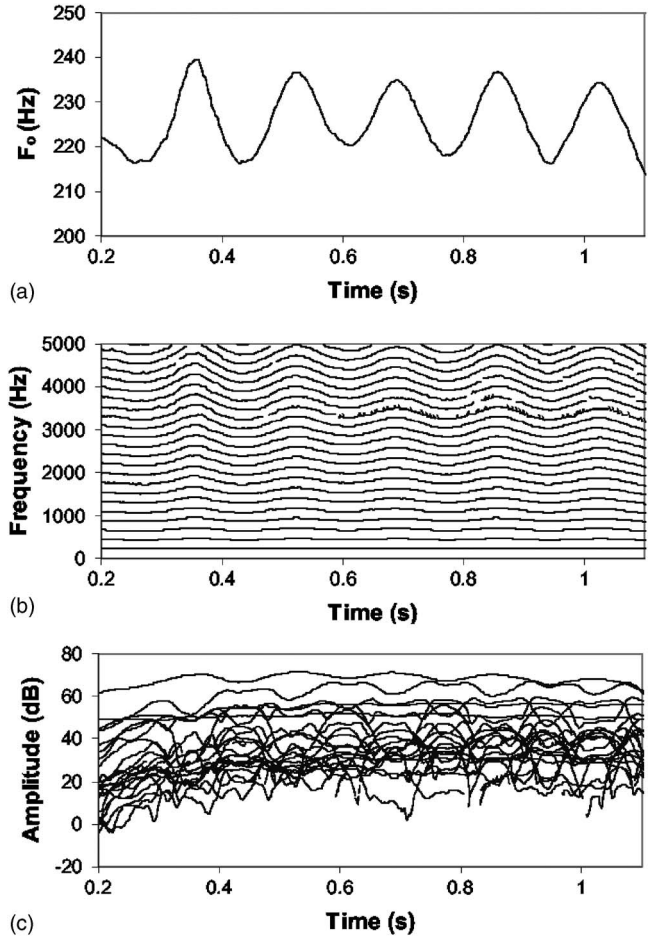


FIG. 2. Sinusoidal model features. The signal corresponds to a tenor recording, singing the vowel [a] with a F_0 of 220 Hz. The selected analysis frame size is 6 ms. (a) F_0 (b) IF (c) IA.

harmonics, respectively. Finally, the stochastic part is estimated by removing the deterministic part from the original signal. For more details of this analysis procedure readers are referred to the works of McAulay and Quatieri[15] or Serra and Smith.[16]

Going back to the recorded samples, the vocal vibrato recordings correspond to clean solo recordings and only voiced sounds, since there is no consonant in the musical exercise. Moreover, they correspond to the normal phonation, where the aspiration noise is minimum,[24] so that the whole signal can be modeled deterministically. In Fig. 2 results obtained for a representative tenor recording are shown.

From Fig. 2 it is apparent that all the signal partials are harmonically related to the fundamental frequency, and all of them follow the vibrato pattern, which is more evident as the harmonic order increases. Regarding the amplitude of the partials, in some cases it is possible to see a semi-periodic pattern but, in contrast to the frequency variation of the partials, no correlation among their IA is obvious.

C. Frequency and amplitude of the partials

In this subsection, the frequency and amplitude variations of the partials associated with vibrato will be analyzed, as well as the intensity of the sound, and some conclusions about their behavior during vibrato will be drawn.

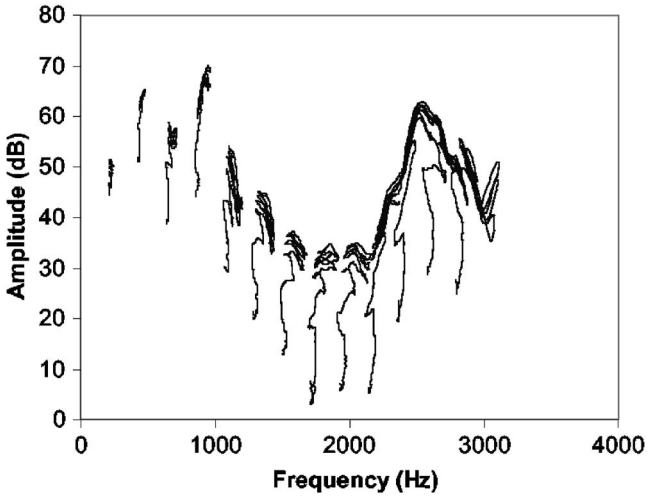


FIG. 3. AM-FM representation. The signal corresponds to a tenor recording, singing the vowel [a] with a F_0 of 220 Hz.

In Fig. 2 the frequency and amplitude variations of a set of partials of a tenor recording are shown. They are representative of a typical behavior during vocal vibrato. From Fig. 2(c) it is clear that, unlike the frequency variation of the partials, the amplitude variation does not follow a regular pattern, as all the harmonics show a different amplitude variation; however, in most of the partials it is still possible to see a semi-periodic pattern in their amplitude variation synchronous with the frequency variation.

In order to show the implicit correlation between the amplitude and frequency variations, the amplitude variation of the partials shown in Fig. 2(c) are now represented in Fig. 3 versus their frequency variation appearing in Fig. 2(b). This representation has been used by several authors, as already explained, and has received different names, *average frequency characteristics*,[12] *composite transfer function*,[13] or *AM-FM representation*.[14] All these names imply that this information reflects the dependence on both vocal tract and the spectral variation in the glottal source.

However, by observing Fig. 3 it is not possible to see how the GS and the VTR behave during vibrato, since the effect of the two elements is combined. However, the AM-FM representation shows some common features concerning the IA of the partials, which were mathematically described recently.[25] The IA of a partial, $A_i(n)$, was modeled as the product of three elements:

$$A_i(n) = \alpha_i \beta_i(n) \gamma_i(n) \tag{4}$$

where α_i is the relative weight of the harmonic compared to the fundamental, $\beta_i(n)$ is a time-varying function representing the intensity change of the partial, which models the amplitude variation independently of the spectral envelope tracing, and $\gamma_i(n)$ is a time-varying function representing those variations related to the spectral envelope traced by the AM-FM representation. These three elements can be identified by observing the AM-FM representation of Fig. 3. From Fig. 3 it is apparent that when the IF of the partials shows its periodic pattern, their IA traces a local spectral envelope, which is related to the $\gamma_i(n)$ parameter of the IA, representing the local spectral envelope traced by each partial.

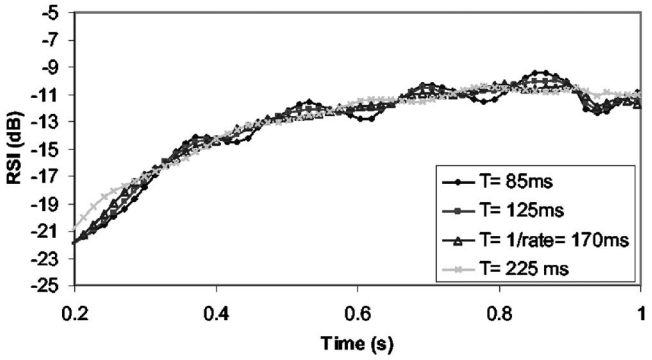


FIG. 4. Relative sound intensity. The signal corresponds to a tenor recording, singing the vowel [a] with a F_0 of 220 Hz.

In spite of this mathematical description, the main conclusion about the amplitude variations during vibrato is that a direct relationship with the voice production features cannot be extracted by a simple observation of the AM-FM representation, since the amplitude of the partials will be affected by all the voice production elements. However, the AM-FM representation gives at least a clue that the amplitude behavior of the partials is not arbitrary, but rather involves the specifics of vocal vibrato production.

On the other hand, regarding the intensity variations, they can be easily characterized based on the former analysis. The intensity of the sound is defined as the sound energy transmitted per unit time through a unit area, and it is proportional to the square of the sound pressure.

We have already mentioned that, during the recording process, no calibration was made for determining the exact value of vocal intensity, because the main concern of the measurement was to avoid the quantization error. Therefore, in order to quantify the sound intensity, the *relative sound intensity* (RSI) will be calculated from the recorded signal samples as

$$\mathrm{RSI}(n)[\mathrm{dB}] = 10\log_{10}\left(\frac{1}{N}\sum_{k=n-(N/2)}^{n+(N/2)} s^2(k)\right), \tag{5}$$

where N is the window length and $s(k)$ are the samples of the acoustical signal. According to this definition, the RSI will be used for a relative quantitative description of the sound intensity evolution during vibrato.

Considering the RSI definition of (5), the IA model of (4), and depending on N values, the resulting intensity may or may not show a periodic variation pattern. If the window length were selected such that the $\gamma_i(n)$ component of the partials is compensated for, the resulting RSI would not show periodic variations, as it is shown in Fig. 4, where the N values are represented by the T values in seconds, for a sampling frequency, f_s, of 44.1 kHz, and $T=N/f_s$. In Fig. 4 four RSI calculations are shown, corresponding to different N values. Comparing these results the RSI does not show periodic variations when the window length is fitted to one vibrato period.

To conclude, we can say that sound intensity variations are possible depending on the averaging window selected to calculate this magnitude. These variations will be a consequence of the frequency and amplitude variations of the partials.

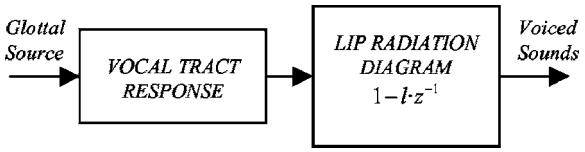


FIG. 5. Block diagram of the noninteractive source-filter model.

III. NONINTERACTIVE SOURCE-FILTER MODEL

Once the main features of the acoustical analysis of vibrato have been determined, the voice production process will be described. In this way, we will determine which voice production elements must be considered in order to understand the production of vocal vibrato. For this purpose, the noninteractive source-filter model will be considered as the voice production model,[19] because it is simpler than physical models and will make it easier to link the sinusoidal model to the voice production mechanisms. Thus, a brief review of this voice production description will be provided, as well as some important key points about the signal analysis associated to this model.

A. Model definition

The noninteractive source-filter model can be represented by the block diagram in Fig. 5. There, the voice production is modeled by the glottal source (GS) that is linearly modified by the vocal tract response (VTR) and the lip radiation impedance, which is approximated by a derivative system. Typically, the lip radiation system is usually combined with the GS, in such a way that the glottal source derivative (GSD) is considered as the vocal tract excitation. It is important to note that this model is denominated as the noninteractive source-filter model since both voice production elements are considered independent. This assumption does not exactly hold in natural voice production.[26–28]

Additionally, and in voiced sounds, the GS excitation is a periodic pulsed signal, determining the vocal texture. The most extended glottal pulse characterization defines five independent parameters: fundamental frequency, F_0, or fundamental period T_0, amplitude of voicing, Av, open quotient, O_q, asymmetry coefficient, α, and return phase interval T_a, which is related to the spectral tilt, f_t. All of the proposed GS waveform mathematical models try to include these parameters, such as for instance the LF model[29] or the KLGLOTT88.[24]

In Fig. 6, an idealized GS pulse and its derivative are shown, illustrating the five parameters defined. From this figure, it is clear that O_q controls the time interval where the GS is not null; α, defined as $\alpha=T_P/O_qT_0$, determines the symmetry degree of the glottal pulse; and T_a parametrizes the closure abruptness of the waveform.

Regarding the VTR, it is normally modeled by an all-pole autoregressive (AR) filter

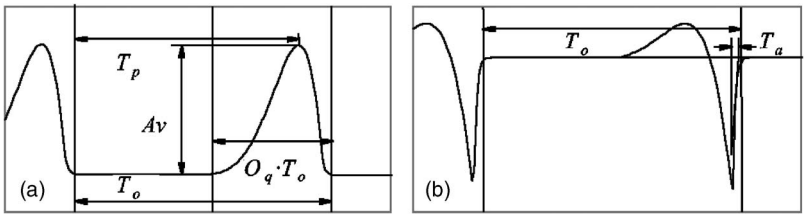


FIG. 6. Glottal pulse parametrization. (a) GS. (b) GSD.

$$H(z) = \frac{G}{1 - \Sigma_{k=1}^{p} a_k z^{-k}}, \tag{6}$$

where a_k are the prediction coefficients and G is the gain of the all-pole filter. This model corresponds to nonnasal sounds, as it is the case in our particular recordings.

B. Analysis procedure

As illustrated in Fig. 5, the noninteractive source-filter model is a very simple voice production one, but the analysis techniques associated to it allow extracting from the acoustic signal the information of the GS and the VTR, which is not the case in other voice production models. However, when this model is considered, a thorough analysis should be carried out, particularly in singing voice analysis, for a correct interpretation of the results. Therefore, some key issues pertaining to this model and its associated analysis procedures will be highlighted.

In natural voice production, the physical system is composed by two resonant cavities (subglottal and supraglottal cavities), which are connected by a valve, the glottis, where the vocal folds are located. In voiced sounds, the vocal folds are opened and closed providing the harmonic nature to the air flow, but, additionally, the vocal tract response varies during a single fundamental period because both cavities are connected and disconnected along every fundamental period, resulting in a nonlinear element. This effect is one of the effects associated to the source-tract interaction,[26,30] and it is not included in the noninteractive source-filter model.

On the other hand, the complete analysis procedure can be decomposed into two different steps: the first one is represented by the inverse filtering techniques, which allow us to separate both the GS and the VTR based on the acoustic signal, and the second one corresponds to a parametrization step, where the GS and VTR parameters are obtained in order to reduce to a few numerical values both voice production elements.

Regarding the inverse filtering techniques, there are many different approaches available in the literature,[31–35] however, since all of them assume a simplified voice production description, the resulting GS waveform will be affected in a different way by the so-called formant ripple resulting from the source-tract interaction and incomplete cancellation of the formants. Additionally, all of them have a fundamental frequency dependence, which limits their application as the fundamental frequency of the signal increases, and thus some singing voice signals.[36]

On the other hand, the GS parametrization is an important analysis step, particularly when a large number of signals are to be analyzed, and there are available several possible methods:[37] We have, for instance, direct estimation methods, where the source parameters are obtained by measuring time domain landmarks. However, they are not very robust because those landmarks are not easy to determine in inverse filtered signals, particularly the opening instant. Additionally, there are fitting estimation methods, where a mathematical model is fitted to the inverse filtered GSD. However, all of these approaches have a high computational load. Alternatively, the normalized amplitude quotient (NAQ) has been recently proposed[38] as a suitable parametrization method. It is amplitude and fundamental frequency normalized, avoids time domain landmarks measurement, and has a minimum computational cost. It is defined as the quotient of the maximum of the GS and the minimum of the GSD. Moreover, it has been shown that it is also a global parameter as it depends on the three above-mentioned GS parameters.[39]

To summarize these paragraphs, it is important to note that the noninteractive source-filter model allows for a simple voice production description, since it is a signal model. However, there are different factors that must be taken into consideration for a correct interpretation of the obtained results.

In order to illustrate the application of this kind of technique to the voice material described in Sec. II A, the GSD and VTR corresponding to a representative baritone recording are shown in Fig. 7. In this particular case three representative inverse filtering techniques have been selected: the analysis by synthesis (AbS) approach,[31] the glottal spectrum based (GSB) inverse filtering,[34] and the closed phase covariance (CPC).[35] From Fig. 7 we can conclude that the three techniques provide similar results. Regarding the GSD, three fundamental periods are represented, illustrating a similar pulse shape to that in Fig. 6. Concerning the VTR, it can be seen that, as this recording corresponds to a male singing voice recording, the highest formants are concentrated in one frequency region, ranging from 2000 to 3000 Hz, which is known as the singer's formant.[22]

It is interesting to note that the inverse filtering analysis is a short time analysis (the window length is three or four fundamental periods), compared to the slow fundamental frequency variations of vibrato (a typical vibrato rate is about 5 Hz). Then, it can be said that these techniques are insensitive to the presence or absence of vibrato, since they are based on short time window analysis.

Therefore, one question arising is how VTR and GSD parameters evolve during vibrato. Additionally, by comparing Fig. 3 to Fig. 7(b), the AM-FM representation is very similar to the VTR of Fig. 7(b). However, in the case of the

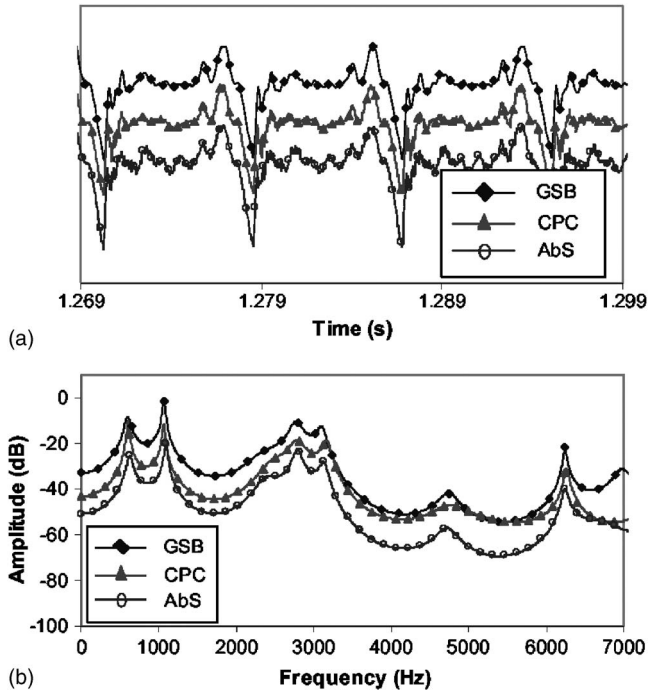


FIG. 7. Inverse filtering results. The signal corresponds to a baritone recording, singing the vowel [a] with a F_0 of 123 Hz. (a) GSD. (b) VTR.

AM-FM representation, no source-filter separation has been made, and thus both elements are melted in this representation.

It can be concluded that in natural vocal vibrato nothing can be said, *a priori*, about what is happening with the GSD and VTR during vibrato, but, whatever happens, the resulting global AM-FM representation provides a spectral envelope very similar to the VTR obtained by inverse filtering.

IV. NONINTERACTIVE VIBRATO PRODUCTION MODEL

After analyzing the acoustic signal corresponding to vibrato, assuming the sinusoidal model, and reviewing the voice production process based on the noninteractive source-filter model, it might seem difficult to establish a relationship between these two different mathematical descriptions of the acoustic signal and, even more, to identify the individual contribution of the voice production elements to the amplitude variation of the partials. With this in mind, a noninteractive vibrato production model will be proposed in order to describe, and help to understand, how the GSD and the VTR behave during vibrato such that the specific amplitude variation of the partials comes out. This model will allow us to assure the goodness of the analysis carried out in natural singing voice recordings. In particular, this signal model will permit us to generate synthetic signals that will be subsequently analyzed in the same way as natural singing voice recordings, which will help to infer what is happening in vocal vibrato production.

Before proposing the vibrato production model, some basic assumptions will be made regarding the behavior of GSD and VTR during vibrato. Assuming that the RSI, calculated by using a window length equal to one vibrato period, does not change during a few vibrato cycles, we have the following.

(1) The GSD characteristics, or glottal pulse shape features, O_q, α, and f_t, remain also constant during vibrato, and only the fundamental frequency of the voice changes. This assumption is justified by the fact that, perceptually, there is no phonation change during a single note.
(2) The VTR does not appreciably change along with vibrato. This assumption supports the fact that vocalization does not change along the note.

Taking into account these assumptions, the proposed noninteractive vibrato production model is represented by the block diagram of Fig. 8.

As illustrated in Fig. 8, for a given vowel where the RSI does not significantly change, the GSD parameters, controlling the glottal pulse shape, along with the VTR features, remain constant while the fundamental frequency of the excitation varies.

This model anticipates that a long-term relationship can be established between the GS and VTR and the AM-FM representation: Taking into account that the GSD features remain constant during vibrato, the AM-FM representation of each harmonic should represent a local section of the VTR, and each representation will be shifted (linearly distorted) in amplitude depending on the GSD spectral shape. Therefore, by removing the GSD effect from the AM-FM representation, only the VTR will be left.

This relationship is graphically shown in Fig. 9 for a synthetic signal described by the proposed model. For this particular example, the GSD has been modeled according to the LF model, and its parameters correspond to a normal phonation glottal pulse shape. The VTR filter corresponds to vowel [a] of a baritone, and the three vibrato parameters remain constant during the note: the intonation 100 Hz, the extent value of 10 Hz, and a rate value of 5 Hz. The resulting signal has been analyzed by both inverse filtering (GSB inverse filtering algorithm), where the presence or absence of vibrato has no influence on the algorithm, and by a sinusoidal model, where the IA and IF of each harmonic need to be measured. Results obtained for this simulation are shown in Fig. 9.

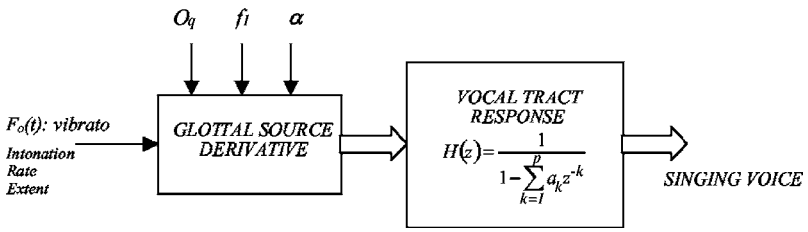


FIG. 8. Noninteractive vibrato production model.

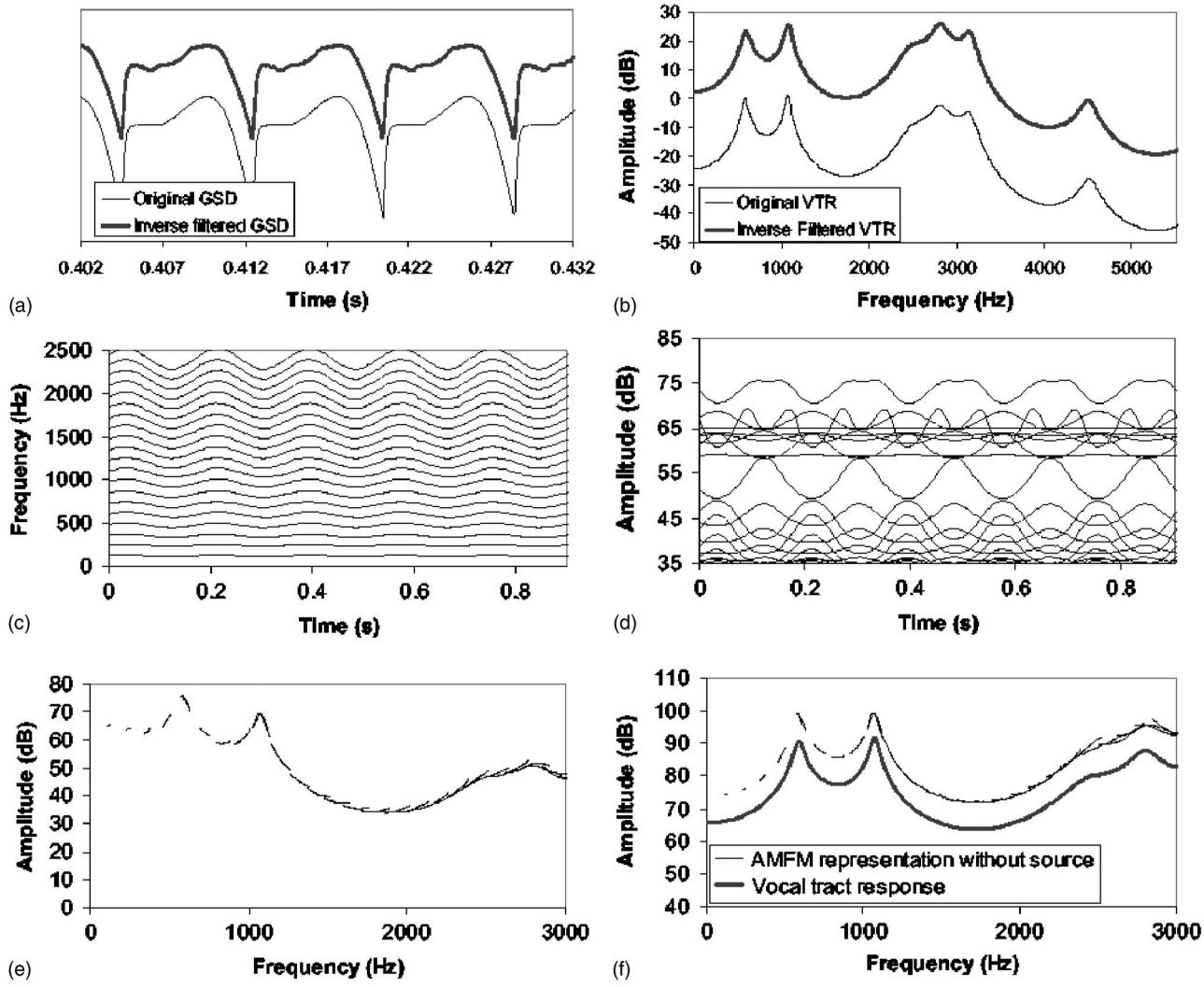


FIG. 9. Relationship between voice production and sinusoidal model for the synthetic signal. Inverse filtering results. (a) GSD. (b) VTR. Sinusoidal modeling results. (c) IF. (d) IA. (e) AMFM representation. (f) AMFM representation without source effect.

In Figs. 9(a) and 9(b) inverse filtering results are shown for a short window analysis. Since fundamental frequency is low, GSD and VTR are well separated. In Figs. 9(c) and 9(d), sinusoidal modeling results are shown. The frequency variations of the signal harmonics are clearly observed and the resulting amplitude variation, too. On the other hand, in Fig. 9(e) the AM-FM representation of the partials is shown. Taking into account the AM-FM representation of every partial, and comparing it to the VTR shown in Fig. 9(b), it is possible to conclude that local information of the VTR is provided by this representation. However, since no source-filter decomposition has been carried out, each AM-FM representation is offset in amplitude depending on the GSD spectral features. This effect results from keeping GSD parameters constant during vibrato. Comparing Figs. 9(e) and 9(b), it can be guessed that if the GSD magnitude spectrum were removed from the AM-FM representation of the harmonics, the resulting AM-FM representation would be very similar to the VTR. The result of this operation is shown in Fig. 9(f), and it can be seen that the compensated AM-FM representation is very close to the VTR.

For this noninteractive vibrato production model, the individual contribution of the GS and the VTR on the amplitude variation of the partials has been identified. In particular, when inverse filtering works, the GSD effect can be removed from the AM-FM representation provided by the sinusoidal model and the VTR information is isolated.

At this point, the relationship between the two signal models, noninteractive source-filter model and sinusoidal model, has been established for a synthetic signal where vibrato has been included under two assumptions stated at the beginning of the section. Now the question is if this relationship also holds in natural singing voice, where many other effects are present. Therefore, both kinds of signal analyses will now be applied to natural singing voice recordings. In order to set up conditions similar to the simulated signal, some precautions have been taken in the recording process:

(1) The musical context has been selected in order to control intensity variations of the sound, as detailed in Sec. II A.
(2) Recordings have been made in a studio, where reverberations are reduced. Under these conditions the

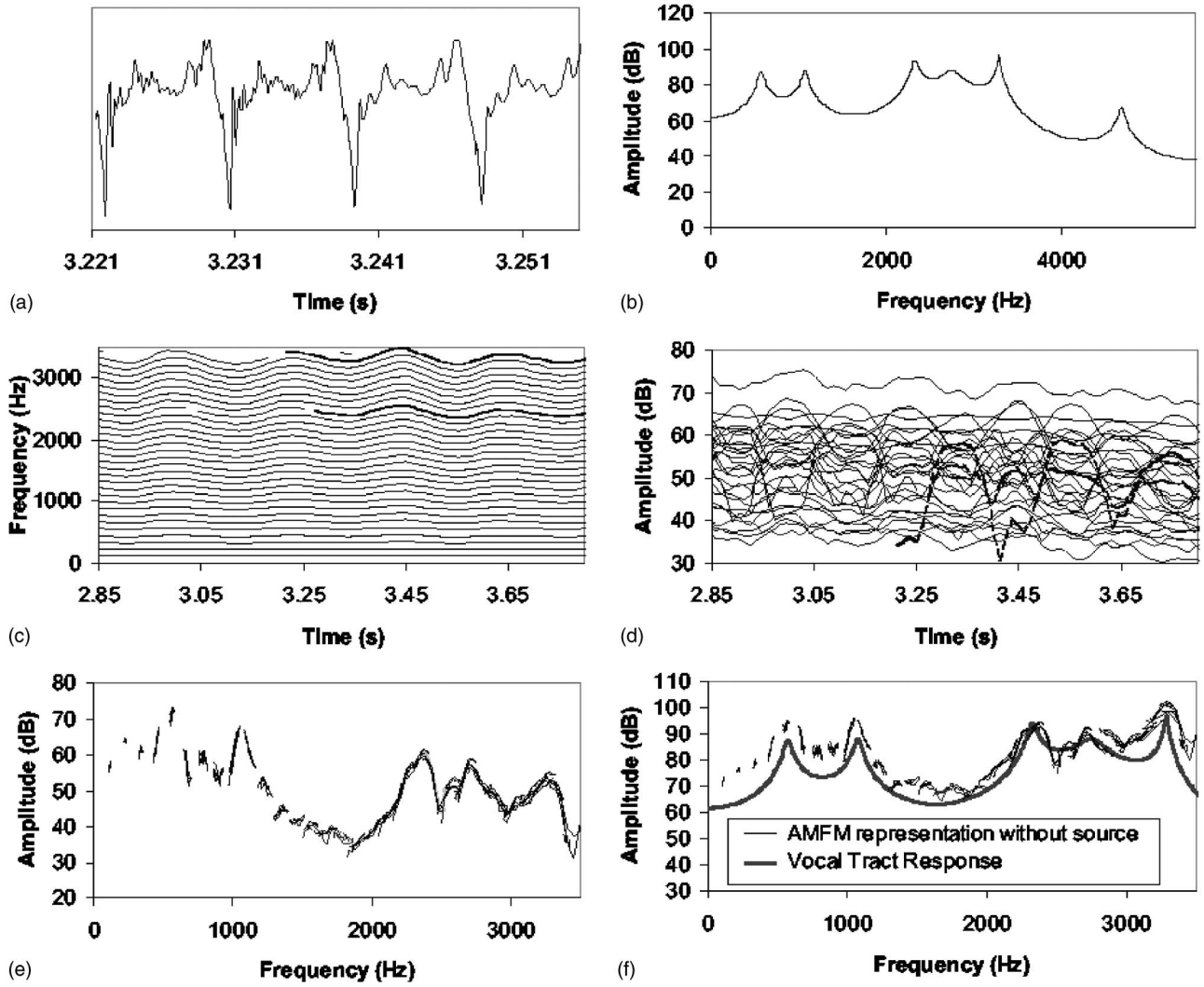


FIG. 10. Relationship between voice production and sinusoidal model for natural singing voice. Inverse filtering results. (a) GSD. (b) VTR. Sinusoidal modeling results. (c) IF. (d) IA. (e) AMFM representation. (f) AMFM representation without source effect.

AM-FM representation will present slight variations from the actual VTR, but it is still possible to carry out the same study.[14]

In Fig. 10 these analysis results are shown for a baritone recording, characterized by a F_0 of 107 Hz and a vocal tract configuration corresponding to vowel [a]. Unlike the measurements shown in Figs. 9(a) and 9(b), there is no reference for the original GSD and VTR. By comparing Figs. 9(c) and 10(c), the IF variation is similar in simulation and natural singing voice. However, the vibrato extent in this baritone recording is lower than in synthetic signal. In the case of the IA, natural singing voice results are obviously not as regular as synthetic ones. This is because of reverberation and irregularities of natural voice. Regarding the RSI of the sound, there are not large variations on the IA and, so, for one or two vibrato cycles, it can be considered constant.

In this situation, the AM-FM representation of the harmonics, shown in Fig. 10(e), is very similar to the synthetic signal AM-FM representation of Fig. 9(e), though the already mentioned irregularities are present. Now, the so-obtained GSD spectrum will be used to extract from the AM-FM representation the VTR information. The result of this operation is shown in Fig. 10(f) and, as in the case of the synthetic signal, the compensated AM-FM representation is very close to the VTR.

In this way, it can be concluded that, as in the case of the synthetic signal, the individual contribution of the voice production elements has been identified from the AM-FM representation corresponding to natural vocal vibrato, since the compensated AM-FM representation is similar to the VTR obtained by inverse filtering. However, the matching is not as close as for the synthetic signal.

For not limiting ourselves to a unique signal analysis, the same analysis procedure has been applied to a representative set of signals of the database described in Sec. II A, and their corresponding results are shown in Fig. 11. For this particular set of signals, recordings corresponding to the three male singers have been selected (tenor 1, tenor 2, and baritone), considering also different vocal tract configurations ([a], [e], [i]). Additionally, in order to analyze the behavior of the model in different situations, three synthetic signals have been added to the comparison, representing also

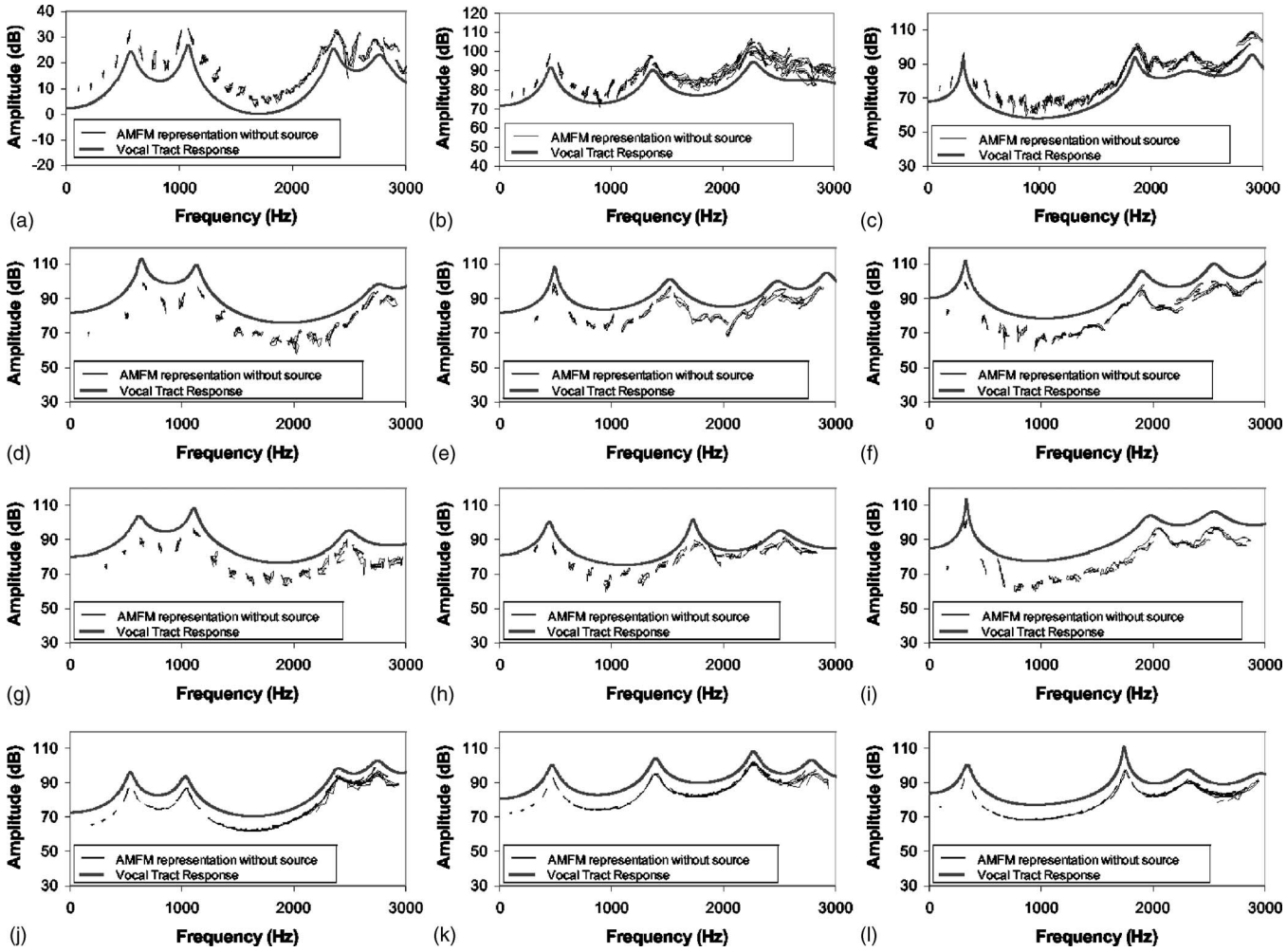


FIG. 11. Compensated AM-FM representation for a representative set of recordings. For each singer three different vocal tract configurations have been selected, as well as for the noninteractive vibrato production model. The results corresponding to different voices have been distributed in different rows, and each column corresponds to the same vocal tract configuration. Baritone recordings: (a) vowel [a], (b) vowel [e], and (c) vowel [i]. Tenor 1 recordings: (d) vowel [a], (e) vowel [e], and (f) vowel [i]. Tenor 2 recordings: (g) vowel [a], (h) vowel [e], and (i) vowel [i]. Synthetic signals: (j) vowel [a], (k) vowel [e], and (l) vowel [i].

different vowels. In order to give more details of these recordings, in Table I the values of the three fundamental frequency parameters (intonation, vibrato rate, and vibrato extent) are collected. It can be seen that all the signals have a low fundamental frequency, so that the main problems associated to the inverse filtering techniques are avoided, and in particular the fundamental frequency of the baritone recordings is slightly lower than the one corresponding to the tenor recordings. In the case of the synthetic signals, the fundamental frequency was chosen to be 100 Hz, the vibrato rate 5.5 Hz, and the vibrato extent 84.46 cents.

TABLE I. Fundamental frequency parameters for the representative data set. This data set covers different vocal tract configurations for the three male singers.

		Rate (Hz)	Intonation (Hz)	Extent (cent)
Baritone	[a]	4.47	110.86	48.20
	[e]	4.60	109.92	60.51
	[i]	4.46	108.21	51.69
Tenor #1	[a]	5.98	166.54	43.62
	[e]	5.16	161.60	70.02
	[i]	5.28	166.48	74.99
Tenor #2	[a]	5.89	160.25	27.65
	[e]	5.60	162.27	26.57
	[i]	5.74	161.44	34.50

By observing Fig. 11 the different vocal tract configurations are evident, because of the different formant positions. Additionally, since all the recordings correspond to singing male voices, the singers' formant is in the high-frequency region (above 2000 Hz). Also, the differences between the fundamental frequencies of the signals can be observed on the AM-FM representation, since for the same frequency region (0–3000 Hz) there are more harmonics in the case of the synthetic signal or in the baritone recordings. On the other hand, it can be seen that in all cases, different singers and vocal tract configurations as well as in the synthetic case, the compensated AM-FM representation is very close to the VTR, as was pointed for Fig. 10, which allows us to conclude that the noninteractive vibrato production model can explain, in an approximated way, what is happening in singing voice when vibrato is present.

V. MODEL EVALUATION

In the last section, a noninteractive vibrato production model has been proposed in order to relate the amplitude and frequency variation of the partials to the voice production elements. The key point of the model, as should be clear from our analysis, is the fact that the GSD and VTR features, O_q, α, and f_t, and formants central frequencies and bandwidths, respectively, remain almost constant during vibrato. In order to confirm these assumptions, the proposed model will be evaluated and further analyses will be carried out. To this purpose, the GSD and VTR features will be measured and their time evolution evaluated.

The evaluation procedure will be as follows: First, inverse filtering techniques will be applied for the GSD and VTR estimation in such a way that the analysis window will be moved over time. Later, and using the appropriate VTR and GSD parametrization, the behavior of these elements during vibrato will be characterized in order to check if their parameters remain constant during vibrato or, on the contrary, if they vary.

Moreover, it is important to note that because of the limitation of the inverse filtering techniques for high fundamental frequencies, the evaluation procedure will be applied in recordings corresponding with low fundamental frequencies. However, after this evaluation process the AM-FM representation of natural vibrato recordings will be compared to the one corresponding to the model, in order to analyze what this model predicts for higher fundamental frequencies.

A. Evaluation procedure

In Sec. III B the analysis procedure associated to the noninteractive source-filter model was shortly reviewed, and some relevant notes were provided for a correct interpretation of the obtained results. In this section, this analysis procedure will be used for the evaluation procedure of the noninteractive vibrato production, and, taking into account the above-mentioned details, some decisions have been made. Concerning the inverse filtering techniques, three representative approaches have been selected (the three were addressed in Sec. III B) for not limiting ourselves to a unique, and maybe biased, calculation. Additionally, as has been highlighted before, there are different factors that affect the inverse filtering results (differences among the selected techniques, fundamental frequency dependence,…), which has led us to apply the same analysis procedure to the synthetic signals, in order to determine the accuracy of the estimation procedure. Finally, regarding the VTR and GS parametrization, the normalized amplitude quotient (NAQ) has been chosen as a representative GS parameter, since its measurement is simple and robust and condenses in a single numerical value the whole glottal pulse information.[39] In parallel, and as the VTR representative parameters, the central frequencies of the first two formants, F1 and F2, have been selected.

This analysis procedure has been applied to the representative data set specified in Table I, as well as to the synthetic signals of Sec. IV, and the results corresponding to the vowel [a] are shown in Fig. 12. It is important to note that, in order to make easier the comparison among different singers, vocal tract configurations, and synthetic and natural signals, the absolute values of NAQ, F1, and F2 have been replaced by their relative values in percentage. To that purpose their mean value has been removed and they have been normalized by their average value in the considered interval.

In Fig. 12, different magnitudes for the four representative signals are shown: the acoustic signal, the RSI, the fundamental frequency variation, and the relative values of the NAQ, F1, and F2 and the plots corresponding to the different signals have been grouped in different columns, (a), (b), (c), and (d).

In all cases it can be seen that the considered time interval is long in the sense that it contains several vibrato cycles (at least four), and in all cases the RSI does not change significantly (perfectly constant in the synthetic signal). By observing the NAQ, F1, and F2 variations, it can be seen that they are not perfectly constant in all cases, either natural or synthetic signals. Additionally, by comparing the results corresponding to the three natural vibrato signals and the synthetic one, it can be seen that the maximum value of the NAQ relative value is very similar, around 10% for all signals. Something similar can be said concerning the F1 and F2, where the maximum value is around 5%. By considering that the variations appreciated in Fig. 12 for the NAQ, F1, and F2 have very similar values, it can be concluded that these variations can be mainly attributed to measurement errors.

In order to analyze other vocal tract configurations, the measurements of the NAQ, F1, and F2 obtained along the considered time interval have been reduced to a single numerical value: their standard deviation along the considered time interval. To this purpose, the results corresponding to the representative data set of Sec. IV are shown in Table II for the three singers and the noninteractive vibrato production model.

By analyzing Table II, it can be seen that the standard deviation of the normalized parameters is quite similar in all cases, and less than 10%, regardless of the conditions (different singers and proposed model, vocal tract configurations, different inverse filtering techniques). Additionally, by considering the three different vocal tract configurations for the three male singers, it can be seen that the standard deviation of the normalized NAQ is higher in the case of vowel [i], which corresponds to the lowest first formant central frequency, as can be seen in Fig. 11. This difference might be a consequence of the source-tract interaction, since it increases as the first formant central frequency decreases,[40] and it will be reflected on the inverse filtered GSD waveform. Also, it can be noticed that the standard deviation of the three parameters obtained for the baritone recordings are lower than those obtained for the tenor's recordings, which corresponds to differences in the intonation value, as is the case of the synthetic signals.

To summarize, and according to the results shown in Fig. 12 and Table II, it can be concluded that the GS and VTR parameters do not significantly change during vibrato, which validates the assumptions made when we proposed the

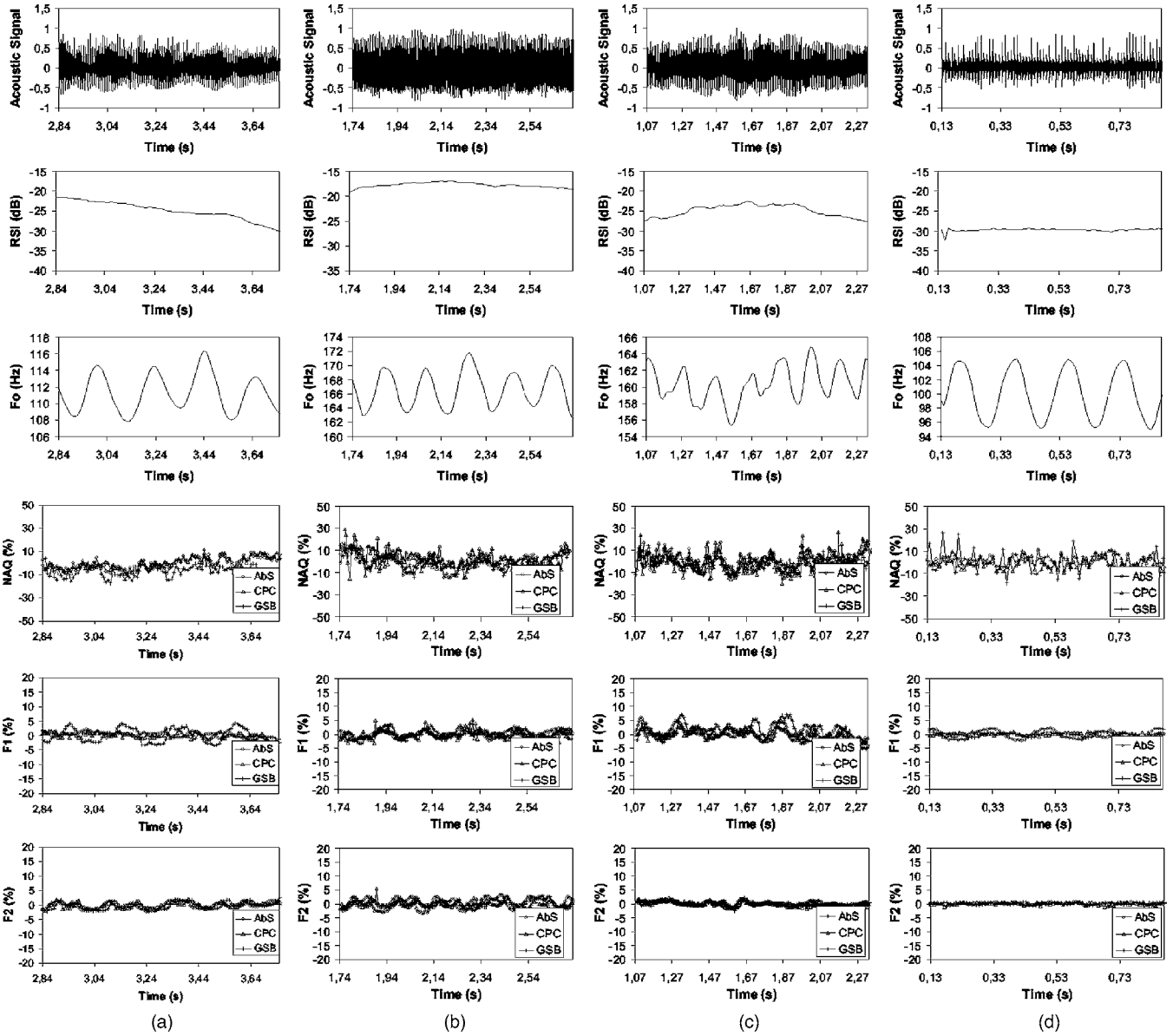


FIG. 12. Model evaluation. The signals associated to the model evaluation process are shown for four representative signals, corresponding to the three singers and the synthetic signal, all of them corresponding to the vowel [a]. For each one of these recordings the represented signals are the acoustic signal, the relative sound intensity, the fundamental frequency, the normalized NAQ, F1, and F2. The plots corresponding to the same voice are grouped by columns: (a) baritone, (b) tenor 1, (c) tenor 2, and (d) synthetic signal.

vibrato production model. This allows us to state that the noninteractive vibrato production model closely represents natural vibrato production.

B. Vibrato in higher pitched signals

As has been mentioned, in Sec. V A an evaluation procedure has been applied in order to validate the assumptions made along the model proposition. This evaluation procedure has been applied in low F_0 recordings such that the possible inverse filtering errors are avoided. However, it would be important to see how the same model performs under different conditions, and in particular for higher F_0 signals. This will allow us to determine if the above extracted conclusions concerning the GS and VTR behavior also hold under different pitch conditions.

Therefore, in this subsection, and in order to compare the noninteractive vibrato production model and natural vibrato signals, the AM-FM representation of the signal will be considered, since the analysis associated to the sinusoidal model has no limitation with the fundamental frequency value.

In this case a representative set of recordings for the vowel [a] has been selected, corresponding to three different F_0 values for the three male singers, as well as for the noninteractive vibrato production model: 166, 216, and 360 Hz, representing low, medium, and high fundamental frequency values (for male voices). In the case of synthetic signals, the VTR has been the same in all cases, corresponding to the vowel [a], and only the fundamental frequency of the signal has been varied, the vibrato rate being 5.5 Hz and the extent 84.46 cents.

TABLE II. Standard deviation of the normalized voice production parameters NAQ, F1, and F2 for the recordings corresponding to the baritone, tenor 1, tenor 2, and the noninteractive vibrato production model. The results correspond to different vocal tract configurations and different inverse filtering techniques results.

		Baritone			Tenor 1		
		sd_{NAQ}	sd_{F1}	sd_{F2}	sd_{NAQ}	sd_{F1}	sd_{F2}
	AbS	4.88	1.17	1.11	6.62	1.42	2.06
[a]	CPC	4.71	1.14	0.97	7.72	1.42	0.90
	GSB	5.65	2.28	1.30	5.84	1.24	3.14
	Abs	4.77	2.73	1.05	6.10	3.20	1.63
[e]	CPC	6.68	1.90	0.99	9.87	2.90	0.94
	GSB	5.02	3.32	1.54	6.39	2.59	1.60
	AbS	8.71	1.13	0.50	6.66	1.37	1.23
[i]	CPC	7.85	1.20	0.39	7.50	1.84	2.07
	GSB	9.32	1.57	0.60	9.20	2.18	1.36
		Tenor 2			Model		
	AbS	7.23	3.06	0.67	3.51	1.43	0.35
[a]	CPC	8.80	1.49	0.79	5.64	0.50	0.47
	GSB	6.12	2.33	0.87	4.37	1.69	0.59
	AbS	5.16	6.57	3.53	3.78	0.92	0.30
[e]	CPC	5.14	7.39	3.41	3.47	0.12	0.07
	GSB	7.38	5.32	3.56	6.41	0.87	0.31
	AbS	9.08	1.23	1.81	3.83	1.51	0.39
[i]	CPC	8.88	3.86	1.23	2.15	0.71	0.09
	GSB	9.79	2.56	1.35	3.76	1.90	0.55

The AM-FM representation of the above-mentioned signals are collected in Fig. 13. In this figure rows represent different singers (two tenors and one baritone) and the synthetic signals, and columns correspond to different fundamental frequency values.

In this case, as was detailed in Sec. II C, the AM-FM representation provides global information, since both voice production elements are included. However, the proposed vibrato production model allows us to understand how this representation is related to the voice production process. The noninteractive vibrato production model predicts that as the fundamental frequency increases, the harmonics of the GSD signal will be located in higher frequencies, and they will be located in different relative positions with respect to the vocal tract resonances. As a result, the AM-FM representation for each partial will correspond to a different local section of the VTR, shifted in amplitude depending on the spectral content of the GSD.

With this in mind, it can be observed in Fig. 13, for both natural and synthetic signals, that as the fundamental frequency increases, the AM-FM representations of the partials are obviously more separated from each other. Additionally, the spectral envelope foreseen in all cases is quite similar, since all correspond to the same vocal tract configuration, vowel [a], and all are of male singing voices. By considering one row of figures, for instance, (a), (b), and (c), it can be seen that, for a given singer, the only difference among the different F_0 values is reflected on the harmonic position, and the location of the AM-FM representation of each partial, which is predicted by the proposed model [panels (j), (k), and (l)]. This points out that there are no additional changes, concerning vibrato production, as the fundamental frequency increases. This behavior can also be observed in the results corresponding to other singers (different rows of figures) and, as a consequence, the conclusions can be extended to other male voices.

To conclude, it can be said that the noninteractive vibrato production model can be used to describe the behavior of the GS and VTR during the vocal vibrato production, at least from the signal point of view. It is clear that for a physiological description of the process other kind of models should be considered.

VI. CONCLUSIONS

In this work an acoustical characterization of vocal vibrato has been carried out. Such characterization has been based on the sinusoidal model parameters, IF and IA of the partials, focused on the amplitude variations and their relationship with the frequency variations. It has been shown that these variations are correlated, which is illustrated by the so called AM-FM representation of the partials. Additionally, we have shown that the amplitude and frequency variation of the partials will be translated into intensity variations depending on the window length imposed for the estimation. In particular, it has been shown that the RSI will not show pitch-related variations when the averaging window length is adjusted to one vibrato period. The main conclusion derived from this acoustical analysis is that the origin of the amplitude variations must be pursued taking into consideration the voice production mechanisms, since the sinusoidal model is a pure signal model.

In order to identify the contribution of the voice production elements, GS and VTR, on the amplitude variation of the partials, a noninteractive vibrato production model has been proposed. Consequently, this model has described the behavior of both voice production features during vibrato. In particular, considering a RSI constant interval, this model assumed that both voice production mechanisms remain almost constant during vibrato, and only the fundamental frequency of the GS changes. As a result, this model predicted that the AM-FM representation of each harmonic corresponds to a local section of the VTR but shifted (i.e., linearly distorted) in amplitude according to the GS spectrum. By comparing both synthetic and natural singing voice signals it was concluded that the above-mentioned prediction holds quite well in natural vocal vibrato.

Next, in order to validate the assumptions that the noninteractive vibrato production model is based on, the time evolution of the GSD and VTR features was evaluated in natural and synthetic signals making use of inverse filtering techniques. From this evaluation process, it was demonstrated that the GSD and VTR do not significantly vary during vibrato, which agrees well with the assumptions made during the model proposition. Additionally, the proposed model was compared to natural vibrato signals for different pitch values, making use of the AM-FM representation, and it was shown that the proposed model predicted similar representations to those obtained for natural vocal vibrato. This

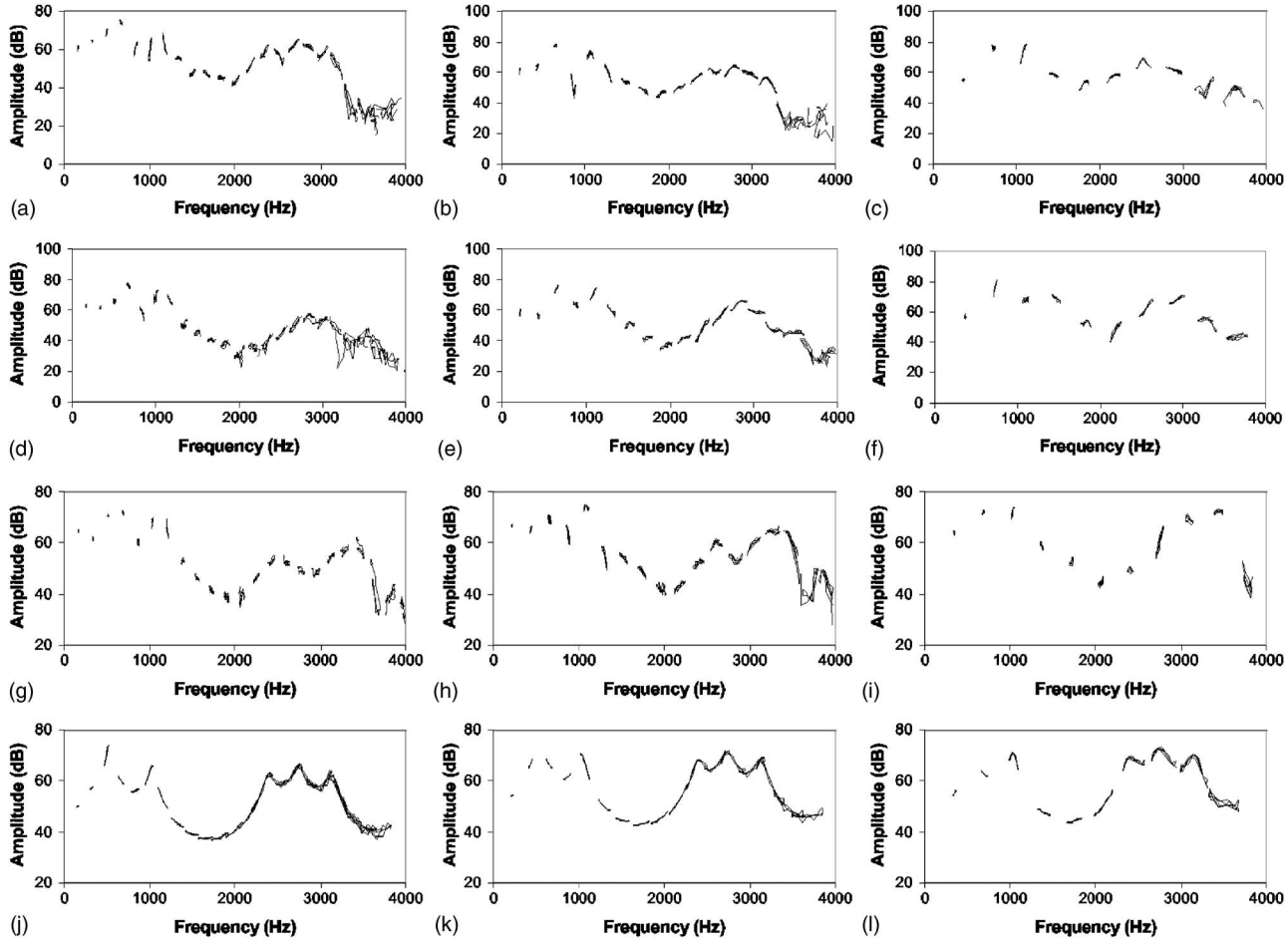


FIG. 13. Vibrato production in higher pitched signals. The proposed noninteractive vibrato production model is compared to natural vocal vibrato for different pitch values and different singers. The selected vocal tract configuration corresponds to vowel [a]. The results corresponding to the same voice are grouped by rows and each column correspond to a similar fundamental frequency value. Baritone recordings: (a) F_0=166 Hz, (b) F_0=216 Hz, and (c) F_0=350 Hz. Tenor 1 recordings: (d) F_0=166 Hz, (e) F_0=216 Hz, and (f) F_0=350 Hz. Tenor 2 recordings: (g) F_0=166 Hz, (h) F_0=216 Hz, and (i) F_0=350 Hz. Synthetic signal: (j) F_0=166 Hz, (k) F_0=216 Hz, and (l) F_0=350 Hz.

validation process has shown that the proposed vibrato production model is able to describe the production of this vocal effect.

Based on the noninteractive vibrato production model, it can be concluded that, during the vibrato production, the pitch variations are generated on the glottal source and these variations, along with the vocal tract filtering effect, reflect the frequency and amplitude variations of the acoustic signal partials.

ACKNOWLEDGMENTS

This work was supported in part by the Ministerio de Educación y Ciencia under Grant FPU, AP2000-4674. The Gobierno de Navarra and the Universidad Pública de Navarra are gratefully acknowledged for financial support. Finally, authors would also like to acknowledge support from Xavier Rodet and Axel Röbel (IRCAM, Paris) and the collaboration from Daniel Erro who implemented some of the algorithms.

[1]J. Sundberg, "Acoustic and psychoacoustics aspects of vocal vibrato," in *Vibrato*, edited by P. H. Dejonckere, M. Hirano, and J. Sundberg (Singular, San Diego, 1995), pp. 35–62.
[2]C. Sheashore, "The vibrato," in *University of Iowa Studies in the Psychology of Music*, Vol. 1 (University Press, Iowa City, 1932).
[3]E. Prame, "Measurements of the vibrato rate of ten singers," J. Acoust. Soc. Am. **96**, 1979–1984 (1994).
[4]E. Prame, "Vibrato extent and intonation in professional western lyric singing," J. Acoust. Soc. Am. **102**, 616–622 (1997).
[5]I. Arroabarren, M. Zivanovic, J. Bretos, A. Ezcurra, and A. Carlosena, "Measurement of vibrato in lyric singers," IEEE Trans. Instrum. Meas. **51**, 660–665 (2002).
[6]C. Dromey, N. Carter, and A. Hopkin, "Vibrato rate adjustment," J. Voice **17**, 168–178 (2003).
[7]J. A. Díaz and H. B. Rothman, "Acoustical comparison between samples of good and poor vibrato in singers," J. Voice **17**, 179–184 (2003).
[8]I. R. Titze, B. H. Story, M. Smith, and R. Long, "A reflex resonance model of vocal vibrato," J. Acoust. Soc. Am. **111**, 2272–2282 (2002).
[9]S. McAdams and X. Rodet, "The role of FM-induced AM in dynamic spectral profile analysis," in *Basic Issues in Hearing*, edited by H. Duifhuis, J. Horst, and H. Wit (Academic, London, 1988), pp. 359–369.
[10]Y. Horii, "Acoustic analysis of vocal vibrato: A theoretical interpretation of data," J. Voice **3**, 36–43 (1989).
[11]R. C. Maher and J. W. Beauchamp, "An investigation of vocal vibrato for synthesis," Appl. Acoust. **30**, 219–245 (1990).
[12]S. Imaizumi, H. Saida, Y. Shimura, and H. Hirose, "Harmonic analysis of the singing voice: Acoustic characteristics of vibrato," in Proc. of the Stockholm Music Acoustics Conf., 28 July–August 1 1993, pp. 197–200.
[13]M. Mellody, F. Herseth, and G. H. Wakefield, "Modal distribution analysis and synthesis of a soprano's sung vowels," J. Voice **15**, 469–482 (2001).
[14]I. Arroabarren, M. Zivanovic, X. Rodet, and A. Carlosena, "Instantaneous frequency and amplitude of vibrato in singing voice," in Proc. of the IEEE ICASSP, 6–10 April, 2003, Hong Kong, China.

[15]R. J. McAulay and Th. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Process. **34**(4), 744–754 (1986).
[16]X. Serra and J. Smith III, "Spectral Modeling Synthesis: A sound Analysis/Synthesis system based on deterministic plus stochastic decomposition," Comput. Music J. **14**, 12–24 (1990).
[17]P. R. Cook, "SPASM, a real-time vocal tract physical model controller; and Singer, the companion software synthesis system," Comput. Music J. **17**, 30–43 (1993).
[18]B. H. Story, "An overview of the physiology, physics and modeling of the sound source for vowels," Acoust. Sci. Technol. **23**(4), 195–206 (2002).
[19]G. Fant, *Acoustic Theory of Speech Production* (Mounton, The Hague, 1960).
[20]J. F. Michel and R. D. Myers, "Vibrato and pitch transitions," J. Voice **1**, 157–161 (1987).
[21]J. F. Michel and R. D. Myers, "The effects of crescendo on vocal vibrato," J. Voice **5**, 292–298 (1991).
[22]J. Sundberg, "Perceptual aspects of singing," J. Voice **8**, 106–122 (1994).
[23]J. O. Smith III and X. Serra, "PARSHL: An Analysis/Synthesis program for non-harmonic sounds based on a sinusoidal representation," in Proc. of the ICMC (1987).
[24]D. H. Klatt and L. C. Klatt, "Analysis, synthesis and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857 (1990).
[25]I. Arroabarren, M. Zivanovic, and A. Carlosena, "Analysis and synthesis of vibrato in lyric singers," in Proc. of the EUSIPCO, 3–6 September 2002, Toulouse, France.
[26]T. V. Ananthapadmanabha and G. Fant, "Calculation of the true glottal flow and its components," Speech Commun. **1**(3–4), 167–184 (1982).
[27]A. Barney, C. Shadle, and P. O. A. L. Davis, "Fluid flow in a dynamic mechanical model of the vocal folds and tract: I. Measurements and theory," J. Acoust. Soc. Am. **105**, 444–455 (1999).
[28]C. Shadle, A. Barney, and P. O. A. L. Davis, "Fluid flow in a dynamic mechanical model of the vocal folds and tract: II. Implications for speech production studies," J. Acoust. Soc. Am. **105**, 456–466 (1999).
[29]G. Fant, J. Liljencrants, and Q. Lin, "A four-parameter model of glottal flow," STL-QPSR **85**, 1–13 (1985).
[30]D. G. Childers and C.-F. Wong, "Measuring and modeling vocal source-tract interaction," IEEE Trans. Biomed. Eng. **41**, 663–671 (1994).
[31]H. Fujisaki and M. Ljungqvist, "Proposal and evaluation of models for the glottal source waveform," in Proc. of the IEEE ICASSP, April 1986, Vol. 11, pp. 1605–1608.
[32]H.-L. Lu and J. O. Smith, "Joint estimation of vocal tract filter and glottal source waveform via convex optimization," in Proc. of the IEEE WASPAA, October 1999, pp. 79–92.
[33]P. Alku and E. Vilkman, "Estimation of the glottal pulseform based on Discrete All-Pole modeling," in Proc. of the ICSLP, September 1994, Yokohama Japan, pp. 1619–1622.
[34]I. Arroabarren and A. Carlosena, "Glottal spectrum based inverse filtering," in Proc. of the EUROSPEECH, 1–4 September 2003, Geneva, Switzerland.
[35]D. Y. Wong, J. D. Markel, and A. H. Gray, "Least squares glottal inverse filtering from the acoustic speech waveform," IEEE Trans. Acoust., Speech, Signal Process. **27**, 350–355 (1979).
[36]N. Henrich, "Etude de la source glottique en voix parlée et chantée: modélisation et estimation, mesures acoustiques et électroglottographiques, perception," Ph.D. thesis, Paris 6 University, November 2001.
[37]H. Strik, "Automatic parametrization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses," J. Acoust. Soc. Am. **103**, 2659–2669 (1998).
[38]P. Alku and T. Bäckström, "Normalized amplitude quotient for parametrization of the glottal flow," J. Acoust. Soc. Am. **112**, 701–710 (2002).
[39]I. Arroabarren and A. Carlosena, "Glottal source parameterization: A comparative study," in Proc. of the VOQUAL, 27–29 August 2003, Geneva, Switzerland.
[40]B. Guerin, M. Mrayati, and R. Carre, "A voice source taking account of coupling with the supraglottal cavities," in Proc. of the IEEE ICASSP, April 1976, Vol. 1, pp. 47–50.

Bayesian analysis of polyphonic western tonal music

Manuel Davy[a)]
LAGIS/CNRS, Cité scientifique, BP 48, 59651 Villeneuve d'Ascq, France

Simon Godsill[b)]
Signal Processing Group, Department of Engineering, University of Cambridge Trumpington Street, Cambridge CB2 1PZ, United Kingdom

Jérôme Idier[c)]
IRCCyN/CNRS, 1 rue de la Noë, BP92101, 44321 Nantes cedex 3, France

(Received 30 August 2004; revised 4 January 2006; accepted 5 January 2006)

This paper deals with the computational analysis of musical audio from recorded audio waveforms. This general problem includes, as subtasks, music transcription, extraction of musical pitch, dynamics, timbre, instrument identity, and source separation. Analysis of real musical signals is a highly ill-posed task which is made complicated by the presence of transient sounds, background interference, or the complex structure of musical pitches in the time-frequency domain. This paper focuses on models and algorithms for computer transcription of multiple musical pitches in audio, elaborated from previous work by two of the authors. The audio data are supposedly presegmented into fixed pitch regimes such as individual chords. The models presented apply to pitched (tonal) music and are formulated via a Gabor representation of nonstationary signals. A Bayesian probabilistic structure is employed for representation of prior information about the parameters of the notes. This paper introduces a numerical Bayesian inference strategy for estimation of the pitches and other parameters of the waveform. The improved algorithm is much quicker and makes the approach feasible in realistic situations. Results are presented for estimation of a known number of notes present in randomly generated note clusters from a real musical instrument database.
 [DOI: 10.1121/1.2168548]

PACS number(s): 43.75.-z, 43.60.-c, 43.60.Pt, 43.60.Uv [ADP] Pages: 2498–2517

I. INTRODUCTION

Polyphonic music modeling is a challenging problem, which includes many possible subtasks, such as simultaneous multiple fundamental frequency estimation, time-varying amplitude/frequency tracking, modeling of inharmonicity, source separation, and inference about instrument-specific structures. Among the numerous approaches proposed in the literature, Bayesian approaches have been surprisingly rare, considering the large quantities of prior information available about musical signals. Musical signals are highly structured, both in the time domain and in the frequency domain (see, for example, Ref. 1 or 2). In the time domain, note transitions and percussive sounds occur at times related to the tempo and beat positions in the music. In the frequency domain, two levels of structure can be considered for harmonic notes. First, each note is composed of a fundamental partial, whose frequency is related to the "pitch" of the note, and overtone partials, whose relative amplitudes and frequencies determine the timbre of the note. This frequency domain description can be regarded as an empirical approximation to the true process, which is in reality a complex nonlinear time-domain system, as pointed out by McIntyre *et al.*[3] and Fletcher and Rossing.[2] Second, several notes played at the same time form chords or polyphony. Finally, higher levels of structure are present in sequences of chords, melodic shape, etc. This structural information can be incorporated into a probabilistic framework, so as to yield a statistical model, termed a *polyphonic Bayesian harmonic model* in this paper.

Numerous fundamental frequency estimation and analysis techniques can be found in the literature. Most apply only to monophonic (single voice) recordings and rely on nonparametric signal analysis tools (local autocorrelation function, spectrogram, etc.). Certain authors have adopted methods with statistical modeling elements, often using iterative procedures to estimate the individual components of a musical signal (see, for example, works by de Cheveigné *et al.*,[4,5] Virtanen and Klapuri,[6,7] Irizarry,[8,9] or Kameoka.[10]). Several existing Bayesian approaches include the work by Kashino *et al.*,[11–13] Raphael,[14,15] and Tabrikian *et al.*,[16] who adopt Bayesian hierarchical models for high level features in music such as chords, notes, timbre, etc., and Sterian *et al.*,[17] who adopt Bayesian tracking ideas for modeling of time-varying frequency partials. In a related vein, Parra and Jain[18] present Kalman filtering estimators for the harmonic plus noise model. A recent development is that of Cemgil *et al.*[19,20] who develop state-space models of musical harmonics and estimate them within an approximate Bayesian procedure.

The complete polyphonic transcription task is a great technical challenge. In this paper, we consider the subtask of multiple fundamental frequency estimation in segments of

a)Electronic mail: manuel.davy@ec-lille.fr
b)Electronic mail: sjg@eng.cam.ac.uk
c)Electronic mail: jerome.idier@irccyn.ec-nantes.fr

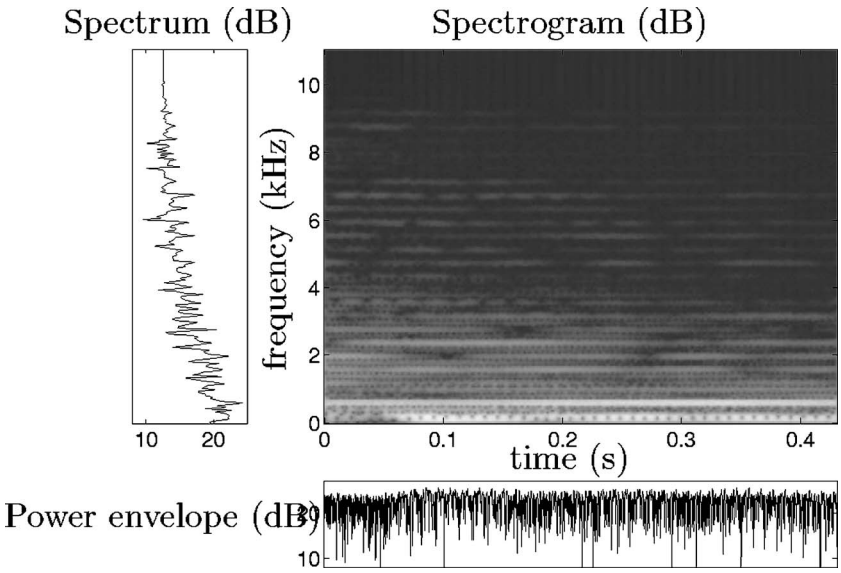


FIG. 1. Spectrogram, spectrum, and power envelope of a musical harmonic signal: a guitar alone plays two notes simultaneously. The spectrogram is computed with a Hamming window with length 11 ms. The harmonic structure, as well as the nonstationarity of partial amplitudes, appear clearly.

audio where it is assumed that no musical note changes occur. Such segments can be obtained efficiently using segmentation algorithms based on time-frequency representations,[21] support vector machines,[22,23] generalized likelihood ratio,[24] or other more music-specific methods.[25,26] We propose to describe these segments using a polyphonic music model where the number of partials, the time-varying amplitudes, the fundamental and overtone partial frequencies, and the noise variances are unknown. This new model is slightly modified compared to the model presented in our previous work (see Refs. 27–30). In particular, the noise is now assumed white (as opposed to colored in the previous Davy and Godsill papers), and the inharmonicity parameter takes a multiplicative form. These modifications were found to make parameter estimation more robust. The main contribution of this paper consists of the parameter estimation algorithm—here a Markov chain Monte Carlo (MCMC) method—which has been completely redesigned. Compared to the previous technique,[29,30] the computational load as been reduced dramatically: the computation time can be up to 10 000 times shorter. This enables us to present systematic tests involving various instruments arranged randomly into "chords," which was feasible only on a smaller scale with the previous approach.

The paper is organized as follows: in Sec. II we introduce the polyphonic Bayesian harmonic model—this involves modifications over our previous models both in terms of the inharmonicity models applied and in the prior structures adopted. In Sec. III, we describe the fast MCMC algorithm used to estimate the model parameters. In Sec. IV, the algorithm convergence is demonstrated, and results obtained with real instruments are reported. Results involving up to four instruments are presented, and a statistical analysis of estimation performance is proposed. Section V gives some conclusions and future work directions.

II. A BAYESIAN MODEL OF HARMONIC MUSIC

In this section, we present a flexible mathematical model of harmonic music. This model is then embedded in a probabilistic framework.

A. General model

Classically, acoustic signals produced by harmonic instruments are almost periodic and can be described as a sum of sinusoids.[2] Each individual note composing the signal is supposed to feature M partials, one of which is a *fundamental partial*, with angular frequency denoted ω_1, and the remaining $M-1$ are *overtone partials* with angular frequencies denoted ω_m, $m=2,\ldots,M$. Note that the fundamental frequency considered here does not coincide exactly with the common definition of fundamental frequency as the inverse of the period of the time domain signal. The overtone partial frequencies are approximately related to ω_1 by $\omega_m \approx m\omega_1$. Various instrument-specific models describe the relation between fundamental and overtone partial frequencies (see for example the piano model in Appendix A). In Fig. 1, the spectrogram, spectrum, and power envelope of a classical guitar record are plotted. The harmonic structure of guitar music appears clearly on the spectrum as well as on the spectrogram, where partials are the horizontal components with regular spacing. In the following, it is assumed that the music signal $y[t]$, $t=1,2,\ldots,N$, includes K notes having M_k partials with frequencies $\omega_{k,m}$ ($m=1,\ldots,M_k$ and $k=1,\ldots,K$).

Musical signals produced by harmonic instruments have the frequency structure described above; however, the partials amplitudes are generally nonconstant (see Fig. 1). An efficient modeling strategy is inspired by Gabor time-frequency representations[31] in which a signal $y[t]$, $t=1,2,\ldots,N$, is projected on a basis of Gabor atoms well localized in time and frequency. Our polyphonic harmonic model relies on a family of real-valued and nonzero phased Gabor atoms. Let $g[t]$ be one of these Gabor atoms with length $2N_g+1$, localized at time t_0 and frequency ω_0. It has two components:

$$g_c[t] = \phi[t]\cos(\omega_0 t) \text{ for } t = t_0 - N_g, \ldots, t_0 + N_g,$$
$$\text{and } g_c[t] = 0 \text{ otherwise}, \quad (1)$$

$$g_s[t] = \phi[t]\sin(\omega_0 t) \text{ for } t = t_0 - N_g, \ldots, t_0 + N_g,$$
$$\text{and } g_s[t] = 0 \text{ otherwise}, \tag{2}$$

where $\phi[t]$ is a parametrized envelope with, e.g., Gaussian shape, Hamming shape, etc. In standard discrete Gabor analysis, atom times $\{t_0\}$ and frequencies $\{\omega_0\}$ are positioned on a regular time-frequency lattice—see Ref. 32 for a fully Bayesian analysis in this setting; see also Ref. 33 for Gabor atom-based analysis of musical audio within a projection pursuit setting. Here, a regular time lattice is kept, but the atom frequencies correspond to the partial frequencies $\omega_{k,m}$, and these are allowed to vary continuously across a predefined range of frequencies. Let $\phi[t]$ be an atom envelope function, let I be a nonzero positive integer, and let $\Delta_t = (N-1)/I$. The Gabor atom-based model used in this paper is written (where ω_s is the sampling frequency)

$$y[t] = \sum_{k=1}^{K} \sum_{m=1}^{M_k} \sum_{i=0}^{I} \phi[t - i\Delta_t] \left\{ a_{k,m,i} \cos\left(\frac{\omega_{k,m}}{\omega_s} t \right) + b_{k,m,i} \sin\left(\frac{\omega_{k,m}}{\omega_s} t \right) \right\} + v[t]. \tag{3}$$

The atom lengths N_g and number of atom time locations $I+1$ have to be set to predefined values so as to capture important fluctuations in the musical signals, such as variations around note attacks, note decays, amplitude variations due to vibrato, and their variability between different instruments. The model presented in Eq. (3) includes relatively few parameters and is computationally tractable, while being sufficiently regularized that ambiguities between low-frequency partials and genuine amplitude modulations are unlikely to occur.

Another interpretation of Eq. (3) is as follows: each basis function $\phi[t - i\Delta_t]$, $i = 0, \ldots, I$, defines a frame, and a harmonic model is defined on each frame. In traditional signal processing approaches, the frames are processed almost independently. Here, all frames are processed together via the model in Eq. (3) which states that the frequencies are the same in each frame; this potentially makes the estimation of unknown parameters more accurate.

Matrix and vector notation for the above models will be used from now on, as follows. Let $\boldsymbol{\beta}$ denote the vector of amplitude parameters with length $2R(I+1)$, where $R = \sum_{k=1}^{K} M_k$ is the overall number of partials (or *model order*),

$$\beta_{2R_k(I+1)+2(M_k i+m)-1} = a_{k,m,i},$$
$$\beta_{2R_k(I+1)+2(M_k i+m)} = b_{k,m,i}, \tag{4}$$

for $i = 0, \ldots, I$, $m = 1, \ldots, M_k$, and $k = 1, \ldots, K$, with $R_k = \sum_{j=1}^{k-1} M_j$ (with the convention that $R_1 = 0$). Similarly, the matrix $\mathbf{D}$ of size $N \times 2R(I+1)$, whose columns are the Gabor atoms located at specified time instants $i\Delta_t$ and partial frequencies $\omega_{k,m}$, has entries defined by

$$\mathbf{D}_{t,2R_k(I+1)+2(M_k i+m)-1} = \phi(t - i\Delta_t)\cos\left(\frac{\omega_{k,m}}{\omega_s} t \right),$$
$$\mathbf{D}_{t,2R_k(I+1)+2(M_k i+m)} = \phi(t - i\Delta_t)\sin\left(\frac{\omega_{k,m}}{\omega_s} t \right), \tag{5}$$

for $i = 0, \ldots, I$, $m = 1, \ldots, M_k$, and $k = 1, \ldots, K$. Let $\mathbf{y} = [y[1], \ldots, y[N]]^t$ and $\mathbf{v} = [v[1], \ldots, v[N]]^t$ denote the vectors of signal samples and noise samples, respectively, where $\cdot^t$ denotes matrix transpose. Then Eq. (3) can be written more compactly as

$$\mathbf{y} = \mathbf{D}\boldsymbol{\beta} + \mathbf{v}. \tag{6}$$

The unknown parameters defining this model are the amplitudes $\boldsymbol{\beta}$, the number of partials for each note $\mathbf{M} = [M_1, \ldots, M_K]$, the partial frequencies $\boldsymbol{\omega} = [\omega_{1,1}, \ldots, \omega_{1,M_1}, \ldots, \omega_{K,1}, \ldots, \omega_{K,M_K}]$, and the variance of the noise vector $\mathbf{v}$, σ_v^2. The total number of unknown parameters is $R(2I+3)+2$. Considering a music signal with $N = 10\,000$ time samples, $K = 2$ notes, and $R = 50$ partials, the number of parameters is 1651 for $I = 15$ (this corresponds to atoms with length $2N_g + 1 = 30$ ms and 50% overlap).

B. Partial frequency models

In music produced by real instruments, inharmonicity will often be significant, i.e., the simple relation $\omega_{k,m} = m\omega_{k,1}$ (for $k = 1, \ldots, K$) is not satisfied. A well-known example is for struck or plucked strings[2] for which it is possible to develop a specific frequency model. However, given that we are here interested in analyzing generic harmonic instrumental music, in which the instruments playing may be unknown *a priori*, we adopt a more flexible form, which can adapt to less predictable inharmonicities found in real instruments:

$$\omega_{k,m} = m\omega_{k,1}(1 + \delta_{k,m}) \quad \text{where } \delta_{k,m} \ll 1 \text{ is to be estimated.} \tag{7}$$

The δ parameters are thus treated as unknown random variables within the proposed estimation scheme. Note that this new inharmonicity model is different from the one proposed in previous works.[29,30] The new model, being multiplicative rather than additive, allows more flexibility for high-frequency partials to deviate from their nominal "ideal" frequency $m\omega_{k,1}$. Note also that for a given k, the $\omega_{k,1}$ and $\delta_{k,m}$ terms uniquely specify the partial frequencies $\omega_{k,m}$. Thus we can work with either parametrization interchangeably, and we will usually refer to the $\omega_{k,m}$ terms directly in the following.

We now embed the polyphonic harmonic model defined in Eq. (3) in a probabilistic framework so as to allow efficient and accurate estimation of the model parameters, given an acoustic signal $\mathbf{y}$.

C. Bayesian harmonic music model

Assume the noise samples $\mathbf{v}$ in Eq. (6) are independent and identically distributed random variables, with zero-mean

Gaussian distribution (whose variance is denoted σ_v^2). Then the likelihood function of the model parameters is

$$p(\mathbf{y}|\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K) = (2\pi\sigma_v^2)^{-N/2}\exp\left[-\frac{1}{2\sigma_v^2}\|\mathbf{y}-\mathbf{D}\boldsymbol{\beta}\|^2\right]. \tag{8}$$

Remark: Here, the noise is assumed white Gaussian. It is also possible to implement a correlated Gaussian (or even non-Gaussian) noise model by appropriate modification of the likelihood Eq. (8). An example of the use of autoregressive coloured noise can be found in our previous work[29] (see also printed comments from discussants). Here we do not explore this option, as simulations showed that autoregressive colored noise could capture some of the partials, thus making computation complicated and degrading convergence of the algorithms.

Since the objective is to estimate the note parameters, one could simply try to find the maximum likelihood (ML) parameter set $(\hat{\boldsymbol{\beta}}_{\mathrm{ML}},\hat{\sigma}^2_{v\mathrm{ML}},\hat{\boldsymbol{\omega}}_{\mathrm{ML}},\hat{\mathbf{M}}_{\mathrm{ML}},\hat{K}_{\mathrm{ML}})$ that maximizes the likelihood function. The ML parameter estimate $(\hat{\boldsymbol{\beta}}_{\mathrm{ML}},\hat{\sigma}^2_{v\mathrm{ML}},\hat{\boldsymbol{\omega}}_{\mathrm{ML}},\hat{\mathbf{M}}_{\mathrm{ML}},\hat{K}_{\mathrm{ML}})$ minimizes the energy of the residual $\mathbf{r}_{\mathrm{ML}}=\mathbf{y}-\hat{\mathbf{D}}_{\mathrm{ML}}\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$, where $\hat{\mathbf{D}}_{\mathrm{ML}}$ is computed with the parameters $(\hat{\boldsymbol{\beta}}_{\mathrm{ML}},\hat{\sigma}^2_{v\mathrm{ML}},\hat{\boldsymbol{\omega}}_{\mathrm{ML}},\hat{\mathbf{M}}_{\mathrm{ML}},\hat{K}_{\mathrm{ML}})$ [see Eq. (8)]. This approach is, however, not satisfactory in practice as it tends to favor solutions with too many partials M_k and notes K (see Refs. 34 and 10). This is a well-known over-fitting problem of ML procedures which can be dealt with in principle using over-fitting penalization terms and methods such as AIC[35] and BIC[36] to determine an appropriate model order.

Here we choose to penalize over-fitting of the data, both by the parameters $\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega}$ and the overall model order R in a fully Bayesian framework. This is achieved by specifying *prior distributions* that encode our prior knowledge or belief about parameter values and their probable variability between different notes, instruments, recording setup, etc. A joint prior probability distribution $p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K)$ is specified, leading via Bayes' theorem to a *posterior distribution*

$$p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K) \times p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K). \tag{9}$$

A possible estimate of the model order parameters is then given by maximum *a posteriori* (MAP), as follows

$$(\hat{\mathbf{M}},\hat{K}) = \arg\max_{(\mathbf{M},K)} p(\mathbf{M},K|\mathbf{y})$$

where

$$p(\mathbf{M},K|\mathbf{y}) = \int p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K|\mathbf{y})d\boldsymbol{\beta}d\sigma_v^2d\boldsymbol{\omega}, \tag{10}$$

while the parameters for a particular model could be estimated for example by the minimum mean squared error (MMSE) estimator (written below) of the amplitudes parameter:

$$\hat{\boldsymbol{\beta}} = \int \boldsymbol{\beta}p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega}|\mathbf{y},\hat{\mathbf{M}},\hat{K})d\boldsymbol{\beta}d\sigma_v^2d\boldsymbol{\omega}. \tag{11}$$

In fact, this approach gives a somewhat simplified viewpoint, since the posterior expectations above may not yield an appropriate estimator of $\boldsymbol{\beta}$ in our case (the ordering of individual notes in $\boldsymbol{\beta}$, σ_v^2, $\boldsymbol{\omega}$, and $\mathbf{M}$ is nonunique, this is the *label switching problem*). Bayesian estimation in label switching contexts is quite difficult. Stephens[37] proposes a relabelling algorithm, and Jasra[38] proposes a review of possible relabelling methods. The technique employed here is simpler—further discussion is left until the MCMC algorithms are presented.

The posterior distribution and any Bayesian estimates made are influenced through Eq. (9) by both the likelihood function, which evaluates the model fit to the data $\mathbf{y}$, and the priors that encode the prior knowledge. The constant of proportionality in Eq. (9) will not be required as it depends only upon the (fixed) data $\mathbf{y}$—see Sec. III.

1. Prior distributions

A hierarchical structure is selected, which has the advantage of expressing $p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K)$, whose structure is complex, as a product of simpler elementary priors:

$$p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K) = p(\boldsymbol{\beta}|\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K)p(\boldsymbol{\omega}|\mathbf{M},K) \times p(\mathbf{M}|K)p(K)p(\sigma_v^2). \tag{12}$$

The individual prior terms above are now detailed. For some discussion of alternative priors, see our previous work.[29]

(*i*) The prior distribution $p(\boldsymbol{\beta}|\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K)$ for the $2R(I+1)$-dimensional amplitude parameter $\boldsymbol{\beta}$ is selected as zero-mean Gaussian, with covariance matrix $(\sigma_v^2/\xi^2)\mathbf{I}$, where $\mathbf{I}$ is an identity matrix and ξ^2 is a scaling parameter that can be interpreted as a signal-to-noise ratio, as pointed out by Andrieu and Doucet.[39] As explained in previous work,[40] selecting a diagonal covariance matrix permits the fast MCMC implementation presented below. Moreover, this diagonal covariance matrix does not degrade noticeably the overall algorithm performance, while enabling dramatic computational savings. Comparing the results presented in our previous work[29] to those presented below shows that the parameter estimation is accurate in both cases, while the computation time is significantly shorter with the new algorithm. The *hyperparameter* ξ^2 used to define the prior of $\boldsymbol{\beta}$ is treated as an additional unknown parameter in our framework, so that the full parameter set is actually augmented to $[\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\xi^2,\mathbf{M},K]$. In the following, we do not explicitly include ξ^2 in the unknown parameter set for the sake of notational simplicity. The prior distribution for ξ^2 is chosen as inverted gamma with small parameters, e.g., $\alpha_\xi=10^{-4}$ and $\beta_\xi=10^{-4}$:

$$p(\xi^2) = \mathcal{IG}(\alpha_\xi,\beta_\xi) \propto \frac{e^{-\beta_\xi/\xi^2}}{\xi^{2(\alpha_\xi+1)}}. \tag{13}$$

(*ii*) The prior distribution for frequencies $p(\boldsymbol{\omega}|\mathbf{M},K)$. In the general model of Eq. (7), the frequency prior structure is

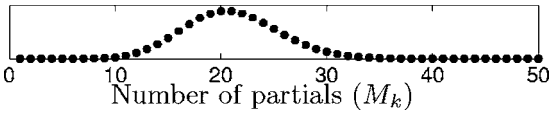


FIG. 2. Prior distribution of the number of partials $p(M_k)$ for any $k=1,\ldots,K$. The Λ_k hyperparameter is set to 20 in this case. This prior favors values of M_k between approximately 10 and 30.

$$p(\boldsymbol{\omega}|\mathbf{M},K)=\prod_{k=1}^{K}\left[p(\omega_{k,1}|M_k)\prod_{m=1}^{M_k}p(\delta_{k,m})\right], \tag{14}$$

where each $p(\delta_{k,m})$ is zero-mean Gaussian with variance σ_δ^2, and, for generality, $p(\omega_{k,1}|M_k)$ is uniform over the whole frequency range $[0,\omega_s/2M_k]$ for each note $k=1,\ldots,K$. Note that $p(\delta_{k,m})$ must be restricted to $[-1/2m,1/2m]$ so that the partial frequencies cannot switch. In practice, σ_δ^2 is a user-defined parameter set to the same very small value for each partial (we used $\sigma_\delta^2=3\times10^{-8}$ in simulations) and the restriction to $[-1/2m,1/2m]$ is numerically true without explicitly truncating $p(\delta_{k,m})$. The maximum possible frequency is $\omega_s/2M_k$ in order to avoid aliasing problems. Other more instrument- or genre-specific frequency priors could easily be envisaged (see Appendix A for a piano example).

(*iii*) The parameter $\mathbf{M}$ tunes the number of partials comprising the notes. Accurate estimation of M_k $(k=1,\ldots,K)$ is critical for correct frequency estimation. On the one hand, too many partials (M too large) generally result in underestimation of the fundamental frequency $\omega_{k,1}$: the estimated $\omega_{k,1}$ is generally one or two octaves below the correct frequency, and many partials with small amplitude appear in between the actual partials. On the other hand, not enough partials (M too small) does not capture enough information, and this may be a problem for signal reconstruction. It is thus crucial to define a prior which penalizes large numbers of partials, while allowing possibly many partials when needed. Following Andrieu and Doucet,[39] we have found that a balanced solution consists of selecting a *Poisson distribution* for each M_k,

$$p(M_k=m|\Lambda_k)=e^{-\Lambda_k}\frac{\Lambda_k^m}{m!}\quad\text{for } k=1,\ldots,K, \tag{15}$$

with $p(\mathbf{M}|K)=\Pi_{k=1}^{K}p(M_k|\Lambda_k)$, where Λ_k is left random with prior $p(\Lambda_k)$. The prior $p(\Lambda_k)$ is a gamma distribution with the same user-defined parameters α_Λ and β_Λ for each note,

$$p(\Lambda_k)=\mathcal{G}(\alpha_\Lambda,\beta_\Lambda)\propto\Lambda_k^{\alpha_\Lambda-1}e^{-\beta_\Lambda\Lambda_k}, \tag{16}$$

where the parameters are selected as $\alpha_\Lambda=1$ and $\beta_\Lambda=2$ to ensure this prior being vague, with infinite mean and variance.

For a given Λ_k, the shape of $p(M_k=m)$ is plotted in Fig. 2, with $\Lambda_k=20$. Again, we do not include it explicitly in the parameter set. In practice, M_k is limited to the range $[M_{\min},\ldots,M_{\max}]$, and $p(M_k=m)=(\Lambda_k^m/m!)/(\Sigma_{m'=M_{\min}}^{M_{\max}}\Lambda_k^{m'}/m'!)$.

This two-level prior is actually equivalent to the one-level prior

$$\begin{aligned}p(M_k=m)&=\int p(M_k=m|\Lambda_k)p(\Lambda_k)d\Lambda_k\\&\propto\int e^{-\Lambda_k}\frac{\Lambda_k^m}{m!}\Lambda_k^{\alpha_k-1}e^{-\beta_\Lambda\Lambda_k}\,d\Lambda_k\\&\propto\frac{\Gamma(m+\alpha_\Lambda)}{\Gamma(\alpha_\Lambda)m!}(\beta_\Lambda+1)^{-m},\end{aligned} \tag{17}$$

where $\Gamma(\cdot)$ is the Gamma function which coincides with the factorial operator $\Gamma(a+1)=a!$ for integer a. In practice, α_λ is set to one, yielding $p(M_k=m)=(\beta_\Lambda+1)^{-m}$, which is a monotonically decreasing function with decreasing rate set by β_Λ. This prior modeling enables a flexible framework: changing the parameters α_Λ and β_Λ defines various families of priors $p(M_k=m)$, with different decreasing profiles.

(*iv*) The number of notes K has a similar two-level prior as M_k, $k=1,\ldots,K$, with hyperparameter denoted Λ'.

(*v*) The noise amplitude parameter prior $p(\sigma_v^2)$ is selected as an inverted Gamma distribution

$$p(\sigma_v^2)\propto(\sigma_v^2)^{-\gamma_0/2-1}e^{-2/\nu_0\sigma_v^2}, \tag{18}$$

where ν_0 and γ_0 have small user-defined values, since this will favor solutions having small residual energy[41]

Using Eq. (9), one can compute the posterior distribution $p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K|\mathbf{y})$. Moreover, the integral

$$p(\boldsymbol{\omega},\mathbf{M},K|\mathbf{y})=\int p(\boldsymbol{\beta},\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K|\mathbf{y})d\boldsymbol{\beta}d\sigma_v^2 \tag{19}$$

can be calculated analytically in our case. In addition, $p(\sigma_v^2|\boldsymbol{\omega},\mathbf{M},K,\mathbf{y})$ and $p(\boldsymbol{\beta}|\sigma_v^2,\boldsymbol{\omega},\mathbf{M},K,\mathbf{y})$ can also be computed analytically. Standard computations lead to

$$\begin{aligned}p(\boldsymbol{\omega},\mathbf{M},K|\mathbf{y})\propto{}&(\gamma_0+\mathbf{y}^t\mathbf{P}\mathbf{y})^{(-N+\nu0)/2}\det(\mathbf{S})^{1/2}p(\boldsymbol{\omega}|\mathbf{M},K)\\&\times\exp\left(-\sum_{k=1}^{K}\Lambda_k\right)\frac{\Lambda'^K}{K!}\prod_{k=1}^{K}\frac{\Lambda_k^{M_k}}{M_k!},\end{aligned} \tag{20}$$

where $\mathbf{P}=\mathbf{I}-\mathbf{D}\mathbf{S}\mathbf{D}^t$ is an N-dimensional square matrix whose computation requires the inversion of the $2R(I+1)$-dimensional square matrix $\mathbf{S}=[\mathbf{D}^t\mathbf{D}+(1/\xi^2)\mathbf{I}]^{-1}$. The posterior distribution of σ_v^2, conditional on the other parameters, is an inverse gamma distribution with parameters $(N+\nu_0)/2$ and $(\gamma_0+\mathbf{y}^t\mathbf{P}\mathbf{y})/2$,

$$\begin{aligned}p(\sigma_v^2|\boldsymbol{\omega},\mathbf{M},\mathbf{y})&=\mathcal{IG}\left(\frac{N+\nu_0}{2},\ \frac{\gamma_0+\mathbf{y}^t\mathbf{P}\mathbf{y}}{2}\right)\\&\propto\exp\left(-\frac{\gamma_0+\mathbf{y}^t\mathbf{P}\mathbf{y}}{2\sigma_v^2}\right)\sigma_v^{-N-\nu_0-2},\end{aligned} \tag{21}$$

and posterior distribution of $\boldsymbol{\beta}$ conditional on the other parameters is a Gaussian distribution in dimension $2R(I+1)$,

$$p(\boldsymbol{\beta}|\sigma_v^2,\boldsymbol{\omega},\mathbf{M},\mathbf{y})=\mathcal{N}(\boldsymbol{\mu},\sigma_v^2\mathbf{S}), \tag{22}$$

where the mean is $\boldsymbol{\mu}=\mathbf{S}\mathbf{D}^t\mathbf{y}$ and the covariance matrix is $\sigma_v^2\mathbf{S}$. Efficient strategies to compute these quantities are presented in Ref. 40.

III. ALGORITHM FOR PARAMETER ESTIMATION

As stated in the text around Eq. (11), parameter estimation of, for example, the frequency parameter $\boldsymbol{\omega}$ requires the evaluation of integrals which cannot be calculated analytically. It is thus necessary to compute such integrals using numerical techniques. Grid-based and other simplistic approaches are not feasible owing to the high dimensionality of the problem and the high accuracy required in the frequency parameters in order to fit the data well.

The approach we adopt instead is a Monte Carlo integration. Assume we are able to generate random samples $(\tilde{\boldsymbol{\beta}}^{(l)}, \tilde{\sigma}_v^{2(l)}, \tilde{\boldsymbol{\omega}}^{(l)}, \tilde{\mathbf{M}}^{(l)}, \tilde{K}^{(l)})$, called Monte Carlo samples, distributed jointly according to $p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K|\mathbf{y})$ for some large number of samples $l=1,\ldots,L$. Then these samples can be used to form Monte Carlo estimates of any of the unknown parameters, as explained in Sec. III B. The actual generation of the Monte Carlo samples is performed using a Markov chain technique presented in Sec. III A.

A. Markov chain Monte Carlo (MCMC)

In order to estimate the unknown parameters, it is necessary to generate samples $(\tilde{\boldsymbol{\beta}}^{(l)}, \tilde{\sigma}_v^{2(l)}, \tilde{\boldsymbol{\omega}}^{(l)}, \tilde{\mathbf{M}}^{(l)}, \tilde{K}^{(l)})$ distributed according to $p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K|\mathbf{y})$ that will be used in the Monte Carlo estimation procedures described in Sec. III B. A general technique consists of generating a *Markov chain* using an iterative algorithm. Monte Carlo estimation together with Markov chain samplers forms the class of MCMC methods.

Generally speaking, Markov chain algorithms produce a series of samples whose distribution asymptotically approaches a given so-called *target distribution* $\pi(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ denotes a parameter vector to be estimated. In this paper, $\boldsymbol{\theta}$ is composed of $\boldsymbol{\beta}$, σ_v^2, $\boldsymbol{\omega}$, $\mathbf{M}$, K, and $\pi(\boldsymbol{\theta}) = p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K|\mathbf{y})$. Two important algorithms are the *Gibbs sampler* and the *Metropolis-Hastings sampler*.[42,43] The MCMC algorithm have been widely used in signal processing applications.[27,28,39,44–46] In this paper, we implement a specially constructed version of Metropolis-Hastings (MH) sampling, whose principle is described in Algorithm 1 for a vector $\boldsymbol{\theta}$ with a fixed number of dimensions N_θ and for the target distribution $\pi(\boldsymbol{\theta})$. Another presentation of the MCMC in the context of acoustics can be found in Ref. 47.

Algorithm 1 consists of sampling the components of $\boldsymbol{\theta}$ one at a time, according to a distribution $q(\boldsymbol{\theta})$. This distribution is called a *proposal*, because it is used to form the proposed candidate parameter value, denoted $\boldsymbol{\theta}^\star$. This candidate is randomly accepted or rejected according to the ratio (23). After some (large number of) iterations, the samples produced $\{\ldots, \tilde{\boldsymbol{\theta}}_j^{(l)}, \tilde{\boldsymbol{\theta}}_j^{(l+1)}, \tilde{\boldsymbol{\theta}}_j^{(l+2)}, \ldots\}$ are distributed according to the distribution $\pi(\boldsymbol{\theta})$, as required.

Algorithm 1: Generic one-at-a-time Metropolis-Hastings algorithm

1. Initialization.

- Set $l \leftarrow 1$

% Step 1.1 The parameter vector $\boldsymbol{\theta}$ is initialized

- For $j=1,\ldots,N_\theta$, do
  - Sample $\tilde{\boldsymbol{\theta}}^{(1)}$ according to some initial proposal distribution $q_1(\boldsymbol{\theta})$

2. While $l<L$, do

%Step 1.2 The parameter vector $\boldsymbol{\theta}$ is sampled one component at a time

- For $j=1,\ldots,N_\theta$, do (where j is the component index in $\boldsymbol{\theta}$)
  - Form the candidate by first setting $\boldsymbol{\theta}^\star \leftarrow \tilde{\boldsymbol{\theta}}^{(l-1)}$, then changing its component #j (denoted $\theta_j^\star$) by sampling it according to the proposal distribution $q(\theta_j|\tilde{\theta}_j^{(l-1)})$
  - Compute the MH ratio defined as follows

$$g(\boldsymbol{\theta}^\star, \tilde{\boldsymbol{\theta}}^{(l-1)}) = \frac{\pi(\boldsymbol{\theta}^\star)}{\pi(\tilde{\boldsymbol{\theta}}^{(l-1)})} \frac{q(\tilde{\theta}_j^{(l-1)}|\theta_j^\star)}{q(\theta_j^\star|\tilde{\theta}_j^{(l-1)})} \tag{23}$$

  - With probability $\alpha=\min[1, g(\boldsymbol{\theta}^\star, \tilde{\boldsymbol{\theta}}^{(l-1)})]$, accept the candidate $\boldsymbol{\theta}^\star$, i.e., set $\tilde{\boldsymbol{\theta}}^{(l)}=\boldsymbol{\theta}^\star$ or, with probability $1-\alpha$, reject the candidate, i.e., set $\tilde{\boldsymbol{\theta}}^{(l)}=\tilde{\boldsymbol{\theta}}^{(l-1)}$
- Set $l \leftarrow l+1$

A key element in MH algorithms is the *proposal distribution* q. This distribution must be simple enough to enable direct sampling, for example a Gaussian or uniform distribution. In principle, MCMC algorithms explore the space of possible values of $\boldsymbol{\theta}$ and focus on regions where the probability mass of $\pi(\boldsymbol{\theta})$ is large. By construction, it is able to switch from one region of the parameter space to another region where $\pi(\boldsymbol{\theta})$ has large probability mass. The computation of the accept/reject ratio involves the ratio $\pi(\boldsymbol{\theta}^\star)/\pi(\tilde{\boldsymbol{\theta}}^{(l-1)})$, which shows that π needs only be known up to a multiplicative constant, as any proportionality constant cancels out. Algorithm 1 is designed for variables $\boldsymbol{\theta}$ with fixed dimension N_θ. This is not the case in our problem since the dimension of $\boldsymbol{\theta}$ depends on $\mathbf{M}$ and K. Special extensions of MH algorithms, known as reversible jump MCMC, are designed to address cases where N_θ is unknown,[43,45,48] and we adopt these in this paper, following previous work.[27–30,39]

B. Computation of parameter estimates

We now assume that a Markov chain of samples has been generated and that only the samples after convergence are kept. They are denoted $(\tilde{\boldsymbol{\beta}}^{(l)}, \tilde{\sigma}_v^{2(l)}, \tilde{\boldsymbol{\omega}}^{(l)}, \tilde{\mathbf{M}}^{(l)}, \tilde{K}^{(l)})$ for $l=1,\ldots,L$. As discussed earlier, estimation of parameters using Bayesian methods can be carried out in many differing ways, such as maximizing the posterior distribution or using posterior mean parameter estimates. Here, however, none of these approaches is suitable. Owing to the label switching problem, there exist many sets of parameters which achieve the same posterior probability: they are all related through a different ordering of the individual notes in the parameter vectors $\boldsymbol{\beta}$, $\boldsymbol{\omega}$, and $\mathbf{M}$ (label switching problem). It is thus risky to compute averages over the Monte Carlo samples,

since we could average together fundamental frequencies of, e.g., the note with true $\omega_{k,1}=440$ Hz with the note with true $\omega_{k',1}=660$ Hz. Moreover, the posterior admits also local maxima corresponding to octave ambiguities, fifth ambiguities, etc., which would also be mixed in the posterior mean estimate. In order to limit this problem, we adopt the following scheme: First, the number of notes K is estimated by maximizing the marginal posterior distribution $p(K|\mathbf{y})$ (*marginal MAP estimation scheme*). This requires the integration of $\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}$ in the full posterior $p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K|\mathbf{y})$. This is computed very simply from the Monte Carlo samples by only considering the $\tilde{K}^{(l)}$ component and setting $\hat{K}$ as the most represented number of notes among the samples $\tilde{K}^{(l)}$. Second, $\widehat{\mathbf{M}}$ is estimated. In theory, the marginal MAP procedure cannot be used to estimate $\mathbf{M}$ because of the label switching problem. However, this solution is practical insofar as the Markov chains generated in simulations do not switch the labels whenever convergence to one of the maxima of $p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K|\mathbf{y})$ has occurred. In simulations, the chain converges to a posterior local maximum and keeps exploring around by adding or removing partials and changing $\boldsymbol{\omega}$ and $\boldsymbol{\beta}$, but without jumping to a completely different posterior maximum where the notes would be ordered differently (this happens if we run millions of iterations, but it does not improve the practical parameter estimation).

The Markov chain's practical difficulty in jumping from one local maximum to another enables, paradoxically, the best possible Bayesian estimates of $\boldsymbol{\beta}$, σ_v^2, $\boldsymbol{\omega}$, and $\mathbf{M}$: the discrete parameter $\widehat{\mathbf{M}}$ is computed by marginal MAP, which consists here of counting the number of Monte Carlo samples such that $\tilde{K}^{(l)}=\hat{K}$ and $\tilde{\mathbf{M}}^{(l)}=\mathbf{M}$ for each possible value of $\mathbf{M}$, and selecting the most represented value as the estimate $\widehat{\mathbf{M}}$. The continuous parameters $\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}$ are computed by marginal MMSE estimation, that is, by averaging over the Monte Carlo samples that satisfy jointly $\tilde{K}^{(l)}=\hat{K}$ and $\tilde{\mathbf{M}}^{(l)}=\widehat{\mathbf{M}}$. For example, the fundamental partial frequency of note k is estimated by

$$\hat{\omega}_{k,1} = \frac{1}{L'} \sum_{l'=1}^{L'} \omega_{k,1}^{(l')}, \tag{24}$$

where the notation l' and L' is used to emphasize the restriction to Monte Carlo samples that satisfy jointly the condition $\tilde{K}^{(l)}=\hat{K}$ and $\tilde{\mathbf{M}}^{(l)}=\widehat{\mathbf{M}}$. It is also possible to evaluate our uncertainty about the value of this parameter via the empirical variance estimate $(1/L')\Sigma_{l'=1}^{L'}[\omega_{k,1}^{(l')}-\hat{\omega}_{k,1}]^2$.

The above Monte Carlo parameter estimation technique is based on the law of large numbers which states that the expectation of a random variable, Eq. (11), can be approximated by the average of numerous samples of this random variable. Note that a suitably defined Monte Carlo integration may only require a few hundred samples $(\tilde{\boldsymbol{\beta}}^{(l)}, \tilde{\sigma}_v^{2(l)}, \tilde{\boldsymbol{\omega}}^{(l)}, \tilde{\mathbf{M}}^{(l)})$, as opposed to the huge number of grid points in Riemann integration for equivalent precision.[43]

C. A fast MCMC algorithm for polyphonic harmonic models

There are many possible options for creating a MCMC algorithm for a complex model such as this. The choices include which parameters to group together in each MH sampling step, which parameters to integrate out or marginalise (known as *Rao-Blackwellization*[43]), which proposal distributions to use, and how to design the most efficient implementations. Here we choose a scheme that has been optimized for computational speed while retaining good convergence properties.

The algorithm we implement in order to generate samples from the target distribution $p(\boldsymbol{\theta}, \mathbf{M}, K|\mathbf{y})$ is in essence a MH algorithm with several reversible jump steps, since the dimensions of $\boldsymbol{\theta}$, $\mathbf{M}$, and K vary together. The algorithm can be described hierarchically:

(i) At the highest level, we either sample the number of notes K or the parameters of the notes and we sample the noise variance σ_v^2 (see Algorithm 2).
(ii) At the level of notes, k is fixed and we want to sample the number of partials M_k (see Algorithm 5 below and Algorithms 7–10 in Appendix C).
(iii) At the level of individual partials, k and m are fixed and we want to sample the frequency $\omega_{k,m}$ and amplitudes $a_{k,m,i}$ and $b_{k,m,i}$, $i=0,\ldots,I$ (see Algorithm 6).

1. General algorithm structure

We now describe the algorithms from the note level to the frequency level. In these algorithms, every MH step has the same general structure as in Algorithm 1 and consists of (1) generating a candidate, (2) computing the MH ratio, and (3) performing the accept/reject test with respect to the target distribution. In order to simplify notation we omit the symbols $\tilde{\cdot}^{(l)}$ and $\cdot^{\star}$ wherever it is clear from the context that we are dealing with Markov chain samples and with candidates.

Algorithm 2: Overall Metropolis-Hastings algorithm for music

1. Initialization.

- Set $l \leftarrow 1$

% Step 2.1 Initialize the parameters $\boldsymbol{\omega}$, σ_v^2, $\boldsymbol{\beta}$, $\mathbf{M}$ and K

- Sample $\tilde{K}_1$ according to some initial distribution $q_{\text{init}}(K)$
- For $k=1,\ldots,\tilde{K}_1$ sample $\tilde{M}_k^{(1)}$ according to its prior distribution Eq. (15)
- Sample $\tilde{\omega}^{(1)}$ according to $q_{\text{init}}(\omega|\mathbf{y})$ where $q_{\text{init}}(\omega|\mathbf{y})$ is the probability distribution proportional to the Fourier spectrum of $\mathbf{y}$ (see Ref. 39 for a similar implementation).
- Sample the noise variance parameter $\tilde{\sigma}_v^{2(1)}$ according to $p(\sigma_v^2|\tilde{\boldsymbol{\omega}}^{(1)}, \tilde{\mathbf{M}}^{(1)}, \mathbf{y})$ given in Eq. (21)
- Sample the amplitudes $\tilde{\boldsymbol{\beta}}^{(1)}$ according to $p(\boldsymbol{\beta}|\tilde{\sigma}_v^{2(1)}, \tilde{\boldsymbol{\omega}}^{(1)}, \tilde{\mathbf{M}}^{(1)}\mathbf{y})$ given in Eq. (22)

2. While $l<L$, do

% Step 2.2 Sample the note parameters $\widetilde{K}^{(l)}$, $\widetilde{\mathbf{M}}^{(l)}$, $\widetilde{\boldsymbol{\omega}}^{(l)}$ and $\widetilde{\boldsymbol{\beta}}^{(l)}$

- with probability μ_K, try to add a note using the *birth move* (see Algorithm 3).
- Otherwise, with probability ν_K, try to remove a note using the *death move* (see Algorithm 4).
- Otherwise, with probability $1-\mu_K-\nu_K$, try the *update move* as follows
  - Set $\widetilde{K}^{(l)} \leftarrow \widetilde{K}^{(l-1)}$
  - For $k=1,\ldots,\widetilde{K}^{(l)}$, update the parameters of note #k (see Algorithm 5).
- For $k=1,\ldots,K$, update the parameters of note #k (Algorithm 5).

% Step 2.3 Sample the noise variance parameter σ_v^2 from $p(\sigma_v^2|\widetilde{\boldsymbol{\omega}}^{(l)},\widetilde{\mathbf{M}}^{(l)},\mathbf{y})$ given in Eq. (21)

- Set $l \leftarrow l+1$,

In Algorithm 2, $\widetilde{\mathbf{D}}^{(l)}$ denotes the matrix $\mathbf{D}$ computed with the parameters $\widetilde{\boldsymbol{\omega}}^{(l)}$, $\widetilde{\mathbf{M}}^{(l)}$, and $\widetilde{K}^{(l)}$ [see Eq. (5)]. The birth and death moves are aimed at adding/removing notes and include a reversible jump accept/reject test, similar in principle to the test in Algorithm 1: a candidate is generated, then it is tested with respect to the sample of iteration $l-1$ via a *reversible jump* MH accept/reject test. The probabilities to try the birth move (μ_K) or the death move (ν_K) are computed at each iteration based on the prior distribution of K, and the current value $\widetilde{K}^{(l-1)}$ (see Appendix B for details). The birth move is presented below, and the proposal distribution q_{note} is presented in Sec. III E.

Algorithm 3: Note birth

- Compute the residual $\mathbf{r}_0=\mathbf{y}-\widetilde{\boldsymbol{\beta}}^{(l-1)}\widetilde{\mathbf{D}}^{(l-1)}$
- Generate a candidate by increasing K, namely $K^{\bigstar} \leftarrow \widetilde{K}^{(l-1)}+1$
- Sample $M_{K+1}^{\bigstar}$ using the prior distribution $p(M_k)$ and set $\mathbf{M}^{\bigstar} \leftarrow [\widetilde{\mathbf{M}}^{(l-1)}, M_{K+1}^{\bigstar}]$
- Sample $\boldsymbol{\omega}_{K+1}^{\bigstar}=[\omega_{K+1,1}^{\bigstar},\ldots,\omega_{K+1,M_{K+1}}^{\bigstar}]$ using the proposal distribution $q_{\text{note}}(\boldsymbol{\omega}|\mathbf{r}_0,M_{K+1})$ and set $\boldsymbol{\omega}^{\bigstar} \leftarrow [\widetilde{\boldsymbol{\omega}}^{(l-1)}, \boldsymbol{\omega}_{K+1}^{\bigstar}]$
- Let $\mathbf{P}^{\bigstar}$ and $\mathbf{S}^{\bigstar}$ be the $\mathbf{P}$ and $\mathbf{S}$ matrices related to candidate note featuring $M_{K+1}^{\bigstar}$ partials with frequencies $\boldsymbol{\omega}_{K+1}^{\bigstar}$. Compute

$$g_{\text{birth}}=\left[\frac{\gamma_0+\|\mathbf{r}_0\|^2}{\gamma_0+\mathbf{r}_0^{\text{t}}\mathbf{P}^{\bigstar}\mathbf{r}_0}\right]^{(N+\nu_0)/2}\frac{p(\boldsymbol{\omega}_{K+1}^{\bigstar},M_{K+1}^{\bigstar},K^{\bigstar})}{p(\widetilde{K}^{(l-1)})}\frac{\nu_{K+1}}{\mu_K}\times\frac{\det(\mathbf{S}^{\bigstar})^{1/2}}{K^{\bigstar}q_{\text{note}}(\boldsymbol{\omega}_{K+1}^{\bigstar}|\mathbf{r}_0,M_{K+1}^{\bigstar})} \quad (25)$$

- With probability $\min(1,g_{\text{birth}})$, accept the candidate: Set $\widetilde{K}^{(l)} \leftarrow K^{\bigstar}$, $\widetilde{\mathbf{M}}^{(l)} \leftarrow \mathbf{M}^{\bigstar}$, $\widetilde{\boldsymbol{\omega}}^{(l)} \leftarrow \boldsymbol{\omega}^{\bigstar}$ and sample $\widetilde{\boldsymbol{\beta}}_{K+1}^{(l)}$ according to $p(\boldsymbol{\beta}_{K+1}|\widetilde{\sigma}_v^{2(l-1)},\widetilde{\boldsymbol{\omega}}_{K+1}^{(l)},\widetilde{M}_{K+1}^{(l)},K=1,\widetilde{I}^{(l)},\mathbf{r}_0)$ given in Eq. (22).

The opposite move is the death move, as detailed in the following algorithm.

Algorithm 4: Note death

- Generate a candidate by decreasing K, namely $K^{\bigstar} \leftarrow \widetilde{K}^{(l-1)}-1$,
- Sample a note index j in $\{1,\ldots,\widetilde{K}^{(l-1)}\}$ with equal probabilities
- Remove the parameters of note #j from $\widetilde{\mathbf{M}}^{(l-1)}$ and $\boldsymbol{\omega}^{(l-1)}$. This yields the candidates $\mathbf{M}^{\bigstar}$ and $\boldsymbol{\omega}^{\bigstar}$.
- Compute the residual $\mathbf{r}_j=\mathbf{y}-\boldsymbol{\beta}_{-j}\mathbf{D}_{-j}$ where $\boldsymbol{\beta}_{-j}$ (resp. $\mathbf{D}_{-j}$) is the vector of amplitudes $\widetilde{\boldsymbol{\beta}}^{(l-1)}$ (resp. the matrix of Gabor atoms $\widetilde{\mathbf{D}}^{(l-1)}$) computed with parameters $\widetilde{K}^{(l)}$, $\widetilde{\mathbf{M}}^{(l)}$, $\widetilde{\boldsymbol{\omega}}^{(l)}$ and $\widetilde{I}^{(l)}$ where the entries related to note #j have been removed.
- Let $\mathbf{P}_j$ and $\mathbf{S}_j$ be the $\mathbf{P}$ and $\mathbf{S}$ matrices related to note #j. Compute

$$g_{\text{death}}=\left[\frac{\gamma_0+\mathbf{r}_j^{\text{t}}\mathbf{P}_j\mathbf{r}_j}{\gamma_0+\|\mathbf{r}_j\|^2}\right]^{(N+\nu_0)/2}\frac{p(K^{\bigstar})}{p(\boldsymbol{\omega}_j,M_j,\widetilde{K}^{(l-1)})}\frac{\mu_{K-1}}{\nu_K}\widetilde{K}^{(l-1)}\times\frac{q_{\text{note}}(\boldsymbol{\omega}_j|\mathbf{r}_j,M_j)}{\det(\mathbf{S}_j)^{1/2}} \quad (26)$$

- With probability $\min(1,g_{\text{death}})$, accept the candidate: Set $\widetilde{K}^{(l)} \leftarrow K^{\bigstar}$, $\widetilde{\mathbf{M}}^{(l)} \leftarrow \mathbf{M}^{\bigstar}$, $\widetilde{\boldsymbol{\omega}}^{(l)} \leftarrow \boldsymbol{\omega}^{\bigstar}$ and remove the components corresponding to note #j from $\widetilde{\boldsymbol{\beta}}^{(l-1)}$ so as to obtain $\widetilde{\boldsymbol{\beta}}^{(l)}$

Algorithm 5, presented below, is aimed at sampling the parameters $\mathbf{M}$, $\boldsymbol{\omega}$, and $\boldsymbol{\beta}$ using a reversible jump MH scheme with $p(\boldsymbol{\beta},\boldsymbol{\omega},\mathbf{M}|\mathbf{y},\sigma_v^2,K)$ as target distribution. The steps involve proposing changes to the numbers of partials in each note, and also the frequencies in each note.

Algorithm 5: Metropolis-Hastings algorithm for note parameter updates

% Step 5.1 Compute the residual $\mathbf{r}_k$ that contains the note #k

- Compute the residual $\mathbf{r}_k=\mathbf{y}-\mathbf{D}_{-k}\boldsymbol{\beta}_{-k}$ where $\boldsymbol{\beta}_{-k}$ (resp. $\mathbf{D}_{-k}$) is the vector of amplitudes $\widetilde{\boldsymbol{\beta}}^{(l-1)}$ (resp. the matrix of Gabor atoms $\widetilde{\mathbf{D}}^{(l-1)}$) computed with parameters $\widetilde{\mathbf{M}}^{(l)}$ and $\widetilde{\boldsymbol{\omega}}^{(l)}$ where the entries related to note #k have been removed.

% Step 5.2 The number of partials M_k is sampled

- Sample a positive integer n uniformly in $[0,\ldots,\max(M_{\max}-\widetilde{M}_k^{(l-1)},\widetilde{M}_k^{(l-1)}-M_{\min})]$ and sample an integer e uniformly in $[-1,0,1]$
- With probability λb_{M_k}, try the *n-increase move*, see Algorithm 7 in Appendix C.
- Otherwise, with probability λd_{M_k}, try the *n-decrease move*, see Algorithm 8 in Appendix C.
- Otherwise, with probability $(1-\lambda)/2$, try the *divide move*, see Algorithm 9 in Appendix C.
- Otherwise, with probability $(1-\lambda)/2$, try the *multiply move*, see Algorithm 10 in Appendix C.
- Otherwise, with probability $\lambda(1-b_{M_k}-d_{M_k})$, try the *note update move* as follows
  - Set $\widetilde{M}_k^{(l)} \leftarrow \widetilde{M}_k^{(l-1)}$
  - Update the partial frequencies $\boldsymbol{\omega}_k$ (see Algorithm 6)

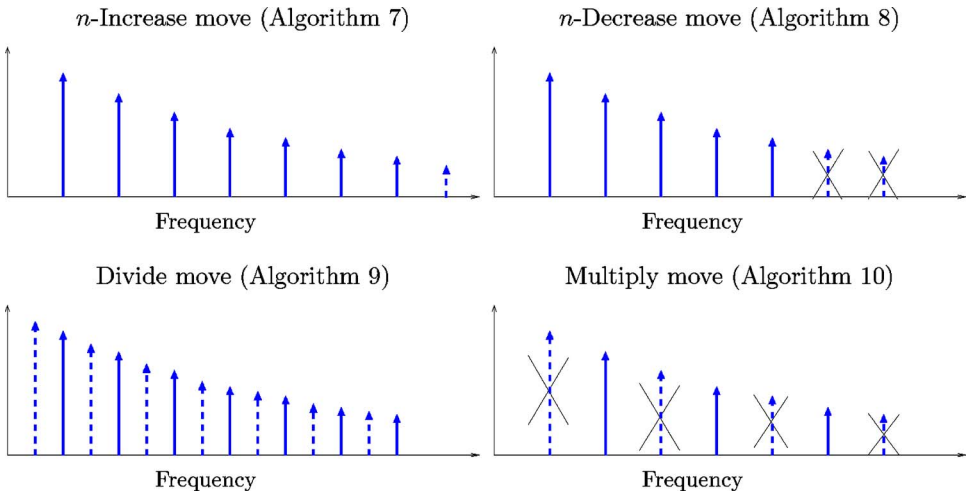


FIG. 3. (Color online) Schematic description of the four moves used to sample the number of partials M_k. From the initial situations, partials are either added (in dashed lines) or removed (in dashed lines, crossed). The partials plotted in solid lines are left unchanged.

In Algorithm 5, the probabilities b_{M_k} and d_{M_k} are computed in the same way as μ_K and ν_K (see Appendix B). The number of partials can be changed by two possible reversible moves as depicted in Fig. 3: either n partials are added (resp. removed) in the n-increase move (resp. n-decrease move) [see Algorithm 7 (resp. see Algorithm 8)] or the number of partials is multiplied by 2 (resp. divided by 2) in the divide move (resp. multiply move) [see Algorithm 10 (resp. see Algorithm 9)]. The multiply and divide moves are synchronized with fundamental frequency division/multiplication by two in order to avoid octave ambiguities: assume a note with frequency 440 Hz is played by an instrument. The notes with frequencies 220 and 880 Hz have partials that are almost superimposed with the partials of the actual note. Without these moves, the Markov chain may be stuck on incorrect fundamental frequencies, namely 220 or 880 Hz, which is not satisfactory. Note that the random parameter e is only aimed at ensuring reversibility of the divide and multiply moves. The parameter λ, selected by users in $[0;1]$, tunes the rate multiply-divide moves/increase-divide-update moves. A small λ is selected to avoid octave errors, because it favors multiply/divide moves. Algorithm 6, aimed at updating partial frequencies, is presented below.

Algorithm 6: Frequency update Metropolis-Hastings algorithm

- With probability λ_{note}, try the *global update*: sample $\boldsymbol{\omega}_k^{\star}$ according to $q_{\text{note}}(\boldsymbol{\omega}_k|\mathbf{r}_k,M_k)$ and perform an accept/reject test with respect to $p(\boldsymbol{\omega}_k,M_k,K=1|\mathbf{r}_k)$ given in Eq. (20) so as to obtain $\tilde{\boldsymbol{\omega}}_k^{(l)}$. Sample the amplitudes according to $p(\boldsymbol{\beta}_k|\tilde{\sigma}_v^{2(l-1)},\tilde{\boldsymbol{\omega}}_k^{(l)},\tilde{M}_k^{(l)},\mathbf{r}_k)$
- Otherwise, try the *local update* as follows: for $m=1,\ldots,M_k$,
  - Compute the residual $\mathbf{r}_{k,m}=\mathbf{r}_k-\mathbf{D}_{k,-m}\boldsymbol{\beta}_{k,-m}$ where $\boldsymbol{\beta}_{k,-m}$ (resp. $\mathbf{D}_{k,-m}$) is the vector of amplitudes $\boldsymbol{\beta}_{-k}$ (resp. the matrix of Gabor atoms $\mathbf{D}_{k,-m}$) where the entries related to partial (k,m) have been removed.
  - Sample $\boldsymbol{\omega}_{k,m}^{\star}$ according to $q_{\text{local}}(\boldsymbol{\omega}|\tilde{\boldsymbol{\omega}}_{k,m}^{(l-1)},\{m\})$
  - Perform an accept/reject test with respect to $p(\boldsymbol{\omega}_{k,m},\mathbf{M}=[1],K=1|\mathbf{r}_{k,m})$ given in Eq. (20) so as to obtain $\tilde{\omega}_{k,m}^{(l)}$
  - Resample the amplitudes $\tilde{\boldsymbol{\beta}}_{k,m}^{(l-1)}$ of partial k,m according to $p(\boldsymbol{\beta}_{k,m}|\tilde{\sigma}_v^{2(l-1)},\tilde{\boldsymbol{\omega}}_{k,m}^{(l)},\tilde{M}_k^{(l)},\mathbf{r}_{k,m})$ given in Eq. (22)

In Algorithm 6, two possible frequency updates are mixed: the global move consists of sampling the candidate $\boldsymbol{\omega}_{k,m}^{\star}$ in all the frequency range, whereas in the local move, the candidate $\boldsymbol{\omega}_{k,m}^{\star}$ is kept close to the previous value $\tilde{\boldsymbol{\omega}}_{k,m}^{(l-1)}$, and the partial frequencies are sampled one at a time. The global move probability λ_{note} is generally set to 0.25, and it tunes the rate of global/local moves so as to explore both the full space and the neighborhood. The proposal distributions q_{note} and q_{local} are presented in Sec. III E below.

D. Fast computations

In Algorithms 2–6, as well as in Algorithms 7–10 presented in Appendix C, we extensively use the following trick: each accept/reject test involves the computation of MH ratios of the form $p(\boldsymbol{\omega}^{\star},\mathbf{M}^{\star},K^{\star}|\mathbf{y})/p(\tilde{\boldsymbol{\omega}}^{(l-1)},\tilde{\mathbf{M}}^{(l-1)},\tilde{K}^{(l-1)}|\mathbf{y})$, which requires the costly inversion of two $2R(I+1)$-dimensional matrices of the form $\mathbf{D}^{\mathrm{t}}\mathbf{D}+(1/\xi^2)\mathbf{I}$ [see Eq. (20)]. However, we notice that only a fraction of the components in $\tilde{\boldsymbol{\omega}}^{(l-1)}$ are changed to form $\boldsymbol{\omega}^{\star}$. Similarly, only the component #k in $\tilde{\mathbf{M}}^{(l-1)}$ is modified to form $\mathbf{M}^{\star}$. More generally, assume we know the parameters that define completely some of the partials. In other words, we know $\omega_{k,m},M_k,a_{k,m,i},b_{k,m,i}$ for all $i=0,\ldots,I$ and a limited set of partial indexes denoted $\{k,m\}_a$. Let $\boldsymbol{\beta}_a$, $\boldsymbol{\omega}_a$, and $\mathbf{M}_a$ be the parameters corresponding to $\{k,m\}_a$ with related matrix $\mathbf{D}_a$. Similarly, we denote $\boldsymbol{\beta}_b$, $\boldsymbol{\omega}_b$, and $\mathbf{M}_b$ the remaining parameters with related matrix $\mathbf{D}_b$. By linearity, Eq. (6) can be written $\mathbf{y}=\mathbf{D}_a\boldsymbol{\beta}_a+\mathbf{D}_b\boldsymbol{\beta}_b+\mathbf{v}$. Introducing the residual $\mathbf{r}_a=\mathbf{y}-\mathbf{D}_a\boldsymbol{\beta}_a$, this is equivalent to $\mathbf{r}_a=\mathbf{D}_b\boldsymbol{\beta}_b+\mathbf{v}$ where the dimension of $\mathbf{D}_b$ and $\boldsymbol{\beta}_b$ is smaller than the original dimension of $\mathbf{D}$ and $\boldsymbol{\beta}$. Clearly, the posterior distribution $p(\boldsymbol{\omega}_b,\mathbf{M}_b|\mathbf{y},\boldsymbol{\beta}_a,\boldsymbol{\omega}_a,\mathbf{M}_a)$ is $p(\boldsymbol{\omega}_b,\mathbf{M}_b|\mathbf{r}_a)$ with exactly the same structure as $p(\boldsymbol{\omega},\mathbf{M}|\mathbf{y})$ in Eq. (20), with the major difference that the matrix $\mathbf{D}^{\mathrm{t}}\mathbf{D}$ to be inverted is now $\mathbf{D}_a^{\mathrm{t}}\mathbf{D}_a$, with smaller size. In practice, the indexes $\{k,m\}_a$ correspond to either one note #k (i.e., M_k partials) as in Algorithm 5, or

to a single partial as in Algorithm 6, thus the inversion of $\mathbf{D}_a^t\mathbf{D}_a$ is dramatically quicker than that of $\mathbf{D}^t\mathbf{D}$ (matrix inversion algorithms complexity scales as N_{mat}^3, where N_{mat} is the matrix dimension). More details about this trick can be found in Ref. 40, where several fast methods are also proposed to compute $\mathbf{r}_a^t\mathbf{P}_a\mathbf{r}_a$, $\det(\mathbf{S}_a)$, and $\boldsymbol{\mu}_a$ and to sample from $\mathcal{N}(\boldsymbol{\mu}_a, \sigma_v^2\mathbf{S}_a)$ based on the Choleski decomposition of $\mathbf{S}_a^{-1}$ (where the notation with subscript a is the obvious extension of the above notations).

In practice, the MH ratios $p(\boldsymbol{\omega}^\star, \mathbf{M}^\star, K^\star|\mathbf{y})/p(\tilde{\boldsymbol{\omega}}^{(l-1)}, \tilde{\mathbf{M}}^{(l-1)}, \tilde{K}^{(l-1)}|\mathbf{y})$ are replaced by similar ratios involving only the set of parameters to be changed, as written in Algorithms 2–10.

E. Proposal distributions

In Algorithms 2–6, the new frequencies are sampled randomly according to several proposal distributions. These distributions are crucial: a bad proposal distribution samples candidates distant from likely solutions which are rejected almost always, yielding an inefficient algorithm.

The local proposal distribution q_{local} consists of sampling the new frequency close to the last accepted one. For the model proposed in Eq. (6), q_{note} may have the following structure: the fundamental frequency $\omega_{k,1}$ is sampled according to $q_{\text{local}}(\omega_{k,1}|\tilde{\omega}_{k,m}^{(l-1)})$ chosen as the Gaussian distribution of mean $\tilde{\omega}_{k,1}^{(l-1)}$ and user defined variance $\sigma_{\text{RW},1}^2$. Then, for $m=1,\ldots,M_k$, the δ detuning parameters are updated one at a time and $q_{\text{local}}(\delta|\tilde{\delta}_{k,m}^{(l-1)}, m)$ is a Gaussian distribution of mean $\tilde{\delta}_{k,m}^{(l-1)}$ and user-defined variance $\sigma_{\text{RW},\delta}^2$. In practice, there is a different q_{local} for each partial frequency model. The distribution $q_{\text{partials}}(\omega|\mathbf{r}_k, \omega_{k,1}, \{\text{list of partials}\})$ samples the partial frequencies listed according to $q_{\text{local}}(\delta|\tilde{\delta}_{k,m}^{(l-1)}, m)$.

F. Convergence issues

The Metropolis-Hastings-type algorithm presented above is designed so as to generate a Markov chain with target distribution $p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K|\mathbf{y})$. It can be shown that this algorithm is built so as to produce an ergodic Markov chain, which ensures convergence whatever the initial sample. Moreover, it can be shown that it converges uniformly geometrically to the desired probability distribution:[49]

$$\|p^{(l)}(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K) - p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K|\mathbf{y})\|_{TV} \leqslant C\rho^l \quad (27)$$

where $p^{(l)}(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K)$ is the actual probability distribution of $\tilde{\boldsymbol{\beta}}^{(l)}$, $\tilde{\sigma}_v^{2(l)}$, $\tilde{\boldsymbol{\omega}}^{(l)}$, $\tilde{\mathbf{M}}^{(l)}$, $\tilde{K}^{(l)}$ and $\|\cdot\|_{TV}$ is the total variation norm; C and $\rho<1$ are constants. In practice, it would be necessary to run millions of iterations to have samples distributed according to the posterior over the full space (including label switching); however, simulations show that a few hundred iterations are enough to explore one of the global maxima. Empirical study of the simulation of similar algorithms can be found in Ref. 40.

All the elements have now been described to enable implementation of the proposed method. Results are presented in Sec. IV.

IV. RESULTS

In this section, we report results obtained by processing real music signals from the McGill Database. In order to illustrate the performance of the approach in monophonic and polyphonic contexts, we investigate the four cases $K=1$, $K=2$, $K=3$, and $K=4$. Signals in cases $K\geqslant 2$ are generated by randomly mixing monophonic signals (i.e., signals with $K=1$ note) from the database. In all experiments, we have extracted signals with duration of 544 ms from the original signals. Signals are downsampled from 44100 to 11 025 Hz, thus the processed music extracts have $N=6000$ samples. For the sake of clarity, signals are numbered as $K.xx$ in the following, where K is the number of notes and xx is the extract number. In all the experiments, one Markov chain with 800 iterations is generated using the algorithm presented above, and the model parameters are estimated as explained above. We would like to stress that **only one** simulation is run. We did not run several simulations and keep the best ones, as this would not provide performance results of practical interest. In these estimations, the 100 final samples in the chain are kept in order not to use the 700 first samples where the chain may not have converged and which may not be distributed according to $p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}|\mathbf{y})$. All results concerning pitch estimation are obtained by converting a frequency to a pitch number τ in semitones, relative to A 440 Hz, as follows,

$$\tau = 12\frac{\log(\omega/2\pi) - \log(440)}{\log(2)}. \quad (28)$$

In the main simulations, we used the partials frequencies model in Eq. (7) and the frequencies prior is as given in Eq. (14). In further comparative simulations, different simpler versions of the model are tested for comparative purposes with earlier work. The model parameters are $I=14$, $M_{\min}=2$, $M_{\max}=30$, and $\sigma_\delta^2=3\times10^{-8}$. The hyperparameter ξ^2, Λ_k and Λ' are incorporated into the set of unknown parameters to be estimated, and are sampled as in Refs. 39 and 50. The algorithm proposal parameters used were $\lambda=0.3$, $\lambda_{\text{note}}=0.25$, $\sigma_{\text{RW},1}^2=0.01/N$, and $\sigma_{\text{RW},\delta}^2=10^{-3}$. Moreover, the notes number K is initialized to $K=1$—thus choosing $q_{\text{init}}(K)=\delta_1(K)$—so as to let the number of notes increase until it reaches the actual number of notes.

On a dual processor Pentium IV (2.6 GHz) computer, and Matlab code, the average computation time is 1.35 s per iteration per note, rather a high load, but note that the segments of music being analyzed are fairly substantial in length (roughly 0.5 s computed in 1 h in total).

In the next subsection, we consider the monophonic case ($K=1$). In Sec. IV B, we address the polyphonic cases ($K=2$ to $K=4$). In both Secs. IV B and IV A, the number of notes is assumed known and is used in the algorithm by setting $K_{\max}$ to the true number of notes. This issue is further discussed in Sec. IV E.

A. Monophonic case: $K=1$

We first present simulations aimed at assessing the Markov chain convergence, then we focus on three typical simulations. Finally, we propose some statistics obtained by processing 20 different music signals.

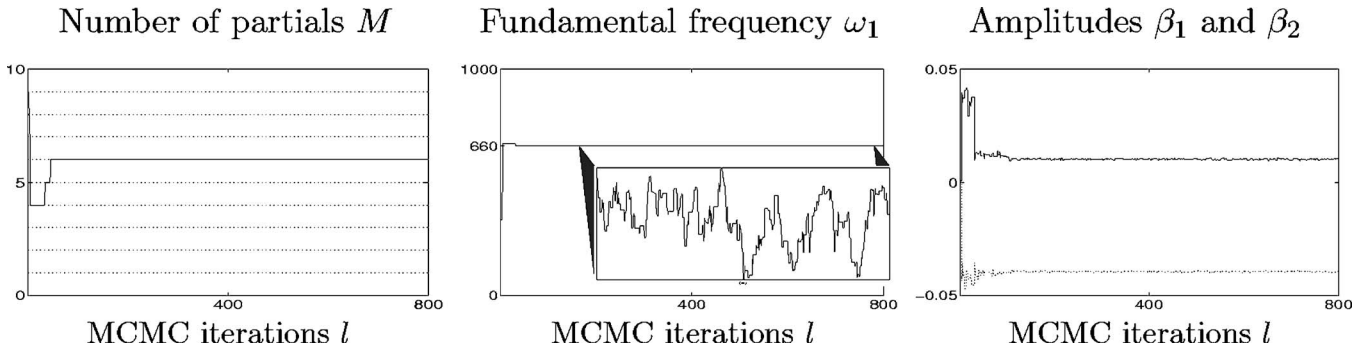


FIG. 4. Monophonic case, test signal 1.1. Convergence of the Markov chain generated by the algorithm in the monophonic case ($K=1$). Some components of the sampled random vector $\tilde{\boldsymbol{\beta}}^{(l)}, \tilde{\sigma}_v^{2(l)}, \tilde{\boldsymbol{\omega}}^{(l)}, \tilde{\mathbf{M}}^{(l)}$ are represented for $l=1,\ldots,1000$. As can be seen, convergence is reached within the 100 first iterations. Samples $l=701$ to $l=800$ are kept for estimation purposes. As illustrated in the central plot, though convergence is reached, the chain concerning ω_1 keeps fluctuating. The true fundamental frequency is 660 Hz ($\tau=7$).

1. Convergence of the sampling algorithm

In order to illustrate the convergence of the above MCMC algorithm, we have plotted the evolution of the random samples generated at each iteration $l=1,\ldots,800$ (see Fig. 4). As can be seen, the chain converges to a steady state after only 100 iterations, and the very final samples can be assumed to be distributed according to $p(\boldsymbol{\beta}, \sigma_v^2, \boldsymbol{\omega}, \mathbf{M}, K=1|\mathbf{y})$. In practice, we keep the final 100 samples (i.e., iterations $l=701$ to $l=800$) to estimate the parameter values (see Sec. III B). In Fig. 4, the magnified portion of $\tilde{\omega}_1^{(l)}$ shows that the parameter keeps fluctuating after convergence is reached: this is a normal behavior which is explained as follows. The Markov chain has converged *in terms of probability density function*; this means that the samples generated are random (thus, the chain fluctuates), but their distribution is fixed. All 20 tests we ran showed very similar behavior to this simulation. In other words, the Markov Chain plotted in Fig. 4 in representative of the 20 Markov chains generated for the 20 tests, whatever the instrument playing, and whatever the played note pitch.

2. Estimation results

The estimation performance is now analyzed. We focus on three test signals, namely test signals #1.1, #1.4, and #1.8. For each test signal, the reconstructed signal $\hat{\mathbf{y}}$ is computed using the model in Eq. (3) with $K=1$ and the estimated parameters $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\omega}}$, $\hat{\mathbf{M}}$ computed using the last 100 samples of the Markov chain. In Fig. 5, we have plotted the original signals $\mathbf{y}$ as well as $\hat{\mathbf{y}}$ and the residuals $\mathbf{y}-\hat{\mathbf{y}}$, in the time domain. The difference between $\mathbf{y}$ and $\hat{\mathbf{y}}$ is quite small in all three cases (i.e., the residual energy is small compared to the energy of $\mathbf{y}$). Figure 6 displays the spectra of $\mathbf{y}$, $\hat{\mathbf{y}}$ and residuals $\mathbf{y}-\hat{\mathbf{y}}$. The model in Eq. (3) has been able to capture all harmonic components except for some high-frequency partials. The amplitudes of these partials are, however, about 20 dB smaller than partials with smaller frequency. Moreover, the flute extract #1.4 displayed in Figs. 5 and 6 shows that our Gabor-based approach models accurately time-varying amplitudes with significant variations.

Table I reports pitch estimation results for the 20 test signals, in terms of which semitone in the scale is closest to the estimated pitch number. As can be seen, there is no pitch estimation error. Estimated pitch values may differ slightly from the exact integer values, because they are computed relative to A 440 Hz. Due to recording artifacts, instrument tuning, etc., the true A frequency may differ from 440 Hz.

B. Polyphonic cases: $K=2$, $K=3$, and $K=4$

In this subsection, we present results obtained in polyphonic contexts. As for monophonic tests, generated Markov chains converge quite quickly to the parameter posterior distribution. (For the sake of brevity, we do not present convergence examples again here.) Table II reports pitch estimation results for the 20 test signals. Similar to the monophonic case, most notes are accurately estimated. There are, however, some octave error cases. Octave errors happen when the estimated pitch $\hat{\tau}$ is one or more octaves (12 note numbers) away from the true pitch. Example errors can be found with extracts numbers #2.10, #2.12–#2.15, and #2.18. For extract number #2.16, the two notes have the same pitch ($\tau=-12$), so it is very difficult for the algorithm to separate them into two superimposed notes. The algorithm actually assigned the correct pitch $\hat{\tau}=-12$ to one note, and set $\hat{\tau}=7$ to the other note. The partials of the note with pitch $\hat{\tau}=7$ are superimposed with partials of the note with $\hat{\tau}=-12$. Beside these problems, no errors occurred.

Figure 7 displays estimation results in the frequency domain for the extracts #2.1 (both notes are correctly estimated), #2.14 (one note is correctly estimated, and the other shows an octave error), and #2.16 (the two notes have the same frequency). The time domain plots are very similar to those presented in Fig. 5 (monophonic case), thus they are not presented here. In these more difficult contexts, the majority of notes are still accurately estimated. There are more octave errors than with $K=2$, but octave errors do not increase much when related to the actual number of notes to be estimated (three or four notes, instead of two). In Fig. 8, we have plotted the spectra corresponding to extracts #3.1 and #4.2. Again, the model has captured most harmonic information, apart from some high-frequency partials. Aside from octave ambiguities, some new error cases occur: they are examined in details in Sec. IV C.

Cases with $K=3$ and $K=4$ notes exhibit the same kind of behavior. In summary, pitch is correctly estimated for 84.6%

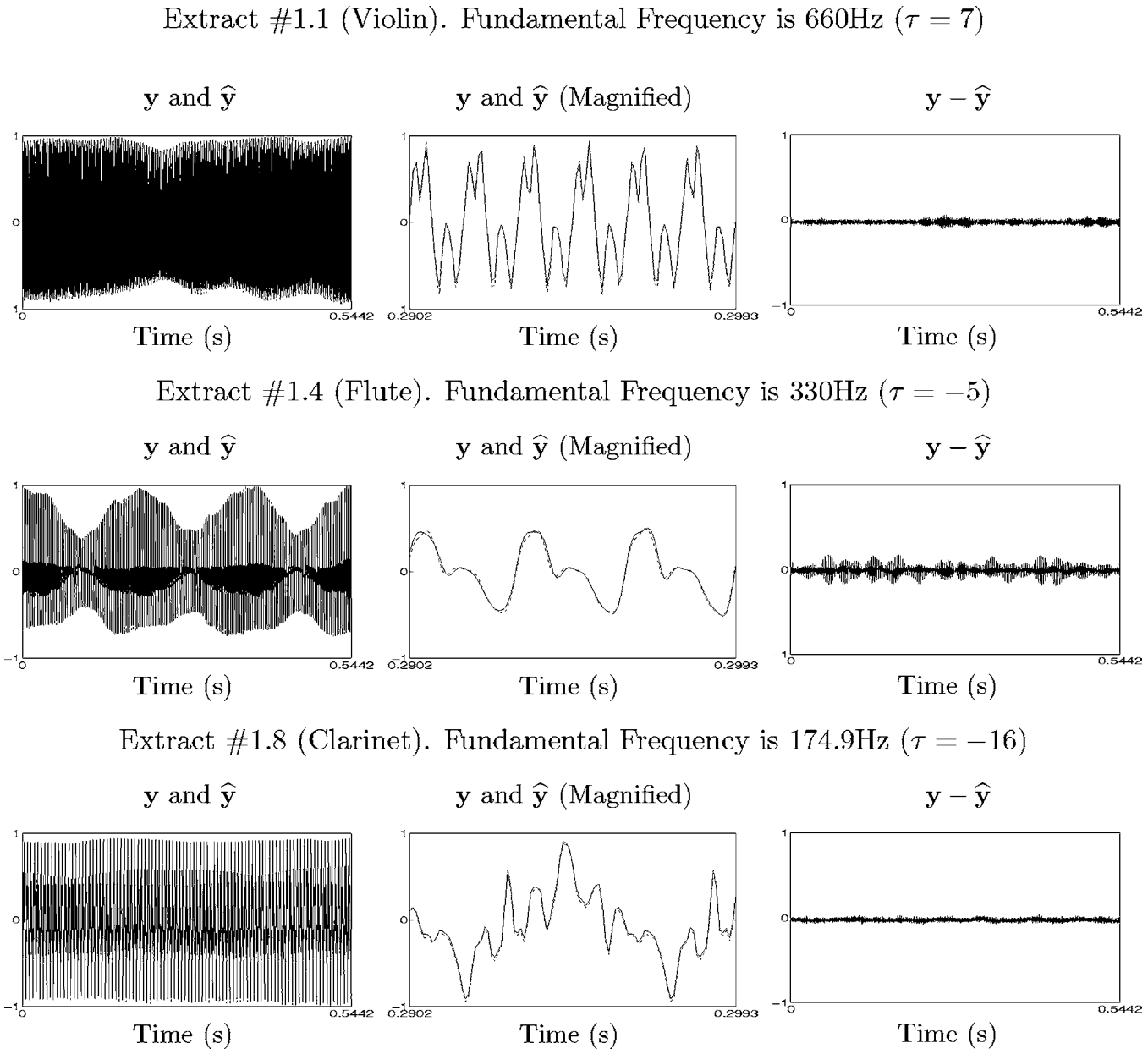


FIG. 5. Monophonic case, estimation results for three test signals. Time domain plots of original signals $\mathbf{y}$ (dashed lines), reconstructed signals $\hat{\mathbf{y}}$ (solid lines), and the residuals $\mathbf{y}-\hat{\mathbf{y}}$.

of notes when $K=2$ (94.9% if we omit octave errors), for 75% of notes when $K=3$ (92.8% if we omit octave errors), and for 71.0% of notes when $K=4$ (80.3% if we omit octave errors).

C. Study of some error cases

As outlined above, some error cases occur when the estimated pitch of a note is not related by octave errors to the true pitch. Among these error cases, we will focus on extracts #3.15 and #4.8. Figure 9 displays the spectra of these extracts, together with the spectra of the estimated notes. As can be seen, the error notes are not located at random frequencies: for extract #3.15, the error note is located at pitch $\hat{\tau}\approx -4$, which is one octave below an already existing note ($\tau=8$). The note at $\tau=-13$ has simply been missed, certainly because its amplitude is lower than that of other existing notes. In order to test the stability of this solution, we have run again the MCMC algorithm three times. The pitch estimation results are $(-4.19,-6.99,8.02)$, $(-3.97,-7.01,1.19)$, $(-13,02,-6.98,8.01)$, and $(-3.99,-7.00,-19.75)$, whereas the true pitches are $(-13,-7,8)$. The corresponding residuals energy, expressed as a fraction of the original signal energy, are 2.85%, 4.00%, 0.09%, and 2.43%. These four solutions are all stable local solutions, and the overall one (found at MCMC run #3) is not reached at each time. It seems in these cases that the algorithm may be trapped in local probability maxima, thus not exploring all of the high-probability regions. Longer MCMC chains could be run to overcome this, but convergence might still not be achieved in a reasonable computation time. Multiple chain approaches are perhaps more promising here, either using informal ideas such as picking the chain with highest posterior probability or lowest

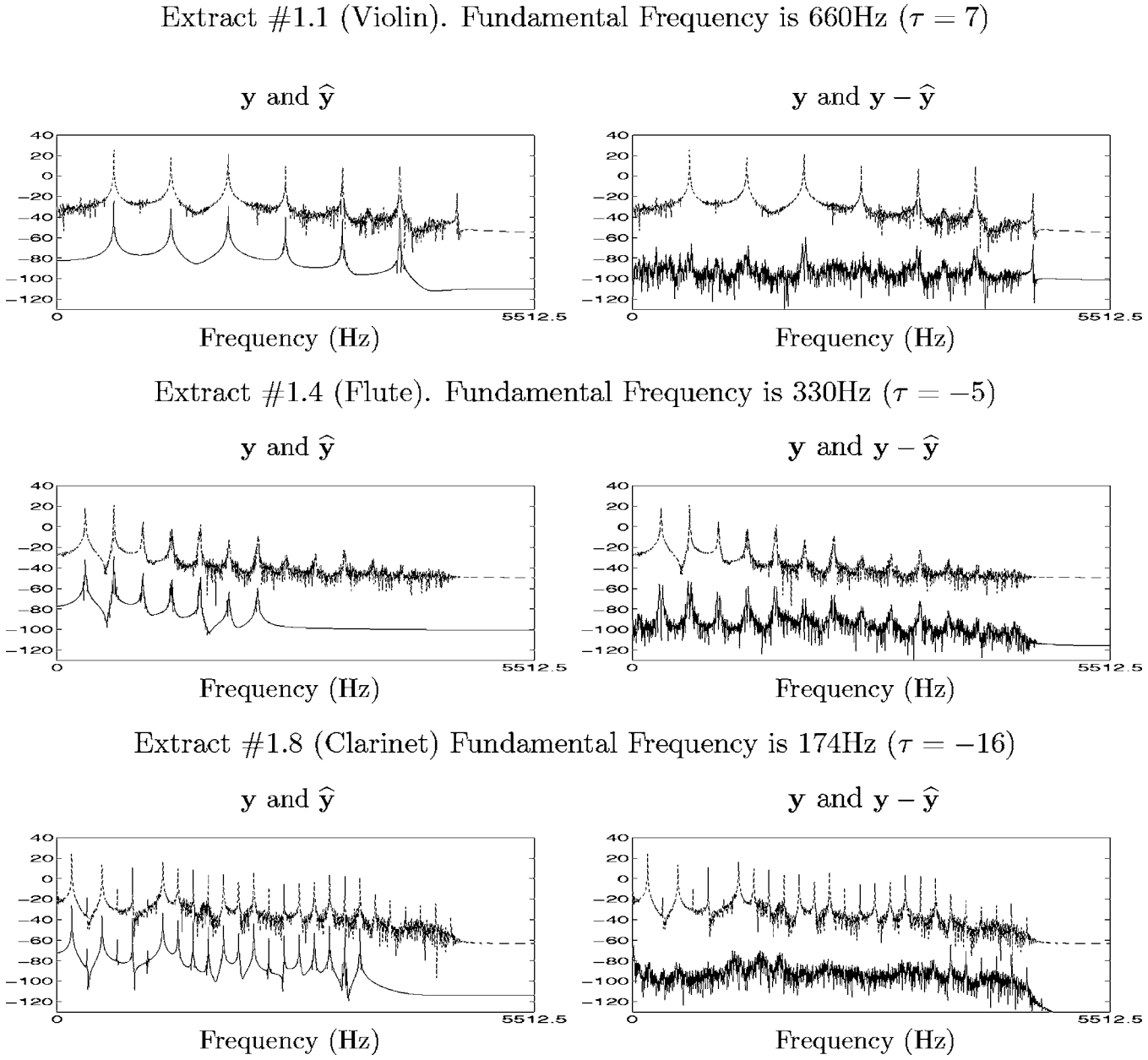


FIG. 6. Monophonic case, estimation results for three test signals. Frequency domain spectra (in dB) of original signals $\mathbf{y}$ (dashed lines), reconstructed signals $\hat{\mathbf{y}}$ (solid lines), and the residuals $\mathbf{y}-\hat{\mathbf{y}}$ (solid lines). The spectra of $\hat{\mathbf{y}}$ and of $\mathbf{y}-\hat{\mathbf{y}}$ are plotted with a −50-dB offset for improved visibility.

TABLE I. Estimation situations and results for 20 monophonic tests.

Extract	Instrument	Pitch τ	Estimated pitch $\hat{\tau}$	Extract	Instrument	Pitch τ	Estimated pitch $\hat{\tau}$
1.1	Violin	7	7.01	1.11	Trumpet	10	9.98
1.2	Fr. Horn	−27	−26.89	1.12	Oboe	13	12.99
1.3	Oboe	−7	−6.98	1.13	Flute	−7	−8.68
1.4	Flute	−5	−4.94	1.14	Flute	−4	−3.89
1.5	Trumpet	−5	−5.03	1.15	Fr. Horn	−27	−26.96
1.6	Trumpet	−9	−8.98	1.16	Violin	−4	−4.09
1.7	Fr. Horn	−17	−17.01	1.17	Oboe	13	12.99
1.8	Clarinet	−16	−15.97	1.18	Oboe	−7	−6.98
1.9	Violin	−9	−9.01	1.19	Flute	−2	−1.91
1.10	Viola	3	2.99	1.20	Violin	−3	−2.94

TABLE II. Estimation situations and results for 20 polyphonic tests, with $K=2$.

Extract	Pitch τ		Estimated pitch $\hat{\tau}$		Extract	Pitch τ		Estimated pitch $\hat{\tau}$	
2.1	1	5	1.00	5.00	2.11	6	−7	5.99	−7.03
2.2	−1	7	−1.06	7.06	2.12	−7	−1	−7.03	−13.05
2.3	−9	−2	−1.96	−2.06	2.13	7	−3	7.07	19.07
2.4	1	11	1.048	11.03	2.14	12	−13	12.00	−25.00
2.5	−7	13	−6.87	13.00	2.15	10	−2	−1.94	−1.90
2.6	−11	10	−11.03	10.06	2.16	−12	−12	−11.99	7.02
2.7	−15	−19	−14.98	−18.96	2.17	−13	−4	−13.26	−3.79
2.8	−13	−4	−12.95	−4.17	2.18	2	8	20.02	8.02
2.9	10	4	10.07	4.07	2.19	−3	3	−2.87	3.03
2.10	−10	5	−10.15	−6.8	2.20	−1	14	−0.91	14.22

residual error, or more formally using multiple-chain MCMC methods and annealing (see Refs. 51 and 52). These promising possibilities will be investigated in our next paper.

Concerning extract #4.8, the error can be explained as follows: though this is not an octave error, we see that the third partial (i.e., the partial such that $m=3$) of the note with $\tau=-8$ is located at $\tau=11$ (this is the pitch of the note that has been missed). Its fourth partial ($m=4$) is located at $\tau=16$, which also corresponds to the third partial ($m=3$) of the note located at $\tau=-3$. This explains the stability of this solution: positioning a note at $\tau=-8$ captures some of the energy of the two notes located at $\tau=\{-3,11\}$. Similar error cases can be found for extracts #4.8, #4.10, #4.11, etc.: they can be easily understood by noticing that the third partial of any note with pitch τ is located at pitch $\tau+19$. No doubt these occasional problems are a combination of misconverged

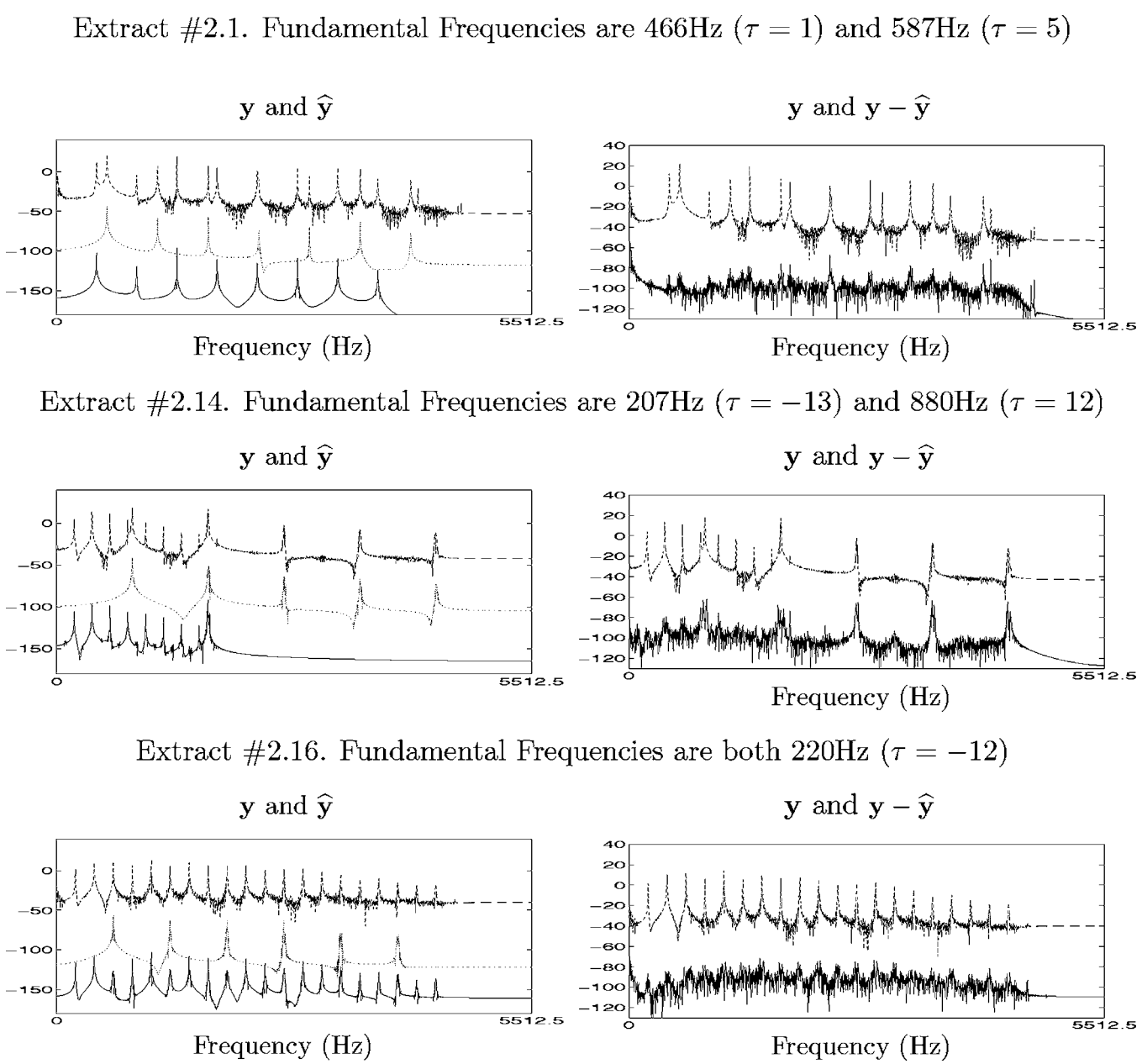


FIG. 7. Polyphonic case ($K=2$), estimation results for three test signals. Frequency domain spectra (in dB) of original signals $\mathbf{y}$ (dashed lines), reconstructed signals $\hat{\mathbf{y}}$ (note #1 in solid lines with −50-dB offset, note #2 in dotted lines with −100-dB offset) and the residuals $\mathbf{y}-\hat{\mathbf{y}}$ (solid lines with −50-dB offset).

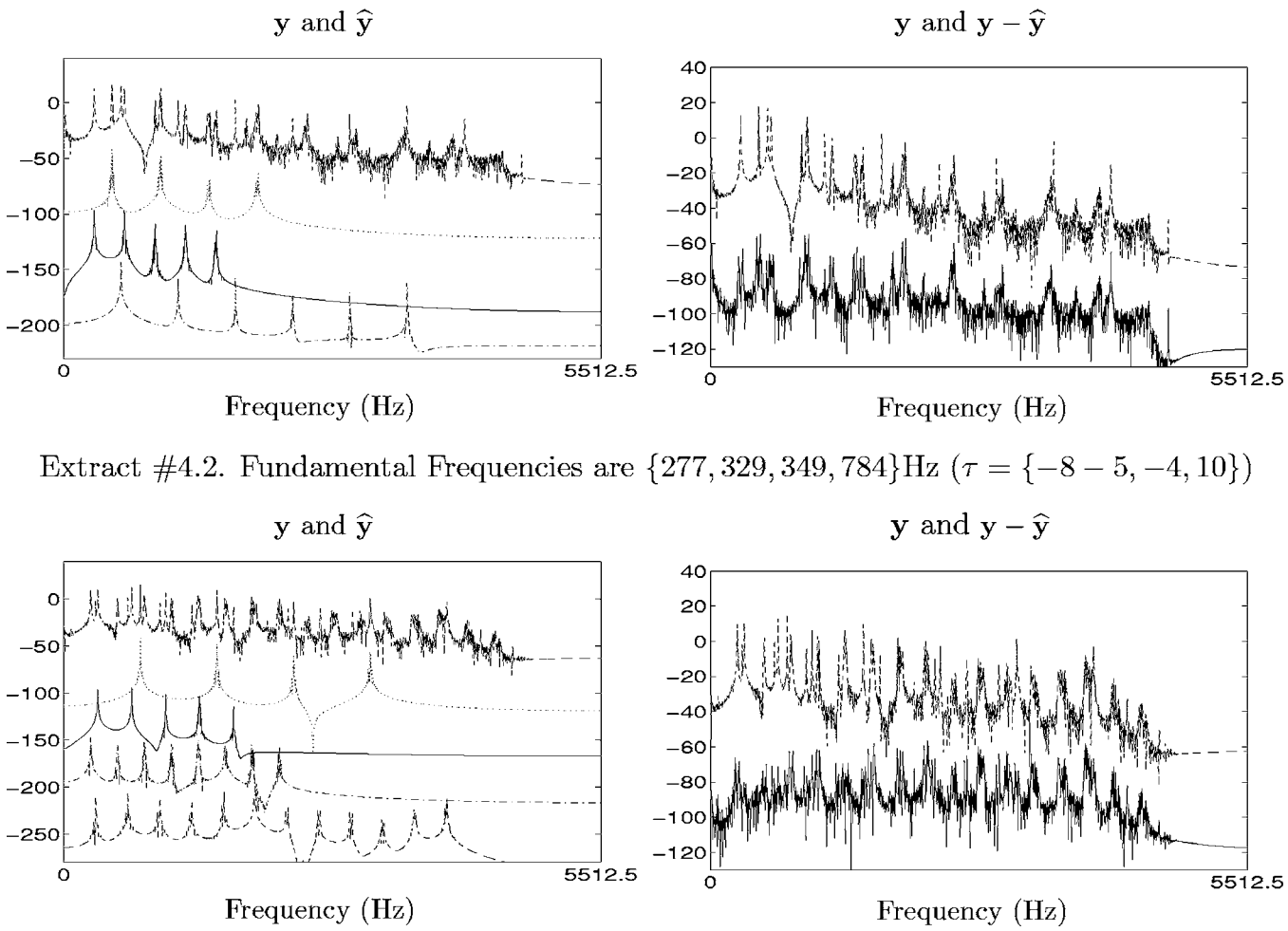


FIG. 8. Polyphonic case ($K=3$ and $K=4$), estimation results for two test signals. Frequency domain spectra (in dB) of original signals $\mathbf{y}$ (dashed lines), reconstructed signals $\hat{\mathbf{y}}$ (note #1 in solid lines with −50-dB offset, note #2 in dotted lines with −100-dB offset, note #3 in dash-dotted lines with −150-dB offset, and note #4 in dashed lines with −200-dB offset) and the residuals $\mathbf{y}-\hat{\mathbf{y}}$ (solid lines with −50-dB offset).

MCMC runs and/or ambiguous data for our models. We have found no straightforward solutions to this problem, although one possibility is to look at the estimated partial amplitudes: for the error note located at pitch $\tau=-8$, partials $m=1$, $m=2$, and $m=5$ have low amplitude compared to partials $m=3$, $m=4$, $m=6$, and $m=7$. In notes produced by natural instruments, such an amplitude profile would be quite unlikely and so priors based on more regular amplitude laws could be incorporated, thus penalizing such erroneous solutions.

D. Performance from the auditory viewpoint

Aside from pitch estimation, a very promising feature of this Bayesian method is the possibility to separate several notes from a mixture, even if only one microphone is used to record the extract (monophonic extracts as opposed to stereophonic extracts). A very promising result is that there is almost no audible difference between $\mathbf{y}$ and $\hat{\mathbf{y}}$ in most cases. Moreover, in the case of octave errors, the corresponding reconstructed note typically sounds very close to the original note. Thus, from the auditory viewpoint, the performance of the method is higher than from the pitch estimation viewpoint. We note that Virtanen and Kalpuri[53] have used their frequency domain models to perform separation of notes in a chord.

E. Estimation of *K*

In the results presented above, the number of notes is assumed known. In principle, though, the model and the algorithm are designed to estimate K. In practice, however, letting K as a free parameter systematically leads to too many notes with, typically, two notes located at the same position. This may be overcome by imposing small K via the prior over Λ', but this approach fails in examples with many true notes.

The birth/death moves are nevertheless quite important insofar as K is initialized to $K=1$ and slowly increases until reaching $K_{\max}$ (the true K assume known). A good solution to actually estimate K is to run MCMC simulations with increasing $K_{\max}$ from one to, e.g., ten and monitoring the minimum residual energy over several MCMC runs for each $K_{\max}$. When $K_{\max}$ reaches the true number K, the residual

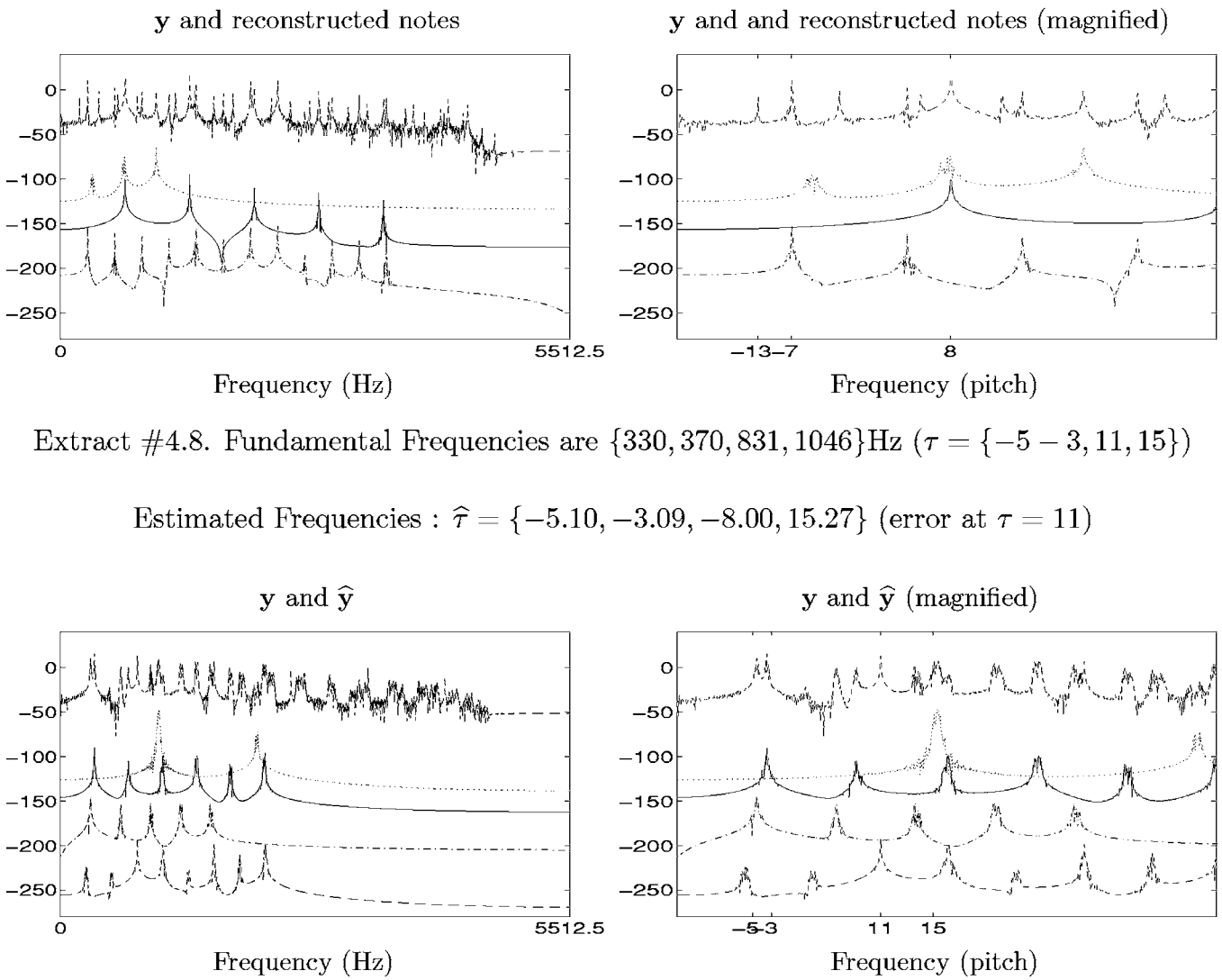


FIG. 9. Polyphonic case (K=3 and K=4), estimation results for two test signals. Frequency domain spectra (in dB) of original signals $\mathbf{y}$ (dashed lines), and reconstructed notes (note #1 in solid lines with −50-dB offset, note #2 in dotted lines with −100-dB offset, note #3 in dash-dotted lines with −150-dB offset, and note #4 in dashed lines with −200-dB offset).

energy suddenly stops decreasing, yielding the actual K. It should be noted that estimating K is a complicated problem, even for trained human listeners.

F. Comparison with previous algorithms

The model presented here is an improved version of previous Bayesian models using MCMC.[27–30] In order to illustrate the improvement in practice, we have processed the 20 three-note extracts with a model with constant amplitudes (that is, our model with a single rectangular window ϕ_0) and no inharmonicity (that is, with $\omega_{k,m}=m\omega_{k,1}$). The remaining algorithm parameters are kept the same. Then we are essentially back to the model by Walmsley *et al.*[27,28] The pitch is correctly estimated for 62.5% of notes (here K=3) and 78.6% if we omit octave errors. This is substantially lower than with the new model (75% and 92.8%). Moreover, the notes are poorly reconstructed with this model: the average residual energy calculated over the 20 extracts is 27.5% of the original signal energy. This is much higher than with the full model (only 5%). This is mainly due to two reasons: (1) the amplitudes are nonconstant with real instruments over time and (2) without the inharmonicity parameters, the model cannot accurately fit partials with order m larger than about 4, because their inharmonicity can be too great [the $\delta_{k,m}$'s in model (7) were found to be nonzero for most of the sounds processed]. A similar improvement in performance with the new model is obtained when comparing with a previous inharmonicity model having additive $\delta_{k,m}$ parameters rather than multiplicative, as we initially proposed in our

previous works:[29,30] for large m, the additive $\delta_{k,m}$ parameter should be large, which is not allowed by the prior distribution of $\delta_{k,m}$.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a novel technique for the estimation and analysis of western tonal (pitched) music. By using a Gabor atomic model together with a specially designed MCMC algorithm, we have shown that it is possible to estimate a high number of parameters from many simultaneously playing notes. Our model includes time-varying amplitudes, inharmonicity (detuned partials), and unknown number of partials. We obtain good pitch estimation results on a database of polyphonic real music extracts containing randomly generated chords with up to four notes. These results are found to be a significant improvement over previous versions of related models. More notes than four are possible with these models, but clearly the performance degrades as the number increases. Nevertheless, in informal trials the methods have been successfully able to identify and separate out the notes in chords of up to eight pitches. In addition to pitch estimation, our technique enables note separation from chords. The separated notes sound perceptually quite similar to the original, unmixed notes, thus giving some confidence in the method's ability to model the individual note components accurately, even when many partials are overlapping between different notes.

We have identified some error cases, and discussed their causes, suggesting improvements to help eliminate them, using multiple chain approaches or more informative amplitude priors, for example. In addition, considering the chords to be in the context of real musical sequences, we can usefully model musical dependencies from frame to frame in order to aid the estimation process. For example, fundamental frequency priors that depend on the previous notes played would build in a useful extra layer of prior information for the models.

APPENDIX A: PIANO MODEL FREQUENCIES DISTRIBUTION

Fletcher and Rossing[2] propose the following piano model, where partial frequencies follow the law:

$$\omega_{k,m} = m\omega_{k,1}\sqrt{\frac{1+m^2B}{1+B}}. \tag{A1}$$

In our harmonic model, one can simply implement this formula with unknown B (a typical value is $B \approx 4\times 10^{-4}$), which yields a piano music model. The frequency prior distribution can be as follows:

$$p(\boldsymbol{\omega}|K) = p(\boldsymbol{\omega}_1|K)p(B_1)\cdots p(B_K), \tag{A2}$$

where $\boldsymbol{\omega}_1=[\omega_{1,1},\ldots,\omega_{K,1}]$ is the vector of fundamental frequencies and B_k $(k=1,\ldots,K)$ are the parameters B in Eq. (A1) for each note. The piano fundamental frequency prior may encode the strong knowledge we have about possible notes, which is determined by strings tuning (equally tempered scale). For simplicity, we assume note fundamental frequencies are independent, which means

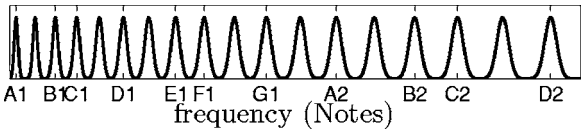


FIG. 10. Prior distribution of the fundamental frequencies $p(\omega_{k,1})$ in the piano model. The spread of the Gaussian function φ_j increases with frequency, i.e., with the strings note labels $j=\{\mathrm{A1},\mathrm{A1\#},\mathrm{B1},\ldots\}$. This is aimed at limiting areas where the prior is zero.

$$p(\boldsymbol{\omega}_1|K) = \prod_{k=1}^{K} p_{\mathrm{plano}}(\boldsymbol{\omega}_{k,1}), \tag{A3}$$

and for each note $k=1,\ldots,K$,

$$p_{\mathrm{piano}}(\omega_{k,1}) \propto \sum_{j=1}^{N_{\mathrm{strings}}} \varphi_j(\omega_{k,1}), \tag{A4}$$

where φ_j is for example a Gaussian function centered on the frequency of string $\#j$, and N_{strings} is the number of piano strings. The shape of the resulting prior is plotted in Fig. 10.

APPENDIX B: PROBABILITIES OF PERFORMING BIRTH/DEATH MOVES

The probabilities $\mu(K)$ and $\nu(K)$ governing the birth and death moves are such that

$$\mu(K) = c\,\min\left\{1, \frac{p(K+1|\Lambda')}{p(K|\Lambda')}\right\} \quad \text{for } K < K_{\max} \quad \text{with } \nu(1)=0, \tag{B1}$$

$$\nu(K+1) = c\,\min\left\{1, \frac{p(K|\Lambda')}{p(K+1|\Lambda')}\right\} \quad \text{for } K > 1 \quad \text{with } \mu(K_{\max})=0, \tag{B2}$$

where $c=0.15$ and $p(K|\Lambda')$ is the prior probability of K.

APPENDIX C: *n*-INCREASE/DECREASE AND MULTIPLY/DIVIDE MOVES

In Algorithm 5, the number of partials can be changed in two ways: either n partials are added (resp. removed) in the n-increase move (resp. n-decrease move) [see Algorithm 7 (resp. see Algorithm 8)] or the number of partials is multiplied by 2 (resp. divided by 2) in the divide move (resp. multiply move) [see Algorithm 10 (resp. see Algorithm 9)].

Algorithm 7: n-Increase move

- Compute the residual $\mathbf{r}_0=\mathbf{y}-\tilde{\mathbf{D}}^{(l-1)}\tilde{\boldsymbol{\beta}}^{(l-1)}$
- Set $M_k^{\star} \leftarrow \tilde{M}_k^{(l-1)}+n$,
- Sample the frequencies $[\omega^{\star}_{k,M_k+1},\ldots,\omega^{\star}_{k,M_k+n}]$ from the proposal distribution $q_{\mathrm{partials}}(\boldsymbol{\omega}|\mathbf{r}_0,\omega_{k,1},\{M_k+1,\ldots,M_k+n\})$
- Set $\boldsymbol{\omega}_k^{\star} \leftarrow [\boldsymbol{\omega}_k^{(l-1)},\omega^{\star}_{k,M_k+1},\ldots,\omega^{\star}_{k,M_k+n}]$.
- Let $\mathbf{P}^{\star}$ and $\mathbf{S}^{\star}$ be the $\mathbf{P}$ and $\mathbf{S}$ matrices related to candidate set of partials from $m=\tilde{M}_k^{(l-1)}+1$ to $m=M_k^{\star}$ with frequencies $[\omega^{\star}_{k,M_k+1},\ldots,\omega^{\star}_{k,M_k+n}]$. Compute

$$g_{\text{inc}} = \left(\frac{\gamma_0 + \|\mathbf{r}_0\|^2}{\gamma_0 + \mathbf{r}_0^{\text{t}} \mathbf{P}^{\bigstar} \mathbf{r}_0} \right)^{(N+\nu_0)/2} \times \frac{p(\boldsymbol{\omega}_k^{\bigstar}, M_k^{\bigstar})}{p(\boldsymbol{\omega}_k^{(l-1)}, \tilde{M}_k^{(l-1)})} \frac{d_{M_k+n}}{b_{M_k}} \det(\mathbf{S}^{\bigstar})^{1/2} \times [q_{\text{partials}}(\omega_{k,M_k+1}^{\bigstar}, \ldots, \omega_{k,M_k+n}^{\bigstar} | \mathbf{r}_0, \omega_{k,1}, \{M_k + 1, \ldots, M_k + n\})_b ig]^{-1} \quad \text{(C1)}$$

- With probability $\min(1, g_{\text{inc}})$, accept the candidate: Set $\tilde{M}_k^{(l)} \leftarrow M_k^{\bigstar}$ and $\tilde{\boldsymbol{\omega}}_k^{(l)} \leftarrow \boldsymbol{\omega}_k^{\bigstar}$ and sample $\tilde{\boldsymbol{\beta}}_{M_k+1,\ldots,M_k+n}^{(l)}$ according to $p(\boldsymbol{\beta} | \tilde{\sigma}_v^{2(l-1)}, \boldsymbol{\omega}_{k,M+n}^{\bigstar}, M = n, K = 1, \mathbf{r}_0)$ given in Eq. (22).

To ensure the theoretical convergence of the algorithm, it is necessary to implement the reverse move, the n-decrease move, presented in Algorithm 8 below.

Algorithm 8: n-Decrease move

- Set $M_k^{\bigstar} \leftarrow \tilde{M}_k^{(l-1)} - n$,
- Remove the frequencies $[\tilde{\omega}_{k,M_k-n+1}^{(l-1)}, \ldots, \tilde{\omega}_{k,M_k}^{(l-1)}]$ in $\tilde{\boldsymbol{\omega}}_k^{(l-1)}$ so as to form $\boldsymbol{\omega}_k^{\bigstar}$
- Compute the residual $\mathbf{r}_{k,M-n} = \mathbf{r}_k - \mathbf{D}_{k,-\{M+1,\ldots,M+n\}} \boldsymbol{\beta}_{k,-\{M+1,\ldots,M+n\}}$ where $\boldsymbol{\beta}_{k,-\{M+1,\ldots,M+n\}}$ (resp. $\mathbf{D}_{k,-\{M+1,\ldots,M+n\}}$) is the vector of amplitudes $\tilde{\boldsymbol{\beta}}^{(l-1)}$ (resp. the matrix of Gabor atoms $\tilde{\mathbf{D}}^{(l-1)}$) computed with parameters $\tilde{\mathbf{M}}^{(l)}$ and $\tilde{\boldsymbol{\omega}}^{(l)}$ where the entries related to partials $\{M_k - n + 1, \ldots, M_k\}$ have been removed.
- Let $\mathbf{P}_n$ and $\mathbf{S}_n$ be the $\mathbf{P}$ and $\mathbf{S}$ matrices related to the partials $M_k - n + 1, \ldots, M_k$ to be removed. Compute

$$g_{\text{dec}} = \left(\frac{\gamma_0 + \mathbf{r}_{k,M-n}^{\text{t}} \mathbf{P}_n \mathbf{r}_{k,M-n}}{\gamma_0 + \|\mathbf{r}_{k,M-n}\|^2} \right)^{(N+\nu_0)/2} \frac{p(\boldsymbol{\omega}_k^{\bigstar}, M_k^{\bigstar})}{p(\boldsymbol{\omega}_k^{(l-1)}, \tilde{M}_k^{(l-1)})} \times \frac{b_{M_k-n}}{d_{M_k}} \frac{q_{\text{partials}}(\tilde{\omega}_{k,M_k-n+1}^{(l-1)}, \ldots, \tilde{\omega}_{k,M_k}^{(l-1)} | \mathbf{r}_{k,M-n}, \omega_{k,1}, \{M_k - n + 1, \ldots, M_k\})}{\det(\mathbf{S}_n)^{1/2}} \quad \text{(C2)}$$

- With probability $\min(1, g_{\text{dec}})$, accept the candidate: Set $\tilde{M}_k^{(l)} \leftarrow M_k^{\bigstar}$, $\tilde{\boldsymbol{\omega}}_k^{(l)} \leftarrow \boldsymbol{\omega}_k^{\bigstar}$ and remove the components corresponding to partials $M_k - n + 1, \ldots, M_k$ from $\tilde{\boldsymbol{\beta}}^{(l-1)}$ so as to obtain $\tilde{\boldsymbol{\beta}}^{(l)}$.

The multiply and divide moves are synchronized with fundamental frequency division/multiplication by 2 in order to detect octave ambiguities: assume a note with frequency 440 Hz is played by an instrument. The notes with frequencies 220 and 880 Hz have partials that are almost superimposed with the partials of the actual note. Without these moves, the Markov chain may be stuck on incorrect fundamental frequencies, namely 220 or 880 Hz, which is not satisfactory.

Algorithm 9: Divide move

- Compute the residual $\mathbf{r}_0 = \mathbf{y} - \tilde{\mathbf{D}}^{(l-1)} \tilde{\boldsymbol{\beta}}^{(l-1)}$
- Set $M_k^{\bigstar} \leftarrow 2\tilde{M}_k^{(l-1)} + e$
- Form $\boldsymbol{\omega}_k^{\bigstar}$ by setting $\omega_{k,2m}^{\bigstar} \leftarrow \tilde{\omega}_{k,m}^{(l-1)}$ for $m = 1, \ldots, \lfloor M_k/2 \rfloor$.
- Sample frequencies $\tilde{\omega}_{k,2m-1}^{\bigstar}$ for $m = 1, \ldots, \lfloor M_k/2 \rfloor$ using the proposal distribution $q_{\text{partials}}(\omega | \mathbf{r}_0, \omega_{k,1}, 2m - 1)$
- Set $\boldsymbol{\omega}_k^{\bigstar} \leftarrow \boldsymbol{\omega}_k^{(l-1)} \cup \{\omega_{k,2m}^{\bigstar} \leftarrow \tilde{\omega}_{k,m}^{(l-1)}\}_{m=1,\ldots,\lfloor M_k/2 \rfloor}$.
- Let $\mathbf{P}^{\bigstar}$ and $\mathbf{S}^{\bigstar}$ be the $\mathbf{P}$ and $\mathbf{S}$ matrices related to candidate set of partials $(k, 2m-1)$ with frequencies $\tilde{\omega}_{k,2m-1}^{\bigstar}$, $m = 1, \ldots, \lfloor M_k/2 \rfloor$. Compute

$$g_{\text{div}} = \left[\frac{\gamma_0 + \|\mathbf{r}_0\|^2}{\gamma_0 + \mathbf{r}_0^{\text{t}} \mathbf{P}^{\bigstar} \mathbf{r}_0} \right]^{(N+\nu_0)/2} \det(\mathbf{S}^{\bigstar})^{1/2} \frac{p(\boldsymbol{\omega}_k^{\bigstar}, M_k^{\bigstar})}{p(\boldsymbol{\omega}_k^{(l-1)}, \tilde{M}_k^{(l-1)})} \times \prod_{m=1,\ldots,\lfloor M_k/2 \rfloor} [q_{\text{partials}}(\omega | \mathbf{r}_0, \omega_{k,1}, 2n - 1)]^{-1} \quad \text{(C3)}$$

- With probability $\min(1, g_{\text{div}})$, accept the candidate: Set $\tilde{M}_k^{(l)} \leftarrow M_k^{\bigstar}$ and $\tilde{\boldsymbol{\omega}}_k^{(l)} \leftarrow \boldsymbol{\omega}_k^{\bigstar}$ and sample $\tilde{\beta}_m$, $m = 1, \ldots, \lfloor M_k/2 \rfloor$ according to $p(\boldsymbol{\beta} | \tilde{\sigma}_v^{2(l-1)}, \tilde{\omega}_{k,2m-1}^{\bigstar}, M = 1, \mathbf{r}_0)$ given in Eq. (22).

As for any reversible jump, the divide move comes together with the reverse, the *multiply move* presented in Algorithm 10 below.

Algorithm 10: Multiply move

- Set $M_k^{\bigstar} \leftarrow (\tilde{M}_k^{(l-1)} - e)/2$,
- Form $\boldsymbol{\omega}_k^{\bigstar}$ by setting $\omega_{k,m}^{\bigstar} \leftarrow \tilde{\omega}_{k,2m}^{(l-1)}$ for $m = 1, \ldots, M_k - e$.
- Compute the residual $\mathbf{r}_{k,\text{odd}} = \mathbf{r}_k - \mathbf{D}_{k,-\{1,3,\ldots\}} \boldsymbol{\beta}_{k,-\{1,3,\ldots\}}$ where $\boldsymbol{\beta}_{k,-\{1,3,\ldots\}}$ (resp. $\mathbf{D}_{k,-\{1,3,\ldots\}}$) is the vector of amplitudes $\tilde{\boldsymbol{\beta}}^{(l-1)}$ (resp. the matrix of Gabor atoms $\tilde{\mathbf{D}}^{(l-1)}$) computed with parameters $\tilde{\mathbf{M}}^{(l)}$ and $\tilde{\boldsymbol{\omega}}^{(l)}$ where the entires related to partials {1, 3,…} have been removed.
- Let $\mathbf{P}_{\text{odd}}$ and $\mathbf{S}_{\text{odd}}$ be the $\mathbf{P}$ and $\mathbf{S}$ matrices related to the partials 1, 3,… to be removed. Compute

$$g_{\text{mul}} = \left[\frac{\gamma_0 + \mathbf{r}_{k,\text{odd}}^{\text{t}} \mathbf{P}_{\text{odd}} \mathbf{r}_{k,\text{odd}}}{\gamma_0 + \|\mathbf{r}_{k,\text{odd}}\|^2} \right]^{(N+\nu_0)/2} \times \det(\mathbf{S}_{\text{odd}})^{-1/2} \frac{p(\boldsymbol{\omega}_k^{\star}, M_k^{\star})}{p(\boldsymbol{\omega}_k^{(l-1)}, \widetilde{M}_k^{(l-1)})} \times \prod_{m=1,3,\ldots,} q_{\text{partials}}(\omega | \mathbf{r}_{k,\text{odd}}, \omega_{k,1}, 2m-1) \tag{C4}$$

- With probability $\min(1, g_{\text{mul}})$, accept the candidate: Set $\widetilde{M}_k^{(l)} \leftarrow M_k^{\star}$, $\widetilde{\boldsymbol{\omega}}_k^{(l)} \leftarrow \boldsymbol{\omega}_k^{\star}$ and remove the components corresponding to partials 1, 3,… from $\widetilde{\boldsymbol{\beta}}^{(l-1)}$ so as to obtain $\widetilde{\boldsymbol{\beta}}^{(l)}$.

Note: In all the ratios g presented above, the Jacobian of the transforms for birth/death, increase/decrease, divide/multiply should appear (see Ref. 43). Here, it equals 1, thus it does not appear.

[1] A. Bregman, *Auditory Scene Analysis* (MIT, Cambridge, USA, 1990).
[2] N. Fletcher and T. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, Berlin, 1998).
[3] M. McIntyre, R. Schumacher, and J. Woodhouse, "On the oscillations of musical instruments," J. Acoust. Soc. Am. **74**, 1325–1345 (1983).
[4] A. de Cheveigné, "Separation of concurrent harmonic sounds: Fundamental frequency estimation and a time-domain cancellation model for auditory processing," J. Acoust. Soc. Am. **93**, 3271–3290 (1993).
[5] A. de Cheveigné and H. Kawahara, "Multiple period estimation and pitch perception model," Speech Commun. **27**, 175–185 (1999).
[6] T. Virtanen and A. Klapuri, "Separation of harmonic sound sources using sinusoidal modeling," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE ICASSP, Istanbul, Turkey, 2000, Vol. **2**, pp. 765–768.
[7] T. Virtanen and A. Klapuri, "Separation of harmonic sounds using linear models for the overtone series," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE ICASSP, Orlando, FL, 2002.
[8] R. Irizarry, "Local harmonic estimation in musical sound signals," J. Am. Stat. Assoc. **96**(454), 357–367 (2001).
[9] R. Irizarry, "Weighted estimation of harmonic components in a musical sound signal," J. Time Ser. Anal. **23**(1), 29–48 (2002).
[10] H. Kameoka, T. Nishimoto, and S. Sagayama, "Separation of harmonic structures based on tied gaussian mixture model and information criterion for concurrent sounds," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE ICASSP, Montreal, Canada, 2004.
[11] K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka, "Organisation of hierarchical perceptual sounds: Music scene analysis with autonomous processing modules and a quantitative information integration mechanism," in *International Joint Conference on Artificial Intelligence*, IJCAI, Montreal, Quebec, 1995, pp. 158–164.
[12] K. Kashino and H. Murase, "A sound source identification system for ensemble music based on template adaptation and music stream extraction," Speech Commun. **27**, 337–349 (1999).
[13] K. Kashino and S. J. Godsill, "Bayesian estimation of simultaneous musical notes based on frequency domain modelling," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE ICASSP, Montreal, Canada, 2004.
[14] C. Raphael, "Automatic transcription of piano music," in *International Conference on Music Information Retrieval*, ISMIR, Paris, France, 2002.
[15] C. Raphael and J. Stoddard, "Harmonic analysis with probabilistic graphical models," in *International Conference on Music Information Retrieval*, ISMIR, Baltimore, MD, 2003.
[16] J. Tabrikian, S. Dubnov, and Y. Dickalov, "Maximum a-posteriori probability pitch tracking in noisy environments using harmonic model," IEEE Trans. Speech Audio Process. **12**(1), 76–87 (2004).
[17] A. Sterian, M. Simoni, and G. Wakefield, "Model-based musical transcription," in *International Computer Music Conference*, ICMC, Beijing, China, 1999.
[18] L. Parra and U. Jain, "Approximate Kalman filtering for the harmonic plus noise model," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE WASPAA, New Paltz, NY, 2001.
[19] A. Cemgil, H. Kappen, and D. Barber, "Generative model based polyphonic music transcription," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE WASPAA, New Paltz, NY, 2003.
[20] A. Cemgil, B. Kappen, and D. Barber, "A generative model for music transcription," IEEE Trans. Speech Audio Process. (to be published).
[21] H. Laurent and C. Doncarli, "Stationarity index for abrupt changes detection in the time-frequency plane," IEEE Signal Process. Lett. **5**(2), 43–45 (1998).
[22] M. Davy and S. Godsill, "Detection of abrupt spectral changes using support vector machines an application to audio signal segmentation," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE ICASSP, Orlando, FL, 2002, Vol. **2**, pp. 1313–1316.
[23] F. Desobry, M. Davy, and C. Doncarli, "An online kernel change detection algorithm," IEEE Trans. Signal Process. **53**(5), 2961–2974 (2005).
[24] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes-Theory and Application* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
[25] J. Paulus and A. Klapuri, "Measuring the similarity of rhythmic patterns," in *International Conference on Music Information Retrieval*, ISMIR, Paris, France, 2002.
[26] C. Duxbury, M. Sandler, and M. Davies, "A hybrid approach to musical note onset detection," in *International Conference on Digital Audio Effects*, DAFX, Hamburg, Germany, 2002.
[27] P. Walmsley, S. J. Godsill, and P. J. W. Rayner, "Multidimensional optimisation of harmonic signals," in *European Signal Processing Conference*, EURASIP, Island of Rhodes, Greece, 1998.
[28] P. J. Walmsley, S. J. Godsill, and P. J. W. Rayner, "Polyphonic pitch tracking using joint Bayesian estimation of multiple frame parameters," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE WASPAA, New Paltz, NY, 1999.
[29] M. Davy and S. Godsill, "Bayesian harmonic models for musical signal analysis," in *Seventh Valencia International Meeting, Bayesian Statistics 7* (Oxford U.P., Tenerife, Spain, 2002).
[30] S. Godsill and M. Davy, "Bayesian harmonic models for musical pitch estimation and analysis," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE ICASSP, Orlando, FL, 2002.
[31] H. Feichtinger and T. Strohmer, *Gabor Analysis and Algorithms: Theory and Applications, Applied and Numerical Harmonic Analysis* (Birkhauser, Boston, 1998).
[32] P. J. Wolfe, S. J. Godsill, and W. Ng, "Bayesian variable selection and regularisation for time-frequency surface estimation," J. R. Stat. Soc. Ser. B. Methodol. **66**(3), 575–589 (2004).
[33] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," IEEE Trans. Signal Process. **51**(1), 101–111 (2003).
[34] J. Rissanen, "A universal prior for integers and estimation by minimum description length," Ann. Stat. **11**(2), 416–431 (1983).
[35] H. Akaike, "A new look at statistical model identification," IEEE Trans. Autom. Control **19**, 716–723 (1974).
[36] G. Schwartz, "Estimating the dimension of a model," Ann. Stat. **6**(2), 461–464 (1985).
[37] M. Stephens, "Dealing with label-switching in mixture models," J. R. Stat. Soc. Ser. B. Methodol. **62**(4), 795 (2000).
[38] A. Jasra, C. Holmes, and D. Stephens, "MCMC and the label switching problem in Bayesian mixture models," Stat. Sci. **20**(1), 50–67 (2005).
[39] C. Andrieu and A. Doucet, "Joint Bayesian detection and estimation of noisy sinusoids via reversible jump MCMC," IEEE Trans. Signal Process. **47**(10), 2667–2676 (1999).
[40] M. Davy and J. Idier, "Fast MCMC computations for the estimation of sparse processes from noisy observations," in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE ICASSP, Montreal, Canada, 2004.
[41] In practice, these parameters are chosen as $\nu_0 = \gamma_0 = 0$. The corresponding prior is no longer a proper probability density function (because its integral over σ_v^2 is infinite), but this is not a problem because the posterior distribution remains a proper probability density function.
[42] *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Chapman & Hall, London, 1996).
[43] C. Robert and G. Casella, *Monte Carlo Statistical Methods* (Springer, New York, 2000).
[44] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration: A Statistical Model-Based Approach* (Springer, Berlin, 1998).
[45] S. J. Godsill, "On the relationship between Markov chain Monte Carlo

methods for model uncertainty," J. Comput. Graph. Stat. **10**(2), 230–248 (2001).
[46]J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill, "Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes," J. Time Ser. Anal. **25**(6), 785–809 (2004).
[47]D. Battle, P. Gerstoft, W. Hodgkiss, W. A. Kuperman, and P. Nielsen, "Bayesian model selection applied to self-noise geoacoustic inversion," J. Acoust. Soc. Am. **116**, 2043–2056 (2004).
[48]P. J. Green, "Reversible jump Markov-chain Monte Carlo computation and Bayesian model determination," Biometrika **82**(4), 711–732 (1995).
[49]For a proof, see Ref. 39, where a related frequency estimation algorithm is analyzed.
[50]M. Davy and S. Godsill, "Bayesian harmonic models for musical pitch estimation and analysis," Technical Report No. CUED/F-INFENG/TR.431, Department of Engineering, University of Cambridge, Cambridge, UK (unpublished).
[51]W. R. Gilks and G. O. Roberts, in *Markov Chain Monte Carlo in Practice*, edited by W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Chapman & Hall, London, 1996); Chap. Strategies for Improving MCMC, pp. 89–110. See Ref. 42.
[52]P. Del Moral and A. Doucet, "On a class of genealogical and interacting metropolis models," in *Séminaire de Probabilités*, Vol. **XXXVII** of *Lecture Notes in Mathematics 1832*, edited by J. Azéma, M. Emery, M. Ledoux, and M. Yor (Springer, Berlin, 2003), pp. 415–446.
[53]T. Virtanen and A. Klapuri, "Separation of harmonic sounds using multipitch analysis and iterative parameter estimation," in *Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE WASPAA, New Paltz, NY, 2001, pp. 83–86.

Investigation of superharmonic sound propagation and imaging in biological tissues *in vitro*

Qingyu Ma
State Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China and Education Technique Department, Nanjing Normal University, Nanjing 210097, China

Dong Zhang, Xiufen Gong,[a] and Yong Ma
State Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China

(Received 20 June 2005; revised 16 January 2006; accepted 21 January 2006)

This article presents both theoretical and experimental studies on the superharmonic generation and its imaging in biological tissues. A superharmonic component is defined as a summation of the third-, fourth-, and fifth-order harmonics. A superharmonic signal is produced using an 8-mm-diam, 2.5-MHz planar piston source that is excited by eight-cycle, 2.5-MHz tone bursts. Axial and lateral field distributions of the superharmonic component and the second harmonic are first calculated based on the nonlinear KZK model and then compared with those experimentally determined at two different source pressures of 0.5 and 1 MPa. Results indicate that the amplitude of the superharmonic component can exceed that of the second harmonic, depending on the axial distance and the fundamental pressure amplitude. Also, the 3-dB beamwidth of the superharmonic component is about 23% narrower than that of the second harmonic. Additional experiments are performed *in vitro* using liver and fatty tissues in transmission mode and produced two-dimensional images using the fundamental, the second harmonic, and the superharmonic signals. Although the clinical applicability of this work still needs to be assessed, these results indicate that the superharmonic image quality is better than that of the other two images. [DOI: 10.1121/1.2177570]

PACS number(s): 43.80.Ev, 43.80.Cs [FD] Pages: 2518–2523

I. INTRODUCTION

Among the significant achievements made in the development of ultrasound imaging techniques over the past decade, those in the fields of ultrasound contrast agent (UCA) and tissue harmonic imaging are most prominent.[1,2] The nonlinear response of microbubbles in UCA to ultrasound irradiation contains not only the fundamental frequency but also the second- and higher-order harmonic frequencies; it is thus possible to separate the perfused and nonperfused tissues.[3,4] Contrast second harmonic imaging shows considerable improvements in image quality compared to that of the fundamental mode, and now is available in commercial ultrasonic diagnosis systems.[5–7] Tissue harmonic imaging uses the harmonic components generated from nonlinear propagation in biological tissues.[8–10] The principles of the nonlinear distortion were discussed theoretically and experimentally as early as the 1980s.[11,12] Christopher investigated the use of higher harmonic components produced by a finite amplitude distortion to improve the resolution of inhomogeneous pulsed echoes in ultrasonic imaging.[13,14] Other investigators reported the nonlinear effects in biological tissues and the acoustic nonlinearity parameter imaging.[15–21] Tissue harmonic imaging is superior to the fundamental imaging mode due to improvement of the spatial resolution and suppression of side lobe levels. It represents major advantages in development of medical imaging techniques.[22] However, two typical drawbacks are associated with this technique. One is that the signal-to-noise ratio (SNR) is low because the amplitude of the second harmonic is about 20 dB lower than that at the fundamental frequency. The other is related to the adverse effect between the bandwidth of the transmitted signal and the axial resolution. Shen and Li indicated that the side lobe levels and the harmonic leakage of the harmonic beam pattern increase with the bandwidth of the transmitted signal, but a trade-off exists between axial resolution and contrast resolution.[23] Because the amplitudes of the third or higher harmonics are even 10 to 70 dB lower than that of the second harmonic, the high-order harmonics are restricted by much lower SNR in imaging, although they can provide higher spatial resolution than the fundamental and second harmonic components. Several techniques have been developed to enhance the amplitude and SNR of the harmonic components. A pulse inversion technique[24–26] was developed to enhance the second harmonic by 6 dB and effectively suppress the fundamental signal. Nonetheless, this technique requires twice as much data acquisition time as the band-pass filtering approach around the second harmonic frequency. In other words, it suffers from frame rate reduction and potential motion artifacts.[27] Bouakaz and de Jong proposed a superharmonic imaging technique,[28,29] in which the superharmonic component was defined as a linear superposition of the third, fourth, and fifth harmonic components. They proved in a theoretical simulation that the superharmonic component possesses both greater energy and better beam

[a]Author to whom correspondence should be addressed. Electronic mail: dzhang@nju.edu.cn

 0001-4966/2006/119(4)/2518/6/$22.50

pattern performance than those of the second harmonic. The superharmonic images obtained by this technique showed an improvement of the contrast performance over the conventional second harmonic image.

This paper furthers the study on superharmonic imaging technique and its comparison with the fundamental and second harmonic both theoretically and experimentally, which enhances the optimization of the applications of this technique. By employing the nonlinear Khokhlov-Zabolotskaya-Kuznetsov (KZK) equation,[30,31] the axial and lateral distributions of the superharmonic component (the linear combination of the third, fourth, and fifth harmonic components) from a planar piston source are theoretically calculated and then examined by the measurement in water. Comparisons between the superharmonic and the second harmonic components are associated with the source amplitude and the propagation distance for optimization of the superharmonic imaging. In addition, two-dimensional superharmonic imaging experiments are performed using porcine liver and fatty tissue samples *in vitro* in transmission mode and compared with the images obtained by the fundamental and the second harmonic imaging techniques.

II. THEORETICAL MODEL AND NUMERICAL IMPLEMENTATION

When a finite amplitude wave propagates in a medium, the nonlinear effects occur and the second and the higher harmonics are generated. The widely accepted KZK model[30,31] is used to calculate the amplitudes of the fundamental, second, third, fourth, and fifth harmonics during the nonlinear sound propagation, and its nondimensional form is

$$\left(4\frac{\partial^2}{\partial\tau\partial\sigma}-\nabla_{\perp}^2-4\alpha r_0\frac{\partial^3}{\partial\tau^3}\right)\bar{p}=2\frac{1}{\sigma_D}\frac{\partial^2}{\partial\tau^2}\bar{p}^2, \tag{1}$$

where $\sigma=z/r_0$ is the normalized axial distance, $r_0=ka^2/2$ is the Rayleigh distance, z is the coordinate along the direction of sound propagation, $k=\omega/c_0$ is the wave number. $r=\rho/a$ is the normalized lateral distance, a is the radius of the transmitter, and ρ is the distance from the center of the source. $\tau=\omega(t-z/c_0)$ is the retarded time, t is the time, ω is the angular frequency, and c_0 is the isentropic sound speed. $\bar{p}=(p-p_0)/\rho_0c_0u_0$, in which p is the sound pressure, p_0 is the static pressure, ρ_0 is the density of the medium, and u_0 is the peak amplitude of the normal velocity on the source. $\nabla_{\perp}^2=(\partial^2/\partial r^2)+(1/r)(\partial/\partial r)$ denotes the nondimensional transverse Laplace operator with respect to r. $\alpha=D\omega^2/2c_0^3$ is the absorption coefficient, in which D is the diffusivity of sound. $\sigma_D=l_D/r_0$, $l_D=c_0/\beta M\omega$ is the shock formation distance of a plane wave, β is the acoustic nonlinear coefficient, and $M=u_0/c_0$ is the Mach number.

Equation (1) is solved numerically using the finite difference method in frequency domain.[32–34] Numerical simulations are performed for a planar piston source with a diameter of 8 mm operating at 2.5 MHz. Distilled water is considered as the medium for ultrasound propagation (Fig. 1). The sound velocity is 1492 m/s at 23 °C (Ref. 35) and the attenuation coefficient is 25×10^{-15} Np m^{-1} Hz^{-2} with a frequency dependency of f^2.[34] The coefficient of nonlinearity

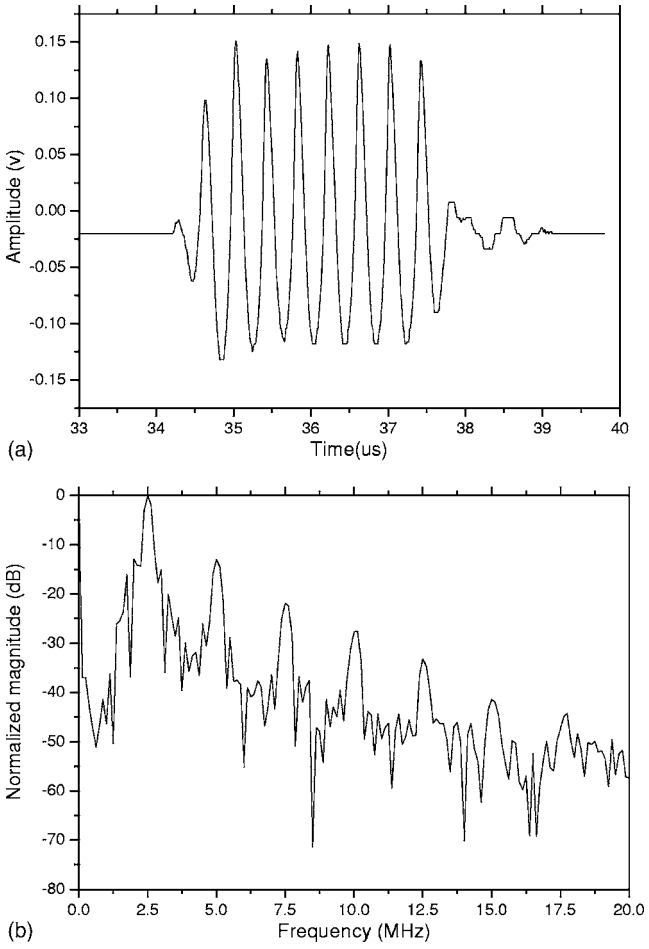


FIG. 1. Waveform (a) and frequency spectrum (b) of the received transmission signal in water at $p_0=0.5$ MPa and $\sigma=0.6$.

β is set to 3.5.[36] The nondimensional axial and lateral grid steps of $\Delta\sigma$ and Δr are equal to 10^{-4} and 5×10^{-3}, respectively. The calculated steps of the axial and lateral spatial spaces are set to 10 000 and 2000; hence the maximum values for axial and lateral directions are 1 and 10, respectively. The maximum number of the calculated harmonic order is 50.

After the field distributions of the second, third, fourth, and fifth harmonics are calculated, the distribution of the superharmonic component is obtained by the linear summation of the absolute amplitudes of the third, fourth, and fifth harmonics. Figure 2 shows the comparison of the axial sound pressure distributions between the superharmonic and second harmonic components at a source pressure of 0.5 MPa. Figure 3 gives the same comparison at a source pressure of 1 MPa. In these two figures, the dashed and solid lines represent the calculated superharmonic and the second harmonic components, respectively, which are both normalized by the maximum sound pressure of the second harmonic. Figure 4 displays the normalized lateral beam patterns of the superharmonic and the second harmonic components at a normalized axial distance of $\sigma=0.6$. A comparison with the experimental data will be discussed in the next section.

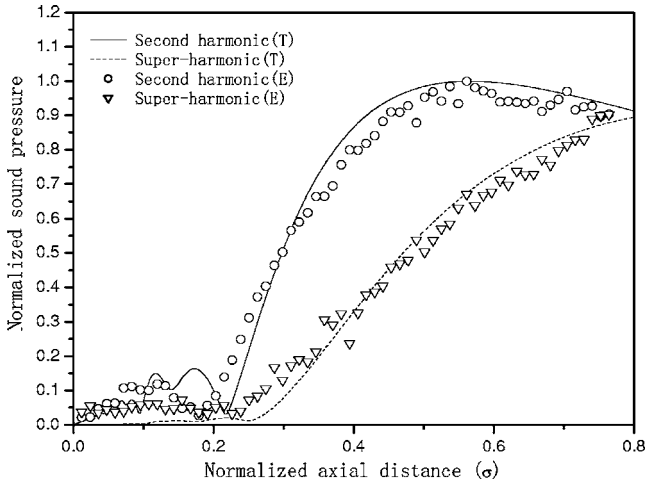


FIG. 2. Comparison of the normalized axial sound pressures between the superharmonic and the second harmonic components at p_0=0.5 MPa (T—theory, E—experiment, normalized by the maximum amplitude of the second harmonic component).

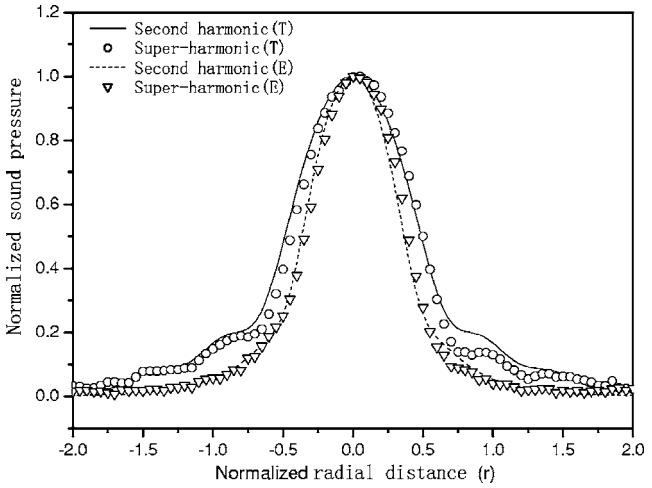


FIG. 4. Comparison of the normalized lateral beam patterns between the superharmonic and the second harmonic components at σ=0.6 (T—theory, E—experiment).

III. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental setup

Figure 5 shows the arrangement of the experimental setup for studying the superharmonic sound propagation and imaging for biological tissues. A programmable function generator (Agilent 33250, Loveland, CO) is controlled by a computer via GPIB interface and generates a tone burst signal (2.5 MHz, 8 cycles, pulse repetition frequency 1 kHz). This signal is first amplified using a broadband 55-dB power amplifier (ENI A150, Rochester, NY) and then sent to a PZT transducer (center frequency 2.5 MHz, radius 4 mm). The Rayleigh distance of the transmitter is r_0=84.2 mm. The received acoustic pressure is detected by a hydrophone (NP1000, PVDF needle hydrophone, Model TNU001A, NTR Systems, Seattle, WA) with an active area 0.6 mm in diameter. Although the spatial averaging of the receiver will reduce the spatial resolution, the beam pattern is relatively wide at the receiving position. The needle hydrophone works from 0.1 to 20 MHz and has a relatively flat frequency response from 1 to 12.5 MHz. In our experiment, the fundamental (2.5 MHz) up to the fifth harmonics (12.5 MHz) is detected by this hydrophone within its effective frequency range. Both the transmitter and the receiver are coaxially fixed and immersed in distilled water. A sound permeable sample container with a thickness of 20 mm is placed between the hydrophone and the transmitter and its rear surface is about 1 mm from the receiver to minimize the influence of sound diffraction. After passing through a 30-dB preamplifier (PFS017A, NTR Systems, Seattle, WA), the received signals are acquired using a digital oscilloscope (Agilent 54810, Colorado Springs, CO). The digitized signals are further processed using a fast Fourier transform (FFT) software. The superharmonic component amplitude is thus obtained by the linear summation of the third, fourth, and fifth harmonic components.

A mechanical scanning system (Newport Motion Controller MM3000, Irvine, CA) is used in two-dimensional motion for the measurement of the superharmonic sound distribution and the superharmonic imaging. The axial and lateral resolutions are 1 μm.

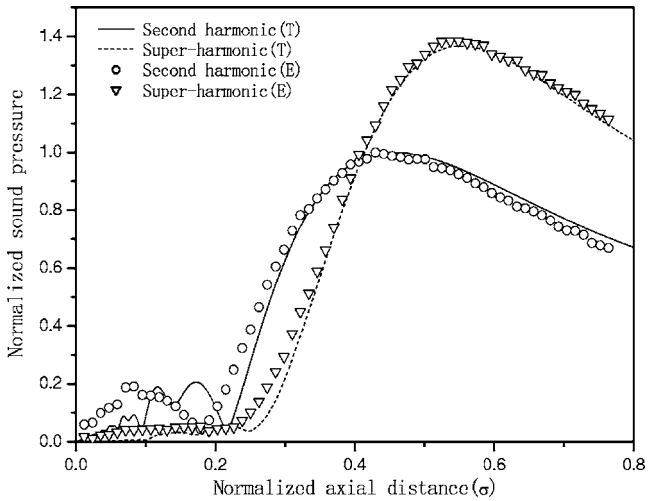


FIG. 3. Comparison of the normalized axial sound pressures between the superharmonic and the second harmonic components at p_0=1 MPa (T—theory, E—experiment, normalized by the maximum amplitude of the second harmonic component).

B. Experimental results and discussions

When the transducer is excited at a source pressure of 0.5 MPa, the transmitted signal is received by the needle

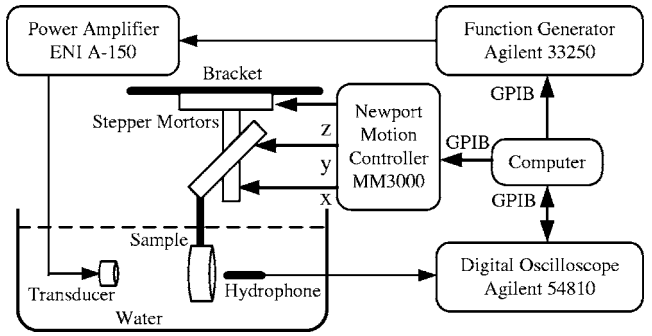


FIG. 5. The diagram of experimental setup.

hydrophone at a normalized axial distance of $\sigma=0.6$ (50 mm). The waveform is shown in Fig. 1(a). The corresponding spectrum is given in Fig. 1(b), which is normalized by the amplitude at the fundamental frequency. It is observed from Fig. 1 that the waveform distortion occurs and the second up to the sixth harmonics are generated during sound propagation. The superharmonic component is obtained by the linear summation of the amplitudes of the third, fourth, and fifth harmonics and its amplitude is about 65% and 15% of those of the second harmonic and fundamental components, respectively.

The superharmonic component is generated from a source with an 8-mm diameter and 2.5-MHz central frequency in water. The axial distribution is measured at a step of $\Delta\sigma=0.011\,89$ (1 mm) from the transducer surface to the normalized distance of $\sigma=0.76$ (64 mm) along the transducer axis. The lateral beam patterns are measured at a step of $\Delta r=0.05$ (0.2 mm) from $r=-5$ (−20 mm) to $r=5$ (20 mm) at the normalized axial distance of $\sigma=0.6$. The transmission waveform is acquired at each scanning point. The amplitudes of the fundamental to the fifth harmonics for each waveform are obtained using a FFT software and the amplitude of the superharmonic component is calculated from the summation of the absolute amplitudes of the third, fourth, and fifth harmonics. Figures 2–4 show the measured axial and lateral field distributions of the superharmonic and the second harmonic components. It is found that the measured results (∇ and $\bigcirc$) coincided well with the corresponding theoretical predictions (solid and dashed lines).

The measured axial distributions of the second harmonic and superharmonic components are normalized by the maximum amplitude of the second harmonic component in Figs. 2 and 3. As shown in Fig. 2, at a source sound pressure of 0.5 MPa, the amplitude of the superharmonic component is lower than that of the second harmonic for the normalized axial distance of $\sigma\leqslant 0.8$ (67.4 mm). As shown in Fig. 3, when the source sound pressure is equal to 1 MPa, the amplitude of the superharmonic component begins to exceed that of the second harmonic for $\sigma\geqslant 0.4$ (33.7 mm), reaches a maximum value at $\sigma=0.54$ (45.5 mm), and then decreases due to the high attenuation, but it is still higher than that of the second harmonic. Therefore the source pressure amplitude and the sound propagation distance are two important variables in obtaining a superharmonic component with higher energy over the second harmonic.

Figure 4 compares the lateral beam patterns of the superharmonic and the second harmonic components from the normalized lateral distance of $r=-2$ (−8 mm) to $r=2$ (8 mm), which are normalized by their corresponding maximum amplitudes. It is observed that the superharmonic component has a narrower main lobe and its 3-dB beam width is about 77% of that of the second harmonic. The narrower beam width of the superharmonic component results in an improved lateral resolution. Compared with those of the second harmonic, the side lobe levels of the superharmonic component are reduced by about 10 dB because the energy is concentrated in the main lobe. This will help reducing artifacts and aberrations in medical imaging.[19]

To explore the optimization of the superharmonic generation, Fig. 6 displays the relationship between the source pressure amplitude and the received pressure amplitudes of the superharmonic and the second harmonic components at the normalized axial distance 0.6. The received sound pressure amplitude of the superharmonic component begins to exceed that of the second harmonic at a source sound pressure 0.6 MPa and is about 4.5 dB higher at 1 MPa. Hence, a source sound pressure higher than 0.6 MPa should be applied in the superharmonic generation if the imaging is performed at a normalized axial distance 0.6.

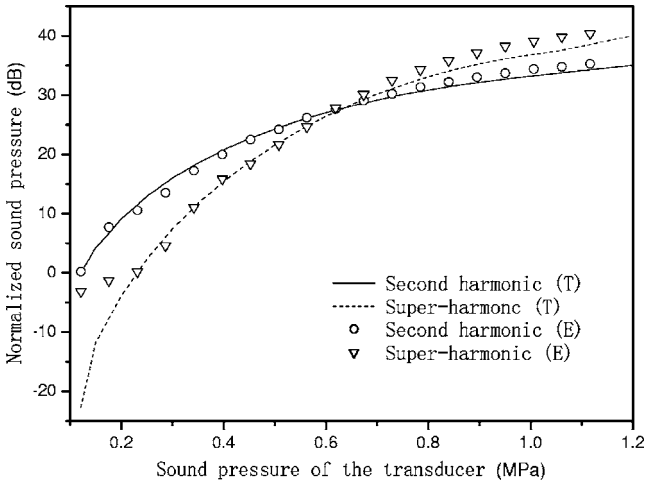


FIG. 6. Sound pressure dependencies of the normalized sound pressures of the superharmonic and the second harmonic components at $\sigma=0.6$ (T—theory, E—experiment).

A tissue superharmonic imaging in the transmission mode is also carried out for porcine liver and fatty tissues *in vitro*. The tissues are obtained from a slaughterhouse and stored in 0.9% saline solution. Imaging experiments are finished within 2 h at 23 °C. To produce a two-dimensional image, the sample is mounted on the scanning system between the transmitter and the needle hydrophone, which are coaxially fixed on the bracket. The sample moves along x and y directions within a 20×20 mm^2 are with a step of 0.4 mm. At each point, the following process for the superharmonic imaging is carried out: (1) the received signal is collected digitally by the computer; (2) the amplitudes of the third, fourth, and fifth harmonics of the received waveform are obtained by means of the FFT software; and (3) the superharmonic component is derived by the linear summation of the third, fourth, and fifth harmonics. Finally, a 50 $\times$50 pixels image is constructed.

Figure 7 shows the two sample models with different tissue combinations that are studied. Figure 7(a) shows the

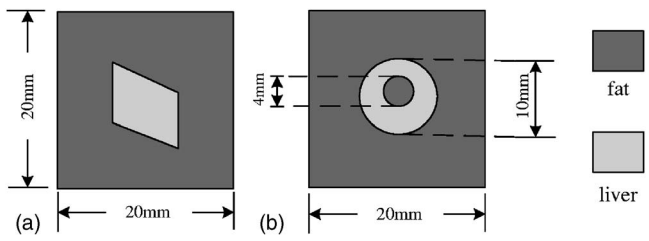


FIG. 7. Sample models with different tissue combinations.

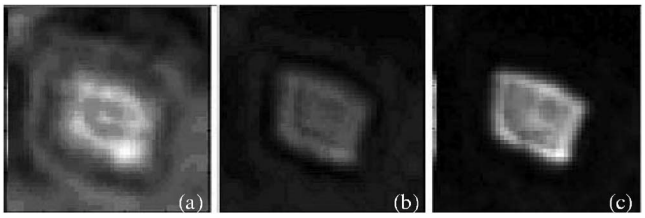


FIG. 8. Reconstructed images of sample model A: (a) fundamental frequency image, (b) second harmonic image, and (c) superharmonic image.

cross section of model A, which consists of an internal portion of a quadrangular porcine liver tissue and an external portion of a porcine fatty tissue. Figure 7(b) shows a three-layer model B with a porcine fatty tissue located at both the external ring and the internal circle, and a porcine liver tissue located at the middle ring.

The corresponding fundamental, second harmonic, and superharmonic images are presented in Figs. 8 and 9, in which the different gray scales correspond to different biological tissues due to the differences in the amplitudes of selected component of the received signal. Because the beam pattern at the superharmonic frequency is improved (3-dB beam width at the fundamental frequency is 1.3 times greater than that at the second-order harmonic component and 1.7 times greater than that at the superharmonic frequency), the superharmonic images exhibit better spatial resolution than the fundamental or the second harmonic images. In the fundamental frequency images Figs. 8(a) and 9(a), it is observed that the dimensions of the liver tissues are about 30% larger than the actual dimensions and the internal fatty tissue is almost invisible in Fig. 8(a). The speckle noise is also obvious due to the artifacts from the side lobes and the sound diffractions. It is also observed that the edges of different tissues are vague and there exists an intergradation of about 2 mm for the poor spatial resolution at the fundamental frequency. With an improvement in spatial resolution, the second harmonic images demonstrate recognizable tissue combinations as shown in Figs. 8(b) and 9(b). The edges of the liver and fatty tissues are confirmed, but the lower amplitude of the second harmonic results in less contrast of the images. In the superharmonic images as shown in Figs. 8(c) and 9(c), both the dimensions and positions of tissues are confirmed by the sharp borders and the contrast of the images is improved. In addition, the speckle noise in the fatty tissue is reduced.

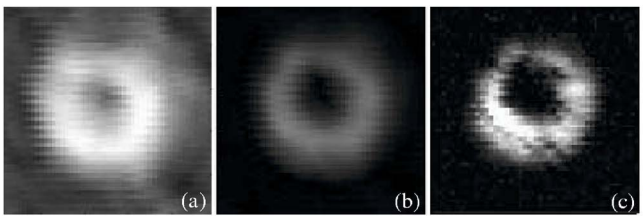


FIG. 9. Reconstructed images of sample model B: (a) fundamental frequency image, (b) second harmonic image, and (c) superharmonic image.

IV. CONCLUSION

Based on the theory of the finite amplitude sound wave, the axial and lateral distributions of the superharmonic component radiated from a planar piston source are theoretically simulated and experimentally measured in the study. Their dependencies on the source sound pressure amplitude and the propagation distance are discussed and compared with those of the second harmonic. When the source sound pressure is higher than a threshold at a propagation distance, the superharmonic component is shown to be superior to the second harmonic due to increased energy and reduced beam width and side lobe levels. It results in improved clarity and contrast for imaging. Furthermore, two-dimensional imaging experiments are performed using liver and fatty tissue samples *in vitro* in transmission mode and compared to the images obtained using the fundamental and the second harmonic components. Although the clinical applicability of this work still needs to be assessed, these results indicate that the superharmonic image quality is better than that of using the fundamental and second harmonic components. Also the use of a focused transducer can obtain a narrower beam pattern, better spatial resolution, and higher energy in the focusing region than the planar transducer. Further study on the superharmonic imaging technique in reflection mode can be carried out and will have a favorable prospect in clinical ultrasound imaging.

ACKNOWLEDGMENTS

The authors wish to thank Professor F. Dunn for his valuable comments in preparing the final version of this paper. This work is supported by the National Natural Science Foundation of China and TWAS (No. 03-390).

[1]F. A. Duck, "Nonlinear acoustics in diagnostic ultrasound," Ultrasound Med. Biol. **28**, 1–18 (2002).
[2]P. N. Burns, D. H. Simpson, and M. A. Lverkiou, "Nonlinear imaging," Ultrasound Med. Biol. **26**, s19–s22 (2000).
[3]N. de Jong, R. Cornet, and C. T. Lancee, "Higher harmonics of vibrating gas filled microspheres. Part one: Simulations," Ultrasonics **32**, 447–453 (1994).
[4]B. B. Goldberg, J. B. Liu, and F. Forsberg, "Ultrasound contrast agents: A review," Ultrasound Med. Biol. **20**, 319–333 (1994).
[5]N. de Jong, A. Bouakaz, and F. J. T. Cate, "Contrast harmonic imaging," Ultrasonics **40**, 567–573 (2002).
[6]M. G. J. Borsboom, C. T. Chin, and N. de Jong, "Nonlinear coded excitation method for ultrasound contrast imaging," Ultrasound Med. Biol. **29**, 285–292 (2003).
[7]F. Forsberg, W. T. Shi, and B. B. Goldberg, "Subharmonic imaging of contrast agents," Ultrasonics **38**, 93–98 (2000).
[8]F. Tranquart, N. Grenier, V. Eder, and L. Pourcelot, "Clinical use of ultrasound tissue harmonic imaging," Ultrasound Med. Biol. **6**, 889–894 (1999).
[9]I. Puls, D. Berg, M. Maurer, M. Schliesser, G. Hetzel, and G. Becker, "Transcranial sonography of the brain parenchyma: comparison of B-mode imaging and tissue harmonic imaging," Ultrasound Med. Biol. **26**, 189–194 (2000).
[10]S. Tanaka, O. Oshikawa, T. Sasaki, T. Ioka, and H. Tsukuma, "Evaluation of tissue harmonic imaging for the diagnosis of focal liver lesions," Ultrasound Med. Biol. **26**, 183–187 (2000).
[11]T. G. Muir and E. L. Carstensen, "Prediction of nonlinear acoustic effects at biomedical frequencies and intensities," Ultrasound Med. Biol. **4**, 345–357 (1980).
[12]E. L. Carstensen, W. K. Law, N. D. Mckay, and T. G. Muir, "Demonstration of nonlinear acoustical effects at biological frequencies and intensities," Ultrasound Med. Biol. **6**, 359–368 (1980).

[13]T. Christopher, "Finite Amplitude Distortion-Based Inhomogeneous Pulse Echo Ultrasonic Imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 125–139 (1997).
[14]T. Christopher, "Experimental investigation of finite amplitude distortion-based, second harmonic pulse echo ultrasonic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**(1), 158–162 (1998).
[15]F. Dunn, W. K. Law, and L. A. Frizzell, "Nonlinear ultrasonic wave propagation in biological materials," IEEE Ultrasonics Symposium 527–532 (1981).
[16]T. G. Muir, "Nonlinear effects in acoustic imaging," Acoust. Imaging **9**, 93–109 (1980).
[17]C. M. Sehgal, G. M. Brown, R. C. Bahn, and J. F. Greenleaf, "Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers," Ultrasound Med. Biol. **12**, 865–874 (1986).
[18]D. Zhang, X. F. Gong, and S. G. Ye, "Acoustic nonlinearity tomography for biological specimens via measurements of the second harmonic wave," J. Acoust. Soc. Am. **99**, 2397–2402 (1996).
[19]X. F. Gong, D. Zhang, J. H. Liu, H. L. Wang, Y. S. Yan, and X. C. Xu, "Study of acoustic nonlinearity parameter imaging methods in reflection mode for biological tissues," J. Acoust. Soc. Am. **116**, 1819–1825 (2004).
[20]M. Fatemi and J. F. Greenleaf, "Real-time assessment of the parameter of nonlinearity in tissue using nonlinear shadowing," Ultrasound Med. Biol. **22**, 1215–1228 (1996).
[21]D. Rugar, "Resolution beyond the diffraction limit in the acoustic microscope: A nonlinear effect," J. Acoust. Soc. Am. **56**, 1338–1346 (1984).
[22]B. Ward, A. C. Baker, and V. F. Humphrey, "Nonlinear propagation applied to the improvement of resolution in diagnostic medical ultrasound," J. Acoust. Soc. Am. **101**, 143–154 (1997).
[23]C. C. Shen and P. C. Li, "Harmonic leakage and image quality degradation in tissue harmonic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 728–736 (2001).
[24]P. Jiang, Z. Mao, and J. C. Lazenby, "A new tissue harmonic imaging scheme with better fundamental frequency cancellation and higher signal-to-noise ratio," IEEE Ultrasonics Symposium 1589–1595 (1998).
[25]D. Hope Simpson, C. T. Chin, and P. N. Burns, "Pulse inversion Doppler: A new method for detecting nonlinear echoes from microbubble contrast agents," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 372–382 (1999).
[26]C. Hohl, T. Schmidt, P. Haage, D. Honnef, G. Blaum, and R. W. Guenther, "Phase-inversion tissue harmonic imaging compared with conventional B-mode ultrasound in the evaluation of pancreatic lesions," Eur. Radiol. **14**(6), 1109–1117 (2004).
[27]C. C. Shen and P. C. Li, "Motion artifacts of phase inversion-based tissue harmonic imaging," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**(9), 1203–1211 (2002).
[28]A. Bouakaz and N. de Jong, "Native Tissue Imaging at Superharmonic Frequencies," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 125–139 (2003).
[29]A. Bouakaz, B. J. Krenning, W. B. Vletter, F. J. Ten Cate, and N. de Jong, "Contrast superharmonic imaging: A feasibility study," Ultrasound Med. Biol. **29**, 547–553 (2003).
[30]E. Zabolotskaya and R. Khokhlov, "Quasi-plane waves in the nonlinear acoustics of confined beams," Sov. Phys. Acoust. **15**, 35–40 (1969).
[31]V. Kuznetsov, "Equation of nonlinear acoustics," Sov. Phys. Acoust. **16**, 467–470 (1970).
[32]S. I. Aanonsen, T. Barkve, J. Naze Tjotta, and S. Tjotta, "Distortion and harmonic generation in the nearfield of a finite amplitude sound beam," J. Acoust. Soc. Am. **75**, 749–768 (1984).
[33]B. Ystad and J. Bernsten, "Numerical solution of the KZK equation for focusing sources," Acta Acust. **3**, 323–330 (1995).
[34]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**, 95–108 (2001).
[35]T. Kamakura, T. Ishiwata, and K. Matsuda, "Model equation for strongly focused finite-amplitude sound beams," J. Acoust. Soc. Am. **107**, 3035–3046 (2000).
[36]R. T. Beyer, "Parameter of nonlinearity in fluids," J. Acoust. Soc. Am. **32**, 719–721 (1960).

Perception of complex sounds in budgerigars (*Melopsittacus undulatus*) with temporary hearing loss

Robert J. Dooling[a)]
Department of Psychology, University of Maryland, College Park, Maryland 20742

Brenda M. Ryals
Department of Communication Sciences and Disorders, James Madison University, Harrisonburg, Virginia 22807

Micheal L. Dent[b)] and Tracy L. Reid
Department of Psychology, University of Maryland, College Park, Maryland 20742

(Received 14 July 2005; revised 11 January 2006; accepted 12 January 2006)

Songbirds and parrots deafened as nestlings fail to develop normal vocalizations, while birds deafened as adults show a gradual deterioration in the quality and precision of vocal production. Beyond this, little is known about the effect of hearing loss on the perception of vocalizations. Here, we induced temporary hearing loss in budgerigars with kanamycin and tested several aspects of the hearing, including the perception of complex, species-specific vocalizations. The ability of these birds to discriminate among acoustically distinct vocalizations was not impaired but the ability to make fine-grain discriminations among acoustically similar vocalizations was affected, even weeks after the basilar papilla had been repopulated with new hair cells. Interestingly, these birds were initially unable to recognize previously familiar contact calls in a classification task—suggesting that previously familiar vocalizations sounded unfamiliar with new hair cells. Eventually, in spite of slightly elevated absolute thresholds, the performance of birds on discrimination and perceptual recognition of vocalizations tasks returned to original levels. Thus, even though vocalizations may initially sound different with new hair cells, there are only minimal long-term effects of temporary hearing loss on auditory perception, recognition of species-specific vocalizations, or other aspects of acoustic communication in these birds.
[DOI: 10.1121/1.2171839]

PACS number(s): 43.80.Nd, 43.80.Lb, 43.66.Gf [JAS] Pages: 2524–2532

I. INTRODUCTION

It is well known that songbirds and parrots rely on hearing to develop and maintain a normal vocal repertoire (see review in Kroodsma and Miller, 1996). Budgerigars (*Melopsittacus undulatus*) are particularly interesting because they learn new vocalizations throughout life, especially in response to changes in their social milieu (see review in Dooling, 1986; Dooling *et al.*, 1987; Farabaugh *et al.*, 1994; Farabaugh and Dooling, 1996; Brittan-Powell *et al.*, 1997; Hile *et al.*, 2000). In the most extreme case of experimental auditory manipulation in budgerigars, permanent deafening by cochlear removal during development (Dooling *et al.*, 1987; Heaton and Brauth, 1999) or in adulthood (Heaton *et al.*, 1999) results in a dramatically impoverished vocal repertoire. In a more subtle auditory manipulation, recent experiments show that the role of auditory feedback in vocal production can also be more immediate. Budgerigars trained to produce particular vocalizations under controlled experimental conditions exhibit the Lombard effect—an increase in vocal intensity in response to an increase in ambient noise. This is suggestive of a sensitive, real-time monitoring of vocal output as occurs in humans (Manabe *et al.*, 1998).

The importance of hearing for the maintenance of a normal adult vocal repertoire in these birds raises the question of the effects of hearing loss (defined by an increase in absolute thresholds) on the discrimination and perception of vocalizations. To this end, we examine the effect of severe, but temporary, hearing loss on the discrimination and recognition of species-specific vocalizations in budgerigars. Numerous reports have shown that following auditory trauma, birds can regenerate hair cells which lead to varying degrees of physiological and behavioral recovery of hearing (e.g., Corwin and Cotanche, 1988; Ryals and Rubel, 1988; Hashino and Sokabe, 1989; Tucci and Rubel, 1990; Hashino *et al.*, 1991; Lippe *et al.*, 1991; Hashino *et al.*, 1992; Marean *et al.*, 1993; Niemiec *et al.*, 1994; Saunders *et al.*, 1995, 1996; Dooling *et al.*, 1997; Marean *et al.*, 1998; Ryals *et al.*, 1999). Recent reports by Rubel and his colleagues show that the Bengalese Finch, an age-dependent vocal learner, demonstrates both auditory and vocal recovery before hair cells have fully regenerated (Woolley and Rubel, 1999; Woolley *et al.*, 2001; Bermingham-McDonogh and Rubel, 2003). Except for a brief and preliminary earlier summary (Dooling *et al.*, 1997), there are no detailed reports on the effect of temporary hearing loss on the discrimination or recognition of vocalizations in birds that exhibit vocal learning throughout adulthood.

[a)]Corresponding author. Electronic mail: dooling@psyc.umd.edu
[b)]Current address: Department of Psychology, University at Buffalo SUNY, Buffalo, NY 14260.

 0001-4966/2006/119(4)/2524/9/$22.50

In the following experiments, we tested whether hearing, auditory discrimination, and the perception and recognition of complex vocalizations were affected by temporary hearing loss. In our tests we measured absolute thresholds, intensity difference limens (IDLs), frequency difference limens (FDLs), the discrimination between natural and synthetic contact calls, and the identification of contact calls in budgerigars before and after injections of kanamycin. Histological evaluations were also carried out to verify the time course of the loss and regeneration of hair cells. Because budgerigars learn new vocalizations throughout life and rely on hearing to maintain a normal adult vocal repertoire, they may provide an important animal model for some aspects of human hearing and speech. It is well known that hearing impairment in humans, for instance, can have a profound effect on speech perception, and the quality of speech production and acoustic communication.

II. GENERAL METHODS

A. Subjects

Fifteen budgerigars were used in the behavioral experiments. These, as well as an additional 14 birds, were also used for histology. Nonbehavioral birds were sacrificed at intervals sufficient to determine the extent of hair cell loss and recovery over time after cessation of drug injection (at 1 day postkanamycin, $n=3$; at 9–14 days postkanamycin, $n=4$; at 30 days postkanamycin, $n=3$; and control birds, $n=4$). The birds were individually housed in an aviary at the University of Maryland and kept on a normal photoperiod correlated with the season. During behavioral testing, the birds were food deprived to 85–90% of their free-feeding weights. Animal care and housing met all standards of the Animal Care and Use Committee at the University of Maryland.

B. Drug injections

The birds were injected with kanamycin for 8–10 days. The single daily intramuscular injections were 100 mg/kg the first day and 200 mg/kg each day after that. The injections were in the pectoris muscle, and the injection site changed from day to day. The birds were weighed daily and injected at approximately the same time every day using a Hamilton (microliter No. 710) 0.1 ml glass syringe with a 30G1/2 needle. Psychophysical tests were carried out before, in some cases during, and after the administration of kanamycin. Long-term psychophysical tests were conducted at approximately biweekly intervals for the next 24 weeks following the end of injections.

C. Behavioral testing apparatus

The testing apparatus has been described in detail previously (Okanoya and Dooling, 1987; Dooling and Okanoya, 1995). A custom-made operant chamber with a food hopper and two response keys consisting of light-emitting diodes and microswitches was used. A standard pigeon grain hopper delivered food with response keys mounted just above the hopper opening. The stimuli were delivered from a loudspeaker mounted above the test cage (Realistic Soft Dome midrange speaker, Model 40-1281A). The entire operant apparatus was suspended in a small animal sound isolation chamber (Industrial Acoustics, Model IAC-3). A microcomputer controlled all experimental events including stimulus delivery and reinforcement contingencies. The behavior of the animals during test sessions was monitored by a video camera.

D. Auditory stimuli and stimulus calibration

Pure tone stimuli ranging from 500 Hz to 5700 Hz (300 ms duration, 5 ms rise/fall times) were used in Experiments 1 and 2. Species-specific contact calls were used in Experiments 2 and 3. These contact calls were recorded from five budgerigars using a Simul-Sync 4-track Teac tape recorder (Model A-3440). To elicit contact calls, birds were isolated individually in small sound-isolation chambers (Industrial Acoustics Company, Model AC1), each fitted with an omnidirectional Realistic Electret Microphone (Model 33-1063) connected to separate tracks of the tape recorder. After at least 30 min of isolation, the doors to the chambers were opened slightly so the birds could hear the faint calls of the other birds. The computer software package SIGNAL (Beeman, 1992) was used to create synthetic versions of the recorded contact calls. The calls were analyzed using overlapping, serial fast Fourier transformations and then synthesized by reconstructing the peak frequency trace and amplitude contour throughout the call (see, for example, Dooling and Okanoya, 1995).

Stimulus calibration was performed using a General Radio Model 1982 sound-level meter with octave band filters. Sound pressure levels (SPLs) in the test chamber were measured by placing the microphone in front of the keys in the position normally occupied by the bird's head during testing. The intensities of test tones and calls were measured several times during the testing phase to ensure they remained constant.

E. Training and testing procedures

The training and testing procedures have been described in detail previously (Okanoya and Dooling, 1991; Dooling and Okanoya, 1995). Briefly, the birds were trained by a standard operant autoshaping program to peck at the left microswitch key (observation key) until a tone was presented. Then, a peck to the right microswitch key (report key) within 2 s of stimulus presentation produced food reinforcement. A failure to peck the report key was recorded as a miss, and a new trial was started. 30% of all trials were sham trials in which no tone was presented. A peck to the report key during a sham trial was recorded as a false alarm. Sessions with a false alarm rate of 16% or higher were discarded.

The birds were tested in two daily sessions typically consisting of about 100 trials each. Stimuli were presented according to the method of constant stimuli and the targets were chosen so that the lowest target tested was below the bird's suspected threshold while the highest targets were well

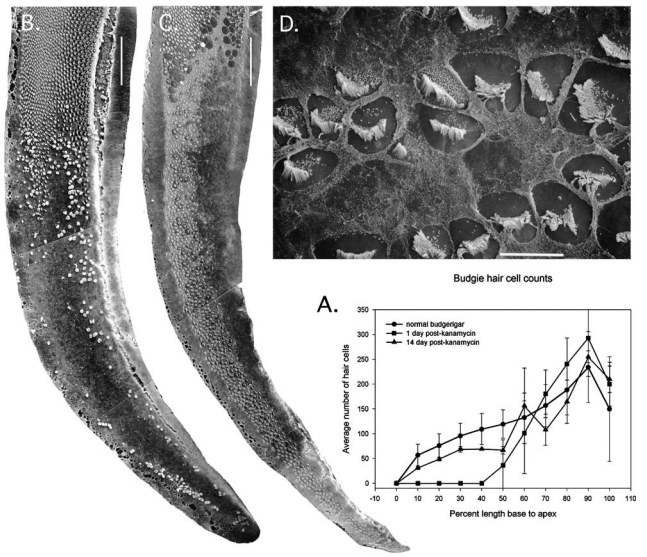


FIG. 1. (A) Average number of hair cells counted from 700× SEM montages at regular intervals from base to apex. Hair cells are completely absent from the basal 40% of the papilla one day after the cessation of kanamycin injections in all birds examined ($n=3$). Hair cell number has increased to nearly normal levels within 14 days of kanamycin cessation ($n=4$). Error bars represent standard deviations. (B) SEM photomicrograph of the basal half of the basilar papilla in a bird one day after the cessation of eight days of kanamycin injections (bar=100μ). Swollen hair cells, ejected from the epithelia are seen as large white blebs on the surface of the papilla. (C) SEM photomicrograph of the basal half of the basilar papilla in a bird 14 days after cessation of 8 days of kanamycin injections (bar=100μ). (D) Higher magnification (bar=10μ) of area on the basilar papilla of a bird 14 days after kanamycin injections. Areas with small microvilli are the swollen surfaces of supporting cells.

above threshold. Threshold was defined as the target level that the bird detected 50% of the time (Dooling and Okanoya, 1995).

III. ANATOMICAL EXPERIMENTS

A. Experiment 1—Histology

1. Methods

Basilar papillae were fixed (3.5% glutaraldehyde followed by 1% osmium tetroxide) and dehydrated to 70% ethanol. Then, the tectorial membrane was removed, dehydrated to 100% and critical point dried in CO_2. Papillae were mounted on aluminum stubs, sputter coated with gold, and viewed at 15 kV using either an Amray 1820 or a Joel 820 scanning electron microscope (SEM). In order to quantify the degree and location of hair cell loss, hair cell counts were made from SEM montages (700×). Montages were divided into ten equidistant intervals (approximately 10% increments of length) from basal to apical tip. Each interval was 60 μm wide (along the basal to apical dimension) and extended from neural to abneural edge; hair cells were counted in each of these ten intervals. This method for hair cell counts from SEM montages is similar to that used by other investigators (see, for example, Gleich and Manley, 1988).

2. Results

Figure 1 shows examples of papillae one day [Fig. 1(B)] and 14 days [Fig. 1(C)] after an 8-day course of kanamycin injections. Hair cells are completely absent in the basal 40% of the papilla immediately after kanamycin injections cease, but return to nearly normal numbers within 14 days [Fig. 1(A)]. While there are no published place-frequency maps for the budgerigar, estimates from other avian species predict that this hair cell loss likely corresponds to hearing for frequencies above 500–1000 Hz (Gleich and Manley, 2000). Even though the hair cell number has nearly returned to normal within 14 days, swelling of support cells continues and small regenerating hair cells continue to be present [see Fig. 1(D)]. Our histological results are comparable to those reported by Hashino *et al.* (1992) and Hashino and Sokabe (1989). The hair cell number continues to increase, but at a slower rate, over the next several weeks. Figure 2(A) shows the number of hair cells present in several of the birds tested behaviorally in the present experiments and sacrificed at either 2–3 months postkanamycin ($n=6$) or 7 months postkanamycin ($n=4$). At 2–3 months postkanamycin treatment, even though hair cell number is within one standard deviation of normal, some hair cell abnormalities (multiple and/or abnormal stereovilli bundles, and abnormal stereovilli bundle orientation) and regenerating hair cells are present [see Fig. 2(B)]. After 7 month survival, no immature hair cells were seen, but stereovilli bundle orientation remained irregular in the area where hair cells had been completely lost [see Fig. 2(C)]. These abnormalities are similar to those reported in other avian hair cell stereovilli after ototoxic drug administration (Cotanche, 1999) and remain in the basal portion of the papilla, even after 7 months.

IV. PSYCHOPHYSICAL EXPERIMENTS

A. Experiment 1 - Absolute thresholds

1. Methods

To track hearing loss, three birds were tested on the detection of pure tones in quiet. Within a ten-trial block, the test tone was presented at seven different predetermined levels using a step size of 3 dB. In addition, three sham trials were included in each block of ten trials. The levels were preselected so that only the quietest one or two tones could not be heard by the birds, whereas the loudest tones were always above threshold. In all, thresholds were measured at six tone frequencies (at 0.5, 1.0, 2.0, 2.86, 4.0, and 5.7 kHz) before and after kanamycin injections. Each bird was run on at least 100 trials weekly at each of the six frequencies following the injections for up to 24 weeks total.

2. Results

As is typical of birds (Marean *et al.*, 1998; Woolley *et al.*, 2001), and as summarized earlier for budgerigars (Dooling *et al.*, 1997), absolute thresholds were elevated at all frequencies following kanamycin injections, but the hearing loss was especially severe at high frequencies as shown in Fig. 3. Within a few weeks of kanamycin treatment, threshold shift was greatest (50–60 dB) at frequencies above 2 kHz and less (10–30 dB) at frequencies below 2 kHz. By 20 weeks following injections, absolute thresholds improved to asymptotic levels of within 10–15 dB of normal at low frequencies at about 20–30 dB at higher frequencies. Recov-

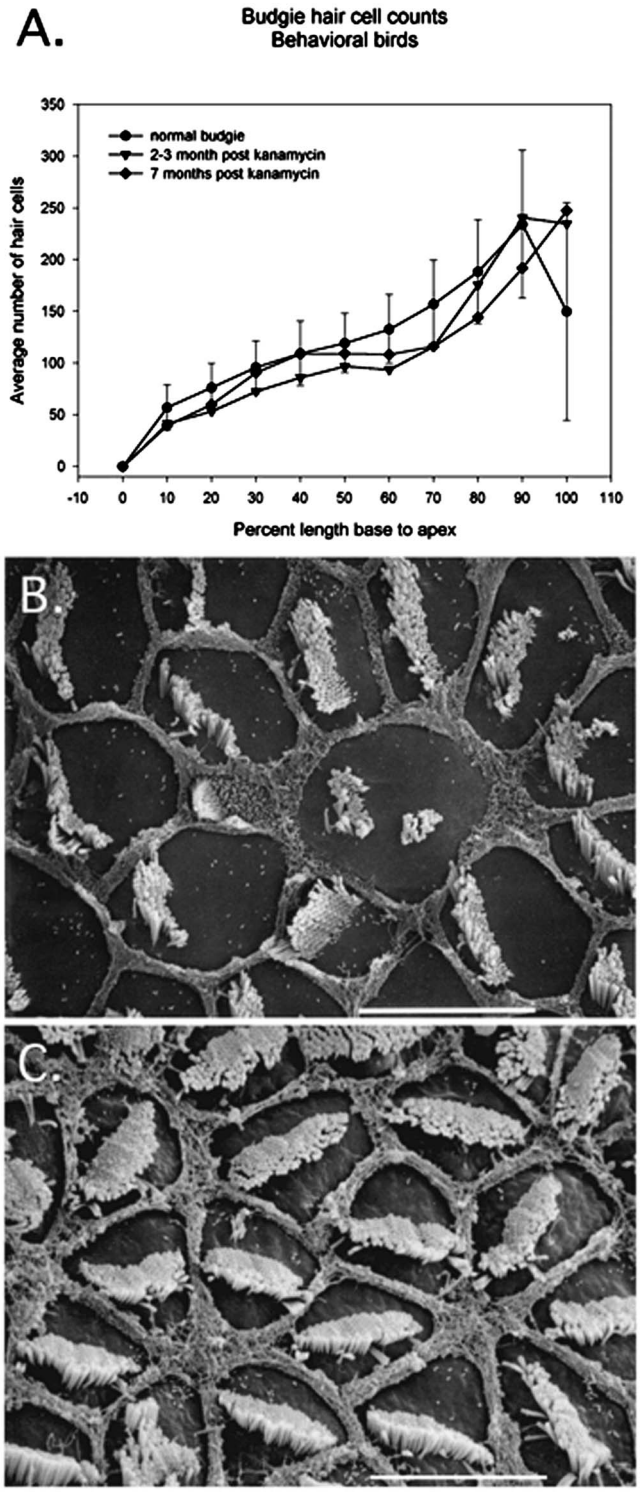


FIG. 2. (A) Average hair cell counts in birds tested behaviorally and allowed to survive either 2–3 months ($n=6$) or 7 months ($n=4$) after kanamycin injections. Hair cell number is within one standard deviation of normal for each group. (B) SEM photomicrograph of hair cell abnormalities (cell with two stereovilli bundles) and regenerating hair cell in a bird sacrificed 2–3 months after kanamycin injections (bar=10μ). (C) SEM photomicrograph of hair cells with abnormal stereovilli bundle polarity 7 months after kanamycin injections. These disordered stereovilli orientation patterns are similar to those previously reported in other avian species (Cotanche, 1999).

ery proceeded most quickly at low frequencies and slower at high frequencies but a permanent threshold shift was evident at all frequencies.

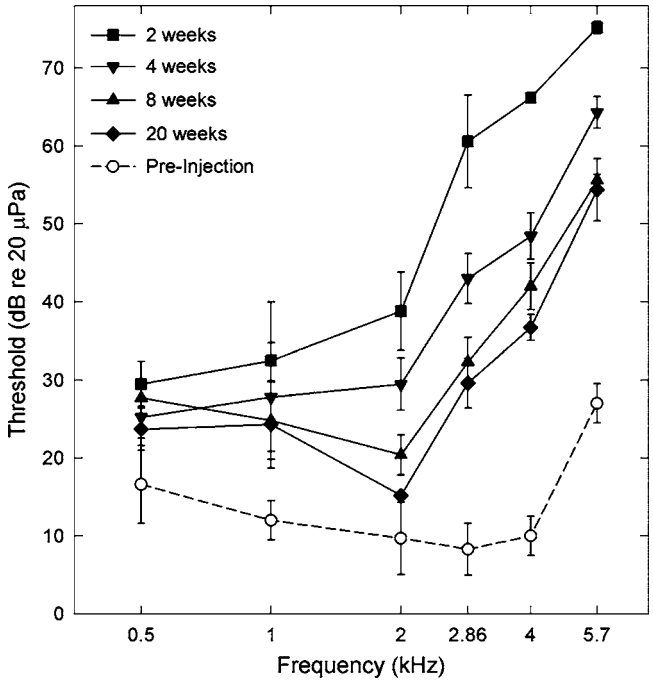


FIG. 3. Hearing loss audiograms for three budgerigars prior to injections (open circles) and 2 weeks (closed squares), 4 weeks (closed upside-down triangles), 8 weeks (closed triangles), and 20 weeks (closed diamonds) following kamanycin treatment. Error bars represent between subject standard errors.

B. Experiment 2—Discrimination tests: Intensity and frequency difference limens and the discrimination of contact calls

1. Methods

For these experiments, three different birds were each tested on three different discrimination tasks. These tasks included IDLs, FDLs, and the discrimination of species-specific contact calls. For all tests, the birds were required to discriminate a change in a repeating background of sound. For the measurement of difference limens, the repeating background consisted of either 1.0 or 2.86 kHz pure tones played at a rate of 2/s at a level of 65 dB SPL [measured as root-mean square (rms) on the fast scale of the sound level meter]. For the IDLs, test tones were presented at seven different predetermined levels increasing in step sizes of 3 dB (three sham trials were included in each block) within a ten-trial block. For the FDLs, the test tones were presented at seven different predetermined frequencies increasing in step sizes of 5 Hz (three sham trials were also included in each block of ten trials). For each bird, a minimum of a 100 trials were collected on IDLs and FDLs immediately before kanamycin injections and approximately every three weeks following injections for a total of 19 weeks.

Another experiment was conducted to determine the relation between the detection and discrimination of standard pure tone stimuli and the discrimination of vocalizations that these birds use in communication. A set of vocalizations were prepared consisting of five different natural species-specific contact calls and their synthetic analogs as shown in Fig. 4(A) (see Dooling and Okanoya, 1995). The birds were tested on a pair-wise discrimination of all possible combina-

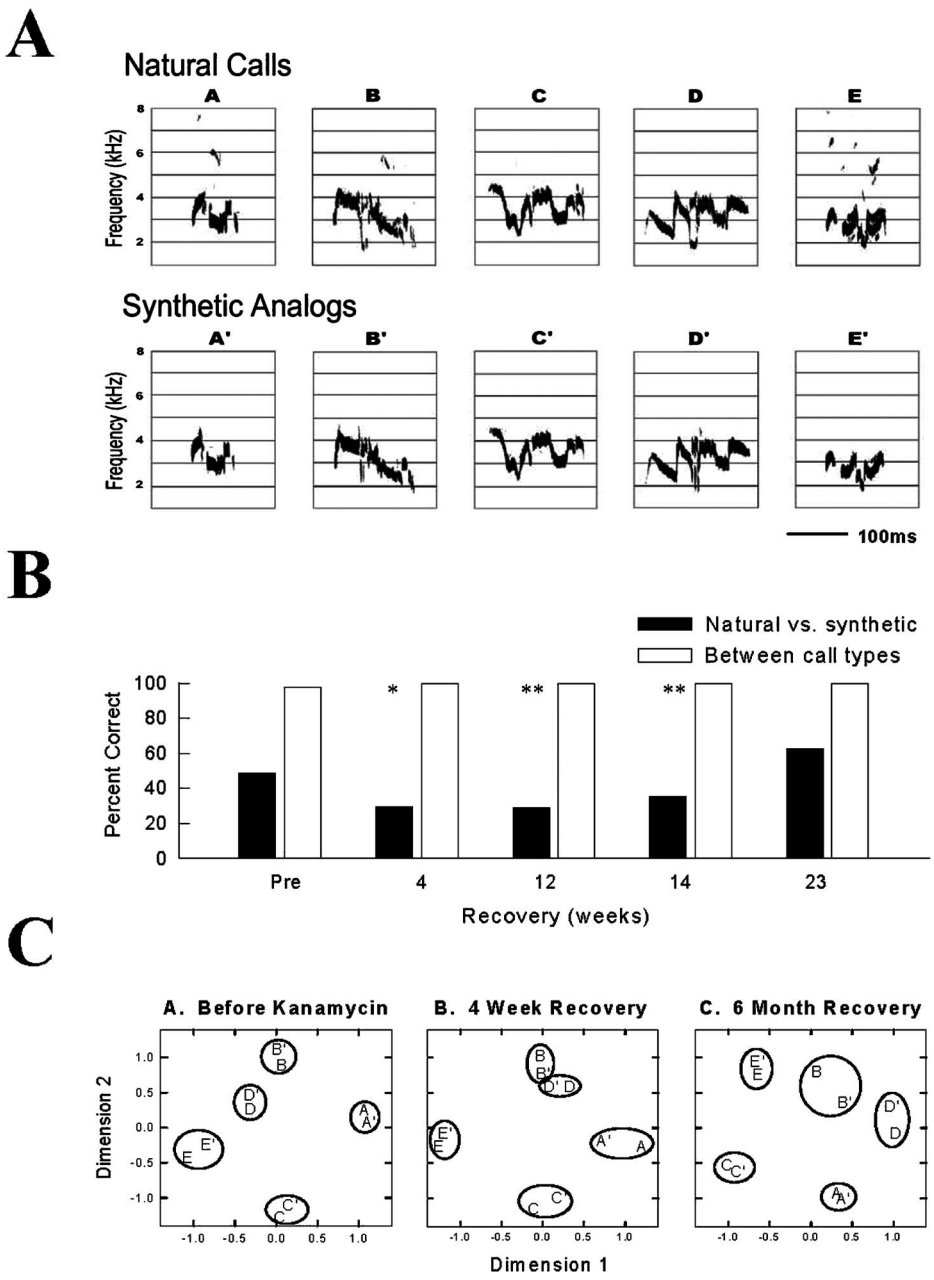


FIG. 4. (A) Sonograms of contact calls from five different birds and their synthetic analogs plotted as frequency by time (taken from Dooling and Okanoya, 1995). (B) The average percent correct for three budgerigars discriminating among the five natural contact calls and their synthetic analogs before treatment with kanamycin and at 4, 12, 14, and 23 weeks after injections. The stars represent significantly different results from the preinjection condition (single star: $p=.05$, double star: $p<.05$). (C) Two-dimensional perceptual maps by MDS from three budgerigars tested before kanamycin treatment (a), after 4 weeks of recovery (b), and after 6 months of recovery (c).

tions of the ten calls shown in Fig. 4(A). This set of calls was designed to allow for both fairly easy discriminations (e.g., between call types as between A and B) and more difficult discriminations (e.g., as between A and A′)—a discrimination that humans find impossible and birds find very difficult (Dooling *et al.*, 1997). During testing, the calls were played out at a rms level of 65 dB SPL measured at the birds' ears. Sessions consisted of 100 trials with one call selected as the repeating background and the other calls selected as targets. This procedure continued until all possible combinations of calls were tested in random order. Each bird was tested on the entire ten-call set both before and approximately every 4 weeks following injections up to about 24 weeks following cessation of kanamycin injections.

2. Results

Though kanamycin treated birds have a substantial hearing loss remaining at 4 weeks following kanamycin injections (as shown in Experiment 1, Fig. 3), FDLs and IDLs are not significantly different from prekanamycin thresholds (see Fig. 5). For all three budgerigars in Experiment 2, FDLs appeared to be only slightly elevated at both frequencies at 4–6 weeks following the cessation of injections. A repeated measures analysis of variance (ANOVA) for each frequency before and at three time periods following kanamycin injections showed that FDLs were not significantly different at any of the subsequent time periods up to 19 weeks at 1.0 kHz $[F(2,6)=2.15, p>0.05]$ or 2.86 kHz $[F(2,6)=0.27, p>0.05]$, IDLs were also only slightly elevated at both frequencies at 4–6 weeks after injections but a repeated measures ANOVA for each frequency showed that IDLs were not significantly different at any of the time periods either at 1.0 kHz $[F(2,6)=1.47, p>0.05]$ or 2.86 kHz $[F(2,6)=1.00, p>0.05]$.

In discriminating among the natural and synthetic contact calls before and after kanamycin injections, all birds

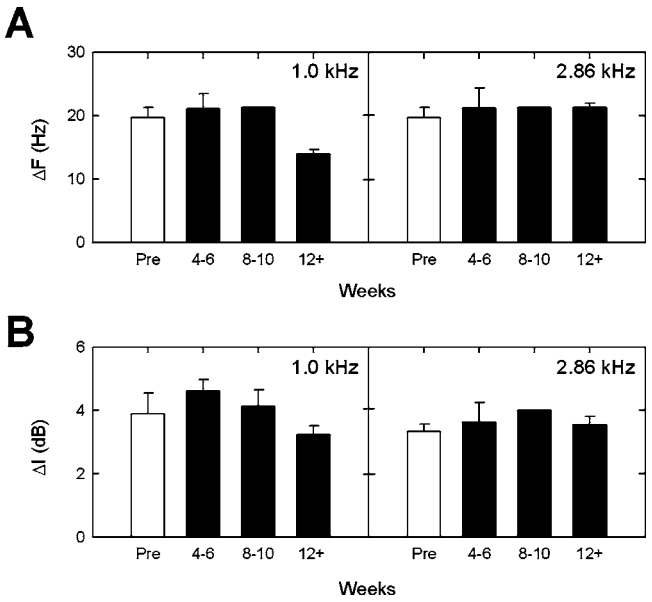


FIG. 5. (A) FDLs and (B) IDLs for 1.0 and 2.86 kHz pure tones before, and at three time periods following injections. There is no negative effect of kanamycin KM treatment on these difference limens and there is even a suggestion of improvement at both frequency and intensity difference limens at 1.0 kHz after 3 months. Error bars represent standard errors.

showed similar patterns. The more difficult acoustic discriminations in the test set as between natural calls and their synthetic analogues were significantly affected for up to 14 weeks. Only at 23 weeks did discrimination return to pre-kanamycin levels [Fig. 4(B)]. Easy disciminations, as between contact calls from different birds, were discriminated at nearly 100% before kanamycin injections and were hardly affected at four weeks following kanamycin treatment. Paired *t*-tests were used to compare each week of testing to pretesting discrimination performance. Discrimination performance at 4 weeks $[t(2)=4.99, p=0.05]$, 8 weeks $[t(2)=4.99, p<0.05]$, and 12 weeks $[t(2)=10.99, p<0.05]$ following kanamycin injections were found to be significantly different from preinjection levels.

These data were also analyzed using a three-way multidimensional scaling (MDS) algorithm (SYSTAT, v7.01, 1997). A matrix of response latencies was constructed where each call in the set was related to every other call in the test set by the response latency. Previous work has shown that a matrix of such response latencies for birds has properties of a similarity matrix in which longer latencies are correlated with greater stimulus similarity (see Dooling and Okanoya, 1995; Dooling *et al.*, 1990). MDS places calls in a multidimensional space in such a way that perceptual similarity is represented by spatial proximity (Dooling and Okanoya, 1995; Shepard, 1980). Two-dimensional MDS solutions are shown in Fig. 4(C) for the combined analysis of the three birds tested on the call set before kanamycin injections, and at postinjection periods of 4 weeks and 6 months. These MDS solutions accounted for 73%, 73%, and 72% of the variance in the birds' response latencies, respectively, before kanamycin treatment, at 4 weeks and at 6 months postinjection.

Before kanamycin treatment, the calls were arranged in perceptual space as pairs, with each natural call close in proximity to its synthetic analog. In other words, the birds discriminated more quickly on trials involving two different contact call types (i.e., calls from different birds such as A versus B) than they did on trials involving a natural contact call and its synthetic analog (i.e., such as A and A′). At 4 weeks following kanamycin treatment, the perceptual space was still significantly distorted by this analysis since two pairs of calls that were previously separated before kanamycin treatment, were now clustered together. At 6 months following kanamycin, however, the birds' perceptual space, in terms of separation of call types, more closely approximated the pattern seen before kanamycin treatment. These results show that frequency and intensity difference limens, as well as relatively easy discriminations between complex sounds such as species-specific vocalizations (contact calls), are relatively unaffected. Even so, a more refined analysis of response latencies which can be used to define perceptual categories show minor perturbations in perceptual space still apparent at 16 weeks.

C. Experiment 3—Recognition tests: Identification of previously familiar contact calls

From the previous experiments, we know that discrimination of frequency and intensity differences in pure tones is relatively unaffected after only four weeks of recovery and discrimination between similar contact call types (as measured by percent correct) also returns to pretreatment levels by that time. However, the analyses of response latencies in discrimination tests with contact calls using MDS also show that the birds' perceptual space for contact calls at this recovery time point is still distorted in that two contact calls that were perceptually distinct prior to kanamycin treatment, sound similar to birds at four weeks following kanamycin.

This suggests a more subtle and complicated effect of hearing loss on the perception of complex sounds in these birds. Taking speech as another example, there are a number of ways to assess the role of hearing loss in speech perception. We know from a variety of perceptual tests that it is one thing to hear speech, or even to discriminate among speech sounds, but it is quite another thing to understand what is being said (e.g., Newby, 1964). The next experiment was designed to explore the phenomenon of complex sound perception during hearing loss from a slightly different perspective. To address this last point, we assessed auditory perception during and after severe, but temporary, threshold shifts using a recognition task involving contact calls. In other words, birds were trained to recognize and classify two different contact calls and then were treated with kanamycin and retested on the recognition of those calls.

1. Methods

In this experiment, six new experimental birds were tested with a different behavioral procedure. This procedure tested the birds on a Go/NoGo task which assessed their ability to classify contact calls before and after kanamycin injections. This task was more difficult than the discrimination task described above in that the birds had to remember

from trial to trial which call was the "Go" call and which call was the "NoGo" call. The natural contact calls used in this test were produced by birds in another flock that the experimental subjects had never heard before. During testing these calls were presented at a rms level of 65 dB SPL.

Testing occurred in the same apparatus as in the detection and discrimination experiments. However, here the budgerigars were required to peck the observation key just once to hear one presentation of one of two contact calls—either a Go call or a NoGo call. A correct response when presented with the Go call was to peck the report key within 2 s to receive food. A correct response when presented with the NoGo call was to withhold responding on the report key for an interval of 5 s. Correct responses on NoGo trials were not rewarded but incorrect responses were punished with a blackout. If the birds pecked the report key within 5 s of the presentation of a NoGo call, the lights in the test chamber were extinguished for 10 s during which another trial could not be initiated.

During initial training, the birds always pecked the report key when presented with either call resulting in an overall performance level of 50% correct (100% correct on the Go trials and 0% correct on the NoGo trials). With continued testing, however, the birds learned to withhold responding when the NoGo contact call was presented. The birds were run in 100 trial sessions until they had achieved a criterion level of at least 85% correct for three successive 100-trial sessions.

Once trained, the six birds were tested daily for five days before and then the ten days during injections. Following the cessation of injections, they were tested again for one session at day 24 and again at 38 days (4 weeks following the end of injections). Daily testing also resumed at day 38 and the number of sessions it took for these birds to reach criterion performance was measured. Four birds were trained and tested on the same pair of contact calls and two birds were tested on a different pair of contact calls.

Prior to being tested, four of the birds were trained to criterion and then given a 2 week pause in testing and then retested on the same task involving the same pair of vocalizations. Two of these birds and two of the others were also retrained to criterion and then given a 4-week pause in testing, and then retested on the same vocalizations. These control tests measuring the birds' performance after a two or four week pause in testing served as a control for the absence of testing during recovery from kanamycin. Past experiments (Park and Dooling, 1985) have shown that birds trained and tested under these conditions maintain much of their performance even after long periods of no testing. An additional group of birds ($n=3$) was trained to criterion on the original pair of contact calls, then switched to a new pair of contact calls, and tested daily until criterion performance was reached on the new vocalizations. This condition also provides important control information for interpreting the effects of hearing loss by kanamycin treatment on vocal recognition by establishing the time it takes to learn to classify a new pair of contact calls.

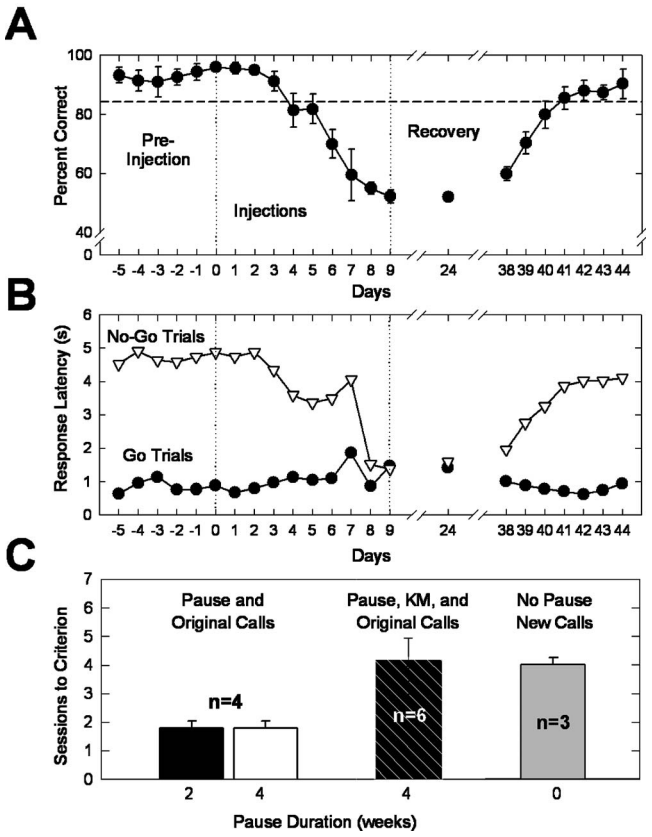


FIG. 6. (A) Average percent correct responses of six budgerigars on a Go/NoGo recognition task involving two contact calls before, during, and after treatment with kanamycin. 4 weeks following the end of kanamycin treatment, classification performance returns to preinjection levels after 4 days of testing. Error bars represent standard errors. (B) The average response latencies to both Go and NoGo stimuli for the six budgerigars before, during, and after kanamycin treatment. (C) Summary of the test sessions required to reach criterion on original calls or new calls with or without antibiotics and a pause in testing. Error bars represent standard errors.

2. Results

Figure 6(A) shows the average percent correct for the six budgerigars on the classification task before, during and after ten days of kanamycin injections. Prior to injections of kanamycin, the birds' performance was well above the 85% criterion level. However, performance falls to chance levels (50%) after several days of kanamycin injections and remains there when assessed in a single 100-trial test session 24 and 38 days later. At 38 days (4 weeks following cessation of kanamycin injections), the birds were retrained and tested daily on the same task in 100-trial sessions. The birds returned to preinjection performance levels in four days.

Response latencies for the Go stimuli remained below 1 s and were relatively unchanged throughout the experiment, attesting to the birds' excellent health, attentiveness, and behavioral responsiveness [Fig. 6(B)]. Response latencies to the NoGo stimuli averaged near 5 s (the duration of the response interval) before and during the first few days of treatment with kanamycin. As testing continued, the birds began to respond by pecking the report key to both Go and NoGo stimuli. Response latencies to the NoGo stimuli approached the levels recorded to the Go stimuli, indicating that the birds were unable to recognize the NoGo stimuli and performance fell to chance responding.

The six kanamycin-treated birds tested after four weeks following kanamycin treatment required an average of over four 100-trial sessions to reach criterion instead of the average of less than two 100-trial sessions to reach criterion following a 4 week pause in testing but with no kanamycin treatment. These results are summarized in Fig. 6(C). The first two control conditions are for the four noninjected birds that were given either a 2-week or a 4-week pause in testing. Because these calls were familiar, the time to criterion was much less than the time required to learn the new classification. Birds performing above criterion on one pair of contact calls (original calls) and switched to another pair (new calls) also required slightly over four 100-trial sessions to reach criterion. These results suggest that previously familiar contact calls do not sound the same to budgerigars who have been treated with kanamycin even though they have undoubtedly regained their full complement of hair cells. Instead, the birds behave as if these calls sound unfamiliar as indicated by the time required to relearn the classification. While the exact cause of this phenomenon is unclear, we know even at 2–3 months postkanamycin treatment, some hair cell abnormalities remain in the basal portion of the papilla including multiple and/or abnormal stereovilli bundles, and abnormal stereovilli bundle orientation, some immature hair cells, and an irregular pattern of hair cells as has been reported in other avian species after ototoxic drug administration (e.g. Cotanche, 1999).

V. DISCUSSION

An earlier study demonstrated that budgerigars show changes in vocalizations produced under operant control around the time when hair cell loss from aminoglycoside administration was greatest (Dooling *et al.*, 1997). Here, we address the effect of a severe but temporary threshold shift on the hearing and auditory perception of vocalizations in budgerigars in more detail. Absolute thresholds in budgerigars recover considerably by 4–8 weeks following aminoglycoside administration. These results are generally similar to those found previously both behaviorally or physiologically in budgerigars as well as several other species of birds including starlings, chicks, and Bengalese finches (Dooling *et al.*, 1997; Hashino and Sokabe, 1989; Marean *et al.* 1993; 1998; Tucci and Rubel, 1990). The time course of hearing recovery and the amount of permanent threshold shift in budgerigars suggests that they may be more sensitive to damage from ototoxic antibiotics than are starlings or Bengalese finches (Marean *et al.* 1993; 1998; Woolley *et al.*, 2001). The present findings also show that frequency and intensity difference limens 4 weeks after kanamycin treatment in budgerigars are already at near normal levels and remain so for the duration of testing out to as long as four months. We could infer from these data using relatively simple sounds that the recovery of auditory function with regeneration of hair cells is nearly complete, and there are only relatively small residual effects after kanamycin-induced trauma.

The present experiments addressed how temporary hearing loss affects the discrimination and the perception of more complex sounds in birds. We expected that the effects of hearing loss on the perception of species-specific vocalizations are more relevant and more complicated than the effects on the detection and discrimination of simple sounds. Previous work has shown that, as in humans, hearing loss in birds is accompanied by an increase in critical ratios (Hashino and Sokabe, 1989) and a broadening of auditory filters and changes in temporal resolution (Marean *et al.*, 1998). As hair cells regenerated, both of these deficits eventually improved to near pre-treatment levels, suggesting that complex sound perception may also return to near normal as new hair cells develop.

In our experiments, discrimination among contact calls of different birds as measured by percent correct performance in an operant task was near 100% before kanamycin treatment and at 4 weeks after kanamycin treatment and remained at high levels out to 6 months. Difficult discriminations, such as between a natural contact call and its synthetic analog (a task that human listeners find challenging) were affected for up to twenty weeks following kanamycin treatment. A more refined analysis using MDS of the response latencies of birds working in this task showed that the perceptual map of contact calls at 4 weeks of recovery still showed overlapping categories of contact calls that were perceptually distinct prior to kanamycin administration. As far as we know, these are the first experiments to examine how perceptual categories of species-specific vocalizations are affected following hair cell loss and regeneration. In budgerigars, at least, the results show a large recovery within a few weeks but some residual perceptual problems that persist for up to 14–16 weeks.

A common refrain of hearing impaired humans is that speech can be heard as well as before the hearing loss occurred but cannot be as easily understood (Newby, 1964). This is an intriguing phenomenon and one that is even more interesting in an organism that has the capability of auditory hair cell regeneration. When trained to recognize two different contact calls, budgerigars completely lose the ability to correctly label these calls when their hair cells have been destroyed by injections with kanamycin. Four weeks into recovery, these birds quickly relearn the classification of previously familiar contact calls to a high level but they do so with a time course that suggests that these previously familiar calls now sound unfamiliar. In other words, although the ability to detect, discriminate, and classify complex acoustic sounds approaches pretreatment levels 4 weeks into recovery, the perceptual world of vocalizations is not the same as before hair cells were lost. These findings have relevance for hearing restoration efforts in humans since they suggest an enduring change in an organism's auditory world of vocalizations in spite of a new set of hair cells and relatively normal hearing otherwise.

ACKNOWLEDGMENTS

The authors thank Beth Brittan-Powell, Amanda Lauer, Marjorie Leek, and Liana Tang for data collection and comments on the manuscript. Support from NIDCD (Grant No. DC-01372) for two of the authors, (R. J. D. and B. M. R.)

and a NIDCD Institutional Training Grant (No. 5T32-DC-00046) for another author (M.L.D) is acknowledged.

Beeman, K. (**1992**). "Signal technology, V2.20.," (Engineering Design, Belmont, MA).

Bermingham-McDonogh, O., and Rubel, E. W. (**2003**). "Hair cell regeneration: Winging our way towards a sound future," Curr. Opin. Neurobiol. **13**, 119–126.

Brittan-Powell, E. F., Dooling, R. J., and Farabaugh, S. M. (**1997**). "Vocal development in budgerigars (*Melopsittacus undulatus*): contact calls," J. Comp. Psychol. **111**, 226–241.

Corwin, J. T., and Cotanche, D. A. (**1988**). "Regeneration of sensory hair cells after acoustic trauma," Science **240**, 1772–1774.

Cotanche, D. A. (**1999**). "Structural recovery from sound and aminoglycoside damage in the avian cochlea," Audiol. Neuro-Otol. **4**, 271–285.

Dooling, R. J. (**1986**). "Perception of vocal signals by budgerigars," Exp. Biol. **45**, 195–218.

Dooling, R. J., Gephart, B. F., Price, P. H., McHale, C., and Brauth, S. E. (**1987**). "Effects of deafening on the contact call of the budgerigar, *Melopsittacus undulatus*," Anim. Behav. **35**, 1264–1266.

Dooling, R. J., Brown, S. D., Park, T. J., and Okanoya, K. (**1990**). "Natural perceptual categories for vocal signals in budgerigars (*Melopsittacus undulatus*)," in *Comparative Perception*, Complex Signals. Vol. **II**, edited byW. Stebbins, M. A. Berkley (Wiley, NewYork).

Dooling, R. J., and Okanoya, K. (**1995**). "The method of constant stimuli in testing auditory sensitivity in small birds," in *Methods in Comparative Psychoacoustics*, edited by G. M. Klump, R. J. Dooling, R. R. Fay, and W. C. Stebbins (Birkhauser-Verlag, Switzerland), pp. 161–169.

Dooling, R. J., Ryals, B. M., and Manabe, K. (**1997**). "Recovery of hearing and vocal behavior after hair cell regeneration," Proc. Natl. Acad. Sci. U.S.A. **94**, 14206–14210.

Farabaugh, S. M., Linzenbold, A., and Dooling, R. J. (**1994**). "Vocal plasticity in budgerigars (*Melopsittacus undulatus*): Evidence for social factors in the learning of contact calls," J. Comp. Psychol. **108**, 81–92.

Farabaugh, S., and Dooling, R. J. (**1996**). "Acoustic communication in parrots: Laboratory and field studies of budgerigars, *Melopsittacus undulatus*," in *Ecology and Evolution of Acoustic Communication in Birds*, edited byD. E. Kroodsma and E. H. Miller (Cornell University Press, Ithaca, NY), pp. 97–118.

Gleich, O., and Manley, G. A. (**1988**). "Quantitative morphological analysis of the sensory epithelium of the starling and pigeon basilar papilla," Hear. Res. **34**, 69–85.

Gleich, O., Manley, G. A. (**2000**).' "The hearing organ of birds and crocodilia," in: *Comparative Hearing: Birds and Reptiles* edited byR. J. Dooling, R. R. Fay, and A. N. Popper (Springer, New York), pp. 70–138, 2000).' "The hearing organ of birds and crocodilia," in: *Comparative Hearing: Birds and Reptiles* edited byR. J. Dooling, R. R. Fay, and A. N. Popper (Springer, New York), pp. 1139–1149.

Hashino, E., and Sokabe, M. (**1989**). "Kanamycin-induced low-frequency hearing loss in the budgerigar (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **85**, 289–294.

Hashino, E., Sokabe, M., and Tanaka, R. (**1991**). "Function-structure correlation during recovery from aminoglycoside ototoxicity in the avian auditory system," in *Noise-Induced Hearing Loss*, edited byA. L. Dancer, D. Henderson, R. J. Salvi, and R. P. Hammernik, Mosby Year Book, St. Louis, MO, pp. 228–236.

Hashino, E., Tanaka, R., Salvi, R. J., and Sokabe, M. (**1992**). "Hair cell regeneration in the adult budgerigar after kanamycin ototoxicity," Hear. Res. **59**, 46–58.

Heaton, J. T., and Brauth, S. E. (**1999**). "Effects of deafening on the development of nestling and juvenile vocalizations in budgerigars (*Melopsittacus undulatus*)," J. Comp. Psychol. **113**, 314–320.

Heaton, J. T., Dooling, R. J., and Farabaugh, S. M. (**1999**). "Effects of deafening on the calls and warble song of adult budgerigars (*Melopsittacus undulatus*)," J. Acoust. Soc. Am. **105**, 2010–2019.

Hile, A. G., Plummer, T. K., and Streidter, G. F. (**2000**). "Male vocal imitation produces call convergence during pair bonding in budgerigars (*Melopsittacus undulatus*)," Anim. Behav. **59**, 1209–1218.

Kroodsma, D. E., and Miller, E. H. (**1996**). *Ecology and Evolution of Acoustic Communication in Birds* (Cornell University Press, Ithaca, New York).

Lippe, W. R., Westbrook, E. W., and Ryals, B. M. (**1991**). "Hair cell regeneration in the chick cochlea following aminoglycoside ototoxicity," Hear. Res. **56**, 203–210.

Manabe, K., Sadr, E. I., and Dooling, R. J. (**1998**). "Control of vocal intensity in budgerigars (*Melopsittacus undulatus*): Differential reinforcement of vocal intensity and the Lombard effect," J. Acoust. Soc. Am. **103**, 1190–1198.

Marean, G. C., Burt, J. M., Beecher, M. D., and Rubel, E. W. (**1993**). "Hair cell regeneration in the European starling (*Sturnus vulgaris*): Recovery of pure-tone detection thresholds," Hear. Res. **71**, 125–136.

Marean, G. C., Burt, J. M., Beecher, M. D., and Rubel, E. W. (**1998**). "Auditory perception following hair cell regeneration in European starling (*Sturnus vulgaris*): Frequency and temporal resolution," J. Acoust. Soc. Am. **103**, 3567–3580.

Newby, H. A. (**1964**). *Disorders of Hearing; in Audiology* (Appleton-Century-Crofts, New York) pp. 45–53.

Niemiec, A. J., Raphael, Y., and Moody, D. B. (**1994**). "Return of auditory function following structural regeneration after acoustic trauma: Behavioral measures from quail," Hear. Res. **79**, 1–16.

Okanoya, K., and Dooling, R. J. (**1987**). "Hearing in Passerine and Psittacine birds: A comparative study of masked and absolute auditory thresholds," J. Comp. Psychol. **101**, 7–15.

Okanoya, K., and Dooling, R. J. (**1991**). "Perception of distance calls by budgerigars (*Melopsittacus undulatus*) and zebra finches (*Poephila guttata*): Assessing species-specific advantages," J. Comp. Psychol. **105**, 60–72.

Park, T. J., and Dooling, R. J. (**1985**). "Perception of species-specific contact calls by budgerigars (*Melopsittacus undulatus*)," J. Comp. Psychol. **99**, 391–402.

Ryals, B. M., and Rubel, E. W. (**1988**). "Hair cell regeneration after acoustic trauma in adult coturnix quail," Science **240**, 1774–1776.

Ryals, B. M., Dooling, R. J., Westbrook, E., Dent, M. L., MacKenzie, A., and Larsen, O. N. (**1999**). "Avian species differences in susceptibility to noise exposure," Hear. Res. **131**, 71–88.

Saunders, S. S., Salvi, R. J., and Miller, K. M. (**1995**). "Recovery of thresholds and temporal integration in adult chickens after high-level 525-Hz pure-tone exposure," J. Acoust. Soc. Am. **97**, 1150–1164.

Saunders, J. C., Doan, D. E., Cohen, Y. E., Adler, H. J., and Poje, C. P. (**1996**). "Recent observations on the recovery of structure and function in the sound damaged chick ear," in *Auditory System Plasticity and Regeneration*, edited byR. J. Salvi, D. Henderson, F. Fiorino, V. Colletti (Thieme Medical Publishers, New York), pp. 62–84.

Shepard, R. N. (**1980**). "Multidimensional scaling, tree-fitting, and clustering," Science **210**, 390–398.

Tucci, D. L., and Rubel, E. W. (**1990**). "Physiological status of regenerated hair cells in the avian inner ear following aminoglycoside ototoxicity," Otolaryngol.-Head Neck Surg. **103**, 443–445.

Woolley, S. M., and Rubel, E. W. (**1999**). "High-frequency auditory feedback is not required for adult song maintenance in Bengalese finches," J. Neurosci. **19**, 358–371.

Woolley, S. M., Wissman, A. M., and Rubel, E. W. (**2001**). "Hair cell regeneration and recovery of auditory thresholds following aminoglycoside ototoxicity in Bengalese finches," Hear. Res. **153**, 181–195.